General Internal Medicine
Boston University School of Medicine
2009 Publications - A-M

1. **Alford DP**. Opioids for chronic pain in patients with substance abuse: Too much, too little or just right? *Invited Commentary.* Pain. 2009;145:267-68.

2. **Alford DP**, **Bridden C**, **Jackson AH**, **Saitz R**, Amodeo M, Barnes HN, **Samet JH**. Promoting substance use education among generalist physicians: An evaluation of the Chief Resident Immersion Training (CRIT) program. Gen Intern Med. 2009;24(1):40-47.

3. **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.

4. **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.

5. **Bertholet N**, Horton NJ, **Saitz R**. Improvements in readiness to change and drinking in primary care patients with unhealthy alcohol use: a prospective study. BMC Public Health, 2009;9:101 PMCID: PMC2679737.

6. **Berz JPB**, Johnston K, Backus B, Doros G, **Rose AJ**, Pierre S, **Battaglia TA**. The influence of black race on treatment and mortality for early-stage breast cancer. Med Care. 2009;47(9):986-92.

7. Bickmore TW, Pfeifer LM, **Paasche-Orlow MK**. Using computer agents to explain medical documents to patients with low health literacy. Patient Educ Couns. Jun 2009;75(3):315-20.

8. Blackburn GL, Hutter MM, Harvey AM, Apovian CM, Boulton HRW, Cummings S, Fallon JA, Greenberg I, Jiser ME, Jones DB, Jones SB, Kaplan LM, Kelly JJ, Kruger RS Jr, Lautz DB, Lenders CM, LoNigro R, Luce H, McNamara A, Mulligan AT, **Paasche-Orlow MK**, Perna FM, Pratt JSA, Riley SM Jr, Robinson MK, Romanelli JR, Saltzman E, Schumann R, Shikora SA, Snow RL, Sogg S, Sullivan MA, Tarnoff M, Thompson CC, Wee CC, Ridley N, Auerbach J, Hu FB, Kirle L, Buckley RB, Annas CL. Expert panel on weight loss surgery: executive report update. Obesity (Silver Spring). May 2009;17(5):842-62.

9. Bokhour BG, Solomon JL, Knapp H, Asch SM, **Gifford AL**. Barriers and facilitators to routine HIV testing in VA Primary Care. J Gen Intern Med. 2009;24(10):1109-14.

10. Braverman J, **Dedier J**. Predictors of Medication Adherence for African Americans Patients Diagnosed with hypertension. Ethnicity and Disease. Autumn 2009;19:396-400.

11. Campbell CA, Hahn RA, Elder R, Brewer R, Chattopadhyay S, Fielding J, **Naimi T**, Toomey T, Lawrence B, Middleton JC, et al. The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms. Am J Prev Med. 2009;37(6):556–69.

12. **Carr PL**, Pololi L, Knight S, Conrad P. Collaborating in academic medicine: Reflections on gender and advancement. Acad Med. 2009;84(10):1447-53.

13. Chaudhry AA, **Gifford AL**, Cofrancesco J Jr, Sosman J, Berkenblit G, Sullivan LE. Update in HIV medicine for the generalist. J Gen Intern Med. 2009;24(2):276-82.

14. **Cheng DM**, Libman H, **Bridden C**, **Saitz R**, **Samet JH**. Alcohol consumption and lipodystrophy in HIV-infected adults with alcohol problems. Alcohol. 2009;43:65-71.

15. **Cheng T**, **Freund KM**, Lane SE, Ganschow P, **Battaglia TA**. Screening for Breast Cancer. http://pier .acrponline.org/physicians/screening/s369/s369.html. In PIER [online database]. Philadelphia, American College of Physicians, 2009.

16. Cremeens J, Nelson D, **Naimi T**, et al. Sociodemographic differences in binge drinking among adults – 14 states, 2004. Morbidity and Mortality Weekly Report. 2009;58(12):301-4.

17. Doe-Simkins M, **Walley AY**, Epstein A, Moyer P. Saved by the nose: bystander-administered intranasal naloxone hydrochloride for opioid overdose. American Journal of PH. May 2009;99(5):788-91.

18. **Dookeran NM**, Burgess JF Jr, Bowman CC, Goetz MB, Asch SM, **Gifford AL**. HIV screening among substance-abusing veterans in care. J Subst Abuse Treat. Oct 2009;37(3):286-91.

19. Douglass MA, Sanchez GM, **Alford DP**, Wilkes G, **Greenwald JL**. Physicians' pain management confidence versus competence. J Opioid Management. 2009;5(3):169-74.

20. Freedman JE, **Hylek EM**. Clopidogrel, genetics and drug responsiveness [Invited Editorial]. N Engl J Med. 2009;360:411-3.

21. Garcia D and **Hylek EM**. Warfarin pharmacogenetics. N Engl J Med. 2009;360:2474-75.

22. **Gerber MR**, Tan AKW. Lifetime intimate partner violence exposure, attitudes and comfort among Canadian health professions students. BMC Res Notes. Sept 2009;2:191. PMID: 19775437.

23. Giordano TP, Hartman C, **Gifford AL**, Backus LI, Morgan RO. Predictors of retention in HIV care among a national cohort of US veterans. HIV Clin Trials. Sep-Oct 2009;10(5):299-305.

24. Goetz MB, Hoang T, Henry SR, Knapp H, Anaya HD, **Gifford AL**, Asch SM, QUERI-HIV/Hepatitis Program. Evaluation of the sustainability of an intervention to increase HIV testing. J Gen Intern Med. Dec 2009;24(12):1275-80.

25. **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.

26. Hausmann LRM, Jeong K, Bost JE, **Kressin NR**, Ibrahim SA. Perceived racial discrimination in health care: A comparison of Veterans Affairs and other patients. American Journal of Public Health 2009;99(7):1-7.

27. **Hermos J**, Winter M, Heeren T, Hingson R. Alcohol-related problems among younger drinkers who misuse prescription drugs: results from the National Epidemiologic Survey of Alcohol and Related Conditions (NESARC). Substance Abuse. 2009;30:118-26.

28. Hironaka LK, **Paasche-Orlow MK**, Young RL, Bauchner H, Geltman PL. Caregiver health literacy and adherence to a daily multi-vitamin with iron regimen in infants. Patient Educ Couns. Jun 2009;75(3):376-80.

29. **Hylek EM**. Antithrombotic prophylaxis in elderly patients with atrial fibrillation. Semin Thromb Hemost. Sep 2009;35(6):548-53.

30. **Hylek EM**. The need for new oral anticoagulants in clinical practice. J Cardiovasc Med 2009;10(8):605-9.

31. **Hylek EM**, **Rose AJ**. Understanding Low INR in Clinical Practice [Invited Editorial]. Thromb Haemost 2009;101:417-18.

32. Jack BW, Chetty VK, Anthony D, **Greenwald JL**, Sanchez GM, Johnson AE, Forsythe SR, O'Donnell JK, **Paasche-Orlow MK**, Manasseh C, Martin S, Culpepper L. A reengineered hospital discharge program to decrease rehospitalization: a randomized trial. Ann Intern Med. Feb 2009;150(3):178-87.

33. **Kapoor A**, Kraemer KL, Smith KJ, Roberts MS, **Saitz R**. Cost-Effectiveness of Screening for Unhealthy Alcohol Use with %Carbohydrate Deficient Transferrin: Results From a Literature-Based Decision Analytic Computer Model. Alcohol Clin Exp Res. Aug 2009;33(8):1-10.

34. Katz RV, Green BL, **Kressin NR,** James SA, Sang MQ, Claudio C, Russell SL. Exploring the 'legacy' of the Tuskegee Syphilis Study: a follow-up study from the Tuskegee Legacy Project. Journal of the National Medical Association. 2009:101(2):179-83.

35. Kertesz SG, Posner MA, **O'Connell JJ**, Swain S, Mullins AN, Shwartz M, **Ash AS**. Post-Hospital Medical Respite Care and Hospital Readmission of Homeless Persons. 2009. J Prev Interv Community. 2009;37(2):129-42.

36. **Kim TW**, **Alford DP**, Holick MF, Malabanan AO, **Samet JH**. Low vitamin d status in methadone maintenance patients. J Addict Med. 2009;3(3):134-38.

37. Knapp PE, Showers KM, Phipps JC, Speckman JL, Sternthal E, **Freund KM**, **Ash AS**, Apovian CM. Self-Monitoring of Blood Glucose with Finger tip vs. Alternative site sampling: Effect on glycemic control in insulin-using patients with type-2 diabetes. Diabetes Technology and Therapeutics. 2009;11(4):219-25.

38. **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. Medical Care. 2009;47(11):1121-28.

39. **Lasser KE**. Update in smoking and mental illness: A primary care perspective. Dual Diagnosis. 2009;5(2):191-96.

40. **Lasser KE**, Murillo J, Medlin E, Lisboa S, Valley-Shah L, Fletcher RH, Emmons KE, Ayanian JZ. A multilevel intervention to promote colorectal cancer screening among community health center patients: results of a pilot study. BMC Family Practice. 2009;10:37.

41. Lisabeth LD, Beiser AS, Brown DL, **Murabito JM**, Kelly-Hayes M, Wolf PA. Age at natural menopause and risk of ischemic stroke: the Framingham Heart Study. Stroke. Apr 2009;40(4):1044-9. PMID: 19233935; PubMed Central PMCID: PMC2682709.

42. Lopes RD, Becker RC, Garcia D, **Hylek EM**, Granger CB, Lourenço DM, Nader HB, Maffei FH, Cesena FHY, Nicolau JC, Timmerman A, Rocha AT, de Paiva EF, Guimaraes HP, Carvalho AC, D'Amico EA, Moreira-Filho CA, Aldrighi JM, Lopes AC. Highlights from the I International Symposium of Thrombosis and Anticoagulation in Internal Medicine, October 23-25, 2008, Sao Paulo, Brazil. J Thromb Thrombolysis. March 2009;28:106-116.

43. Loucks EB, Lynch JW, Pilote L, Fuhrer R, Almeida ND, Richard H, Agha G, **Murabito JM**, Benjamin EJ. Life-course socioeconomic position and incidence of coronary heart disease: the Framingham Offspring Study. Am J Epidemiol. Apr 2009;169(7):829-36.

44. **Lunze K**. Health and human rights: no miracle in postconflict Chechnya. Lancet. 2009;374:1809-10.

45. Mahabadi AA, Massaro JM, Rosito GA, Levy D, **Murabito JM**, Wolf PA, O'Donnell CJ, Fox CS, Hoffmann U. Association of pericardial fat, intrathoracic fat, and visceral abdominal fat with cardiovascular disease burden: the Framingham Heart Study. Eur Heart J. Apr 2009;30(7):850-6.

46. **McCall-Hosenfeld JS**, **Freund KM**, **Liebschutz JM**. Factors associated with sexual assault and time to presentation. Prev Med. 2009;48(6):593-5.

47. **McCall-Hosenfeld JS**, **Liebschutz JM**, Spiro A III, Seaver MR. Sexual assault in the military and its impact on sexual satisfaction in women veterans: A proposed model. J of Women's Health. 2009;18(6):901-9.

48. **Meisel JL**. Differential diagnosis of chest pain (Parts I and II). In: Rind D, Aronson M eds. Up to Date in adult primary care and internal medicine. Published online.

49. **Meisel JL**. (Letter to the editor.) CHADS2 (85x2). Annals Int Med 2009, published electronically at http://www.annals.org/cgi/eletters/151/5/297

50. **Murabito JM**, Keyes MJ, Guo CY, Keaney JF Jr, Vasan RS, D'Agostino RB Sr, Benjamin EJ. Cross-sectional relations of multiple inflammatory biomarkers to peripheral arterial disease: The Framingham Offspring Study. Atherosclerosis. Apr 2009;203(2):509-14.

PAIN® 145 (2009) 267–268

PAIN®

www.elsevier.com/locate/pain

Commentary

Opioids for chronic pain in patients with substance abuse: Too much, too little or just right?

Over the past two decades, opioid analgesic prescriptions for chronic noncancer pain have increased [12]. Medical literature supporting this practice began in the 1980s [13], followed by aggressive marketing of "safe" sustained-released opioids to primary care physicians. Increases in opioid prescribing continued despite lack of strong evidence supporting this practice [9] and subsequent increases in rates of opioid misuse including addiction and overdose [3,16,17].

The paper by Weisner et al. [18] confirms this dramatic increase in long-term opioid analgesic prescribing over an eight year period in two large health plans representing community practices. Importantly, they found that patients with a prior substance abuse history were about 4 times more often prescribed opioids than those without a substance abuse history. This prevalence increased by 7–8 times for patients with an opioid use disorder with over half of these patients prescribed opioids. Patients with a substance abuse history were prescribed opioids with higher potency and at higher dosages. They were also twice as likely to be concurrently prescribed sedative hypnotics. Of concern is the assertion by the authors that primary care physicians prescribing opioids may not have been aware of their patient's substance abuse history. Often these diagnoses were made in mental health and substance abuse treatment settings where restrictive privacy regulations prevent full communication of these diagnoses with primary care physicians. These findings are concerning in light of observational studies that have found that a substance abuse history significantly increases the risk for prescription opioid misuse [11]. This paper adds a great deal to our understanding of the magnitude of high-risk opioid prescribing by community-based physicians.

It is unlikely that all patients with a history of substance abuse share the same level of risk for prescription opioid misuse. A limitation of this study is the inability to differentiate between patients who are in stable recovery from their substance abuse from those who are not. For patients in recovery, relapse prevention theories would suggest that unrelieved pain is more likely to trigger relapse than adequate analgesia. For patients with active substance abuse, the potential risks i.e. prescription opioid abuse and/or diversion, may outweigh any potential benefits. Guidelines state that patients who are "actively using illicit drugs" should be treated in "highly controlled and specialized settings" and co-managed with an addiction expert [4]. Unfortunately addiction experts are not readily available to most primary care physicians and when available, their level of expertise and interest in managing patients with chronic pain is highly variable.

So what is the correct prevalence of opioid prescribing for the management of chronic pain in patients with a history of substance abuse? This is obviously a controversial and contentious issue as exemplified by a survey of state medical boards where the majority of respondents did not consider the use of long-term opioid analgesics in patients with substance abuse to fall within the scope of acceptable medical practice [6]. This question is particularly pertinent since chronic pain and substance abuse can be related phenomena. Over forty years ago, Martin and Inglis [10] observed that patients with opioid addiction self-medicate "an abnormally low tolerance for painful stimuli". The presence of one condition seems to influence the expression of the other. Savage and Schofferman [15] found that persons with addiction and pain have a "syndrome of pain facilitation." Their pain experience is worsened by withdrawal-related sympathetic nervous system arousal, sleep disturbances, and affective changes, all consequences of addictive disease. Supporting a negative effect of addiction on pain tolerance, patients who abuse stimulants and those who abuse opioids have been shown to be less tolerant of pain than their peers in remission [5]. Studies have consistently found that patients with substance abuse histories have an unusually high prevalence of chronic pain [14] and are more likely to have their pain under-treated [1]. This under-treatment has been reported as a reason for initiating and continuing illicit drug use [8]. While opioids are not indicated or effective for all chronic pain, it is unlikely that pain in patients with substance abuse is any less opioid responsive than pain in patients without a history substance abuse. Universally withholding opioid analgesics as well as ignoring their risks in patients with substance abuse would constitute poor clinical care.

Although this study raises concerns about opioid over-prescribing in high risk patients, we do not know what is happening behind the scenes. Are opioids being started appropriately and continued based on improved clinical outcomes? Are patients being closely monitored for signs of prescription opioid misuse or abuse? Pain and addiction society guidelines [4] recommend performing a careful initial assessment including screening for unhealthy substance use. Historically substance use screening tools have been lengthy and impractical for primary care physicians however shorter screening [2] instruments are available. Also recommended are the use of controlled substance agreements to inform patients about the expected benefits and risks of opioids, and that treatment goals include close monitoring for improved function and any addictive behaviors. Because the consequences of opioid analgesic misuse are so serious and because the true risk cannot be reliably predicted, it has been suggested that physicians utilize "universal precautions" when starting and maintaining patients on opioids [7]. That is, monitoring all patients as if they all have the potential for developing prescription opioid abuse.

Because of the uncertain benefits of using long-term opioids for managing chronic pain especially in patients with substance abuse

0304-3959/$36.00
doi:10.1016/j.pain.2009.06.036

histories, primary care physicians should universally adopt guideline-based practices. Implementation of these guidelines will not be easy as they will require system changes including developing specific policies and procedures for universally screening and assessing for substance use disorders, using patient agreements and monitoring for benefits and risks. Physician educational initiatives must address when and how to use opioids safely and when and how to discontinue opioids when there is lack of benefit and/or apparent harm. Finally, primary care physicians must learn communication skills for discussing opioid misuse and abuse with patients. Because these discussions are potentially uncomfortable, they are often delayed, addressed poorly or never addressed at all.

References

[1] Breitbart W, Rosenfeld BD, Passik SD, McDonald MV, Thaler H, Portenoy RK. The undertreatment of pain in ambulatory AIDS patients. Pain 1996;65:243–9.

[2] Brown RL, Leonard TARCS, Saunders LAMSSW, Papasouliotis O. A two-item conjoint screen for alcohol and other drug problems. J Am Board Fam Pract 2001;14:95–106.

[3] Centers for Disease Control and Prevention. Unintentional poisoning deaths: United States, 1999–2004. MMWR 2007;56:93–96. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5605a1.htm [Accessed 20.05.2009.].

[4] Chou R, Fanciullo GJ, Fine PG, Adler JA, Ballantyne JC, Davies P, Donovan MI, Fishbain DA, Foley KM, Fudin J, Gilson AM, Kelter A, Mauskop A, O'Connor PG, Passik SD, Pasternak GW, Portenoy RK, Rich BA, Roberts RG, Todd KH, Miaskowski C. 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.

[5] Compton MA. Cold-pressor pain tolerance in opiate and cocaine abusers: correlates of drug type and use status. J Pain Symptom Manage 1994;9:462–73.

[6] Gilson AM, Maurer MA, Joranson DE. State medical board members' beliefs about pain, addiction, and diversion and abuse: a changing regulatory environment. J Pain 2007;8:682–91.

[7] Gourlay DL, Heit HA, Almahrezi A. Universal precautions in pain medicine: a rational approach to the treatment of chronic pain. Pain Med 2005;6:107–12.

[8] Karasz A, Zallman L, Berg K, Gourevitch M, Selwyn P, Arnsten JH. The experience of chronic severe pain in patients undergoing methadone maintenance treatment. J Pain Symptom Manage 2004;28:517–25.

[9] 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 2007;146:116–27 [see comment].

[10] Martin JE, Inglis J. Pain tolerance and narcotic addiction. Br J Soc Clin Psychol 1965;4:224–9.

[11] Michna E, Ross EL, Hynes WL, Nedeljkovic SS, Soumekh S, Janfaza D, Palombi D, Jamison RN. Predicting aberrant drug behavior in patients treated for chronic pain: importance of abuse history. J Pain Symptom Manage 2004;28:250–8.

[12] Olsen Y, Daumit GL, Ford DE. Opioid prescriptions by U.S. primary care physicians from 1992 to 2001. J Pain 2006;7:225–35.

[13] Portenoy RK, Foley KM. Chronic use of opioid analgesics in non-malignant pain: report of 38 cases. Pain 1986;25:171–86.

[14] Rosenblum A, Joseph H, Fong C, Kipnis S, Cleland C, Portenoy RK. Prevalence and characteristics of chronic pain among chemically dependent patients in methadone maintenance and residential treatment facilities. JAMA 2003;289:2370–8.

[15] Savage SR, Schofferman J. Pharmacological therapies of pain in drug and alcohol addictions. In: Miller N, Gold M, editors. Pharmacological therapies for drug and alcohol addictions. New York: Dekker; 1995. p. 373–409.

[16] Substance Abuse and Mental Health Services Administration, Office of Applied Studies. The NSDUH report: patterns and trends in nonmedical prescription pain reliever use: 2002 to 2005. Rockland, MD; 2007. Available from: http://www.oas.samhsa.gov/2k7/pain/pain.pdf [Accessed 20.05.2009.].

[17] Substance Abuse and Mental Health Services Administration, Office of Applied Studies. Treatment Episode Data Set (TEDS) highlights: 2006 national admissions to substance abuse treatment services. OAS Series #S-40, DHHS Publication No. (SMA) 08-4313. Rockville, MD; 2007. Available from: http://www.oas.samhsa.gov/TEDS2k6highlights/toc.cfm [Accessed 20.05.2009.].

[18] Weisner C, Campbell CI, Ray GT, Saunders K, Merrill JO, Banta-Green C, Sullivan MD, Silverberg MJ, Mertens JR, Boudreau D, Von Korff M. Trends in prescribed opioid therapy for non-cancer pain for individuals with prior substance use disorders. Pain 2009;145:287–93.

Daniel P. Alford
Boston University School of Medicine, Boston Medical Center, Boston, MA 02118, USA
Tel.: +1 617 414 3747; fax: +1 617 414 4676.
E-mail address: Dan.Alford@bmc.org

Promoting Substance Use Education Among Generalist Physicians: An Evaluation of the Chief Resident Immersion Training (CRIT) Program

Daniel P. Alford, MD, MPH[1], Carly Bridden, MA, MPH[1], Angela H. Jackson, MD[1], Richard Saitz, MD, MPH[1,2], Maryann Amodeo, MSW, PhD[3], Henrietta N. Barnes, MD[4], and Jeffrey H. Samet, MD, MA, MPH[1,5]

[1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA; [2]Youth Alcohol Prevention Center and the Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; [3]Center for Addictions Research and Services, Boston University School of Social Work, Boston, MA, USA; [4]Department of Medicine, Cambridge Health Alliance, Harvard Medical School, Boston, MA, USA; [5]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA.

BACKGROUND: Education about substance use (SU) disorders remains inadequate in medical training.

OBJECTIVE: To describe the Chief Resident Immersion Training (CRIT) program in addiction medicine and to evaluate its impact on chief resident (CR) physicians' substance use knowledge, skills, clinical practice, and teaching.

DESIGN: A controlled educational study of CRIT programs (2003, 2004, and 2005) for incoming CRs in generalist disciplines. Intervention CRs were trained to diagnose, manage, and teach about SU. The control CRs sought but did not receive the intervention.

PARTICIPANTS: Eighty-six CR applicants to the CRIT program.

MEASUREMENTS: Baseline and 6-month questionnaires assessing substance use knowledge, skills, clinical practice, and teaching. Outcomes were compared within groups from baseline to follow-up and between groups at follow-up.

RESULTS: The intervention (n=64) and control (n=22) CRs were similar demographically. At 6-month follow-up, the intervention CRs reported a significant increase in SU knowledge, confidence, and preparedness to diagnose, manage, and teach and an increase in SU clinical and teaching practices compared to their baseline and control CRs.

CONCLUSIONS: This intensive training for chief residents (CRs) improved knowledge, confidence, and preparedness to diagnose, manage, and teach about substance use (SU), affecting both the CRs' SU clinical and teaching practices. The CRIT program was an effective model for dissemination of SU knowledge and skills to educators in a key position to share this training with a broader audience of medical trainees. This model holds potential to address other high priority medical, yet under-addressed, content areas as well.

KEY WORDS: medical education; addiction; chief resident training; substance-related disorders.
J Gen Intern Med 24(1):40–7
DOI: 10.1007/s11606-008-0819-2

Received June 5, 2008
Revised September 4, 2008
Accepted September 12, 2008
Published online October 21, 2008

INTRODUCTION

Education about substance use (SU) disorders remains inadequate in medical training.[1] This deficiency persists despite the contribution of SU to disability and premature death,[2] and its prevalence and societal costs.[3–5] Screening and management of SU merit a position in medical curricula that reflects the importance as a mainstream medical problem.[6–10] Many physicians fail to address SU disorders because of discomfort with SU-related patient discussions,[11] deficient knowledge and clinical skills,[12,13] and negative attitudes,[14,15] all resulting in barriers to providing optimal medical care for this population.[16]

Medical educators are starting to address this need for physician training in SU screening, assessment, and management.[17–22] Formal curricula on these subjects have been developed[23,24] and evaluated,[25,26] and recommendations for the medical care of addicted patients have been published.[6,27,28] Nevertheless, dissemination of up-to-date addiction research into generalist practice and into residency curricula remains a significant challenge.[16,29,30]

Substance use education aimed at improving physician trainees' attitudes and clinical practice has been effective.[31,32] Confidence in ability to screen and refer patients is positively associated with perceived responsibility and clinical practice.[33] Wider dissemination of these practices requires creative strategies to develop a workforce that is knowledgeable about state-of-the-art approaches to patient management and motivated to implement such practices in a range of settings.[1,26,34,35] As noted in the Institute of Medicine Report *Improving the Quality of Health Care for Mental and Substance-Use Conditions*,[16] medical educators have not adequately addressed past recom-

mendations to update training of medical professionals, leaving trainees ill equipped in their ability to care for patients with SU disorders.

Chief residents (CRs) play a key role in training future physicians.[36] Not only is their teaching central to medical education,[37,38] but CRs often become change agents in future leadership roles.[39] Despite these pivotal roles, published efforts to advance medical training by capitalizing on the CRs role are rare. A training effort in SU that has influential medical educators and leaders, such as CRs, as its target audience has enormous potential to integrate SU clinical training into medical curricula and practice.[36,40]

The Chief Resident Immersion Training (CRIT) program sought to provide incoming generalist CRs with the scientific foundation of addiction medicine and state-of-the-art substance use (SU) diagnosis and management skills, in order to facilitate integration of SU content into residency program curricula and CR teaching.

METHODS

Study Design and Subjects

This study is a controlled educational evaluation of applicants to the 4-day annual CRIT program in Addiction Medicine from 2003–2005. The programs were advertised at http://www.bumc.bu.edu/CARE and by mailings to generalist residency training program directors and professional organizations. All applicants had accepted CR positions for the following academic year in internal medicine, family medicine, or emergency medicine, with priority given to outstanding candidates who (1) demonstrated the presence of a supportive faculty member to assist in promoting the teaching of substance use (SU) and (2) would have dedicated time for teaching during their CR year. Due to resource limitations, enrollment was limited to no more than 22 attendees per year based on their application, which included a curriculum vitae, personal statement, and letter of recommendation. The intervention group consisted of CRIT program attendees, while the control group consisted of applicants who were either not accepted or accepted but unable to attend. This study was approved by the Boston Medical Center Institutional Review Board.

CRIT Program Description

The CRIT program curriculum was developed by national experts in physician SU education. After the first year in 2002, the curriculum was modified based on attendees' evaluations and an independent evaluator-led attendee focus group. Subsequent program evaluations gave feedback to faculty, resulting in minor teaching and programmatic modifications. The CRIT program was held at a conference center in Cape Cod, Massachusetts.

Curriculum and Faculty. The CRIT curriculum (see Appendix A) utilized principles of adult and experiential learning and provided an in-depth evidence-based synthesis of major advances in the field of addiction medicine. The curriculum included a keynote address on the science of addiction, by a National Institute on Drug Abuse senior scientist. CRIT faculty, all with addiction and medical education expertise, role modeled teaching methods, including didactic case-based presentations, small group discussions, journal clubs, role plays, visits to Alcoholics Anonymous (AA) meetings, and small group conversations with individuals in recovery. Both the AA meeting visits and session with individuals in recovery were preceded by an orientation describing the learning objectives and followed by a debriefing session emphasizing the use of these curricular elements in teaching physician trainees. The use of experiential learning and specifically AA meeting visits[41,42] are curricular elements that have been used successfully in physician training. Attendees received SU resources and teaching tools, including slide presentations, case studies for small group work, instructions and role descriptions for skill practice exercises, up-to-date medical literature, and important SU websites. The program also included instruction on teaching skills, including small group instruction, giving feedback, teaching reluctant learners, and ways to integrate SU issues into teaching.

Substance Use Teaching Project. The CRIT curriculum included the development of a SU teaching project to enhance transfer of learning from the CRIT setting to the CRs' work setting. Prior to arrival, CR attendees were asked to discuss possible SU teaching projects with their residency program director and a faculty mentor, giving consideration to residency program needs and CRs' interests. The CRs explicitly stated project goals, objectives, implementation methods, potential resources and barriers, a timeline, and outcomes using a worksheet (see Appendix B). CRs met with CRIT faculty to further develop an achievable project. To encourage institutional support, a copy of the teaching project was sent to the CR's mentor and residency program director.

Assessments

The outcome evaluation comparing the intervention and control CRs consisted of self-assessment questionnaires completed at the time of the CRIT application (baseline) and 6-months after the CRIT program (follow-up). The majority of questions used five-point Likert-type scales based on work by D'Onofrio et al.[43] CRs were asked to rate their knowledge, skills, and confidence ("not at all" to "very") and specific clinical and teaching practices ("never" to "always"). In addition, the intervention CRs completed a pre- and post-CRIT multiple-choice knowledge exam at the training based on didactic session content and an 11-month questionnaire about SU teaching project implementation and impact. The baseline, 6-, and 11-month follow-up questionnaires were administered via e-mail, and the pre- and post-CRIT knowledge exams were completed at the CRIT course. All were returned to the CRIT independent evaluator. Intervention CRs received a gift worth approximately $10, and control CRs received a $50 honorarium after completing the follow-up questionnaires. CRs used unique IDs for assessments to assure confidentiality.

Outcomes

Substance Use Clinical Knowledge, Skills, and Practice. Multiple outcomes measured CRs knowledge, confidence, and self-reported SU clinical practices. The baseline and 6-month questionnaires assessed self-reported clinical knowledge and

skills (SU neurobiology, screening, readiness to change assessment, referral options, pharmacotherapy, and relapse) and self-reported clinical confidence and practice (confidence in diagnosing SU problems and frequency of using screening tools with new patients, counseling drug and alcohol using patients, and referring drug- and alcohol-dependent patients for specialty treatment). A multiple-choice knowledge exam was administered to attendees immediately before and after the program.

Substance Use Teaching Skills and Practice. Teaching outcomes included self-reported responsibility for, confidence in, and frequency of teaching about SU. Additionally, CRs were asked how prepared they felt to teach about SU compared to other chronic medical conditions (i.e., congestive heart failure and dementia). For the intervention CRs' SU teaching project, outcomes included type of teaching activity, impact on residency program curriculum, and implementation facilitators and barriers.

Statistical Analysis

Baseline comparison of the intervention and control groups was performed using chi-square and t-tests, as appropriate, for analysis of the demographic variables (Table 1). The differences in self-reported knowledge, skills, and practice from baseline to 6-month follow-up were compared within each group using the paired t-test procedure (Table 2). Differences between group medians for outcomes at 6 months were examined using the Wilcoxon signed rank test. A two-sided p-value of 0.05 was considered statistically significant. Mean scores of the intervention group multiple-choice knowledge exams for attendees were compared.

Table 1 . Baseline characteristics of applicants (n=86) to the Chief Resident Immersion Training (CRIT) program in addiction medicine 2003–2005, stratified by group

Characteristic	Intervention n=64 (%)	Control n=22 (%)	p-value
Gender			
Female	30 (47)	11 (50)	0.81
Race			0.58
White	40 (63)	15 (68)	
Black	6 (9)	2 (9)	
Asian	14 (22)	2 (9)	
Hispanic	2 (3)	1 (5)	
Other	2 (3)	2 (9)	
Age in years, mean [SD]	30.3 [2.4]	30.8 [4.4]	0.61
Specialty			0.60
Emergency medicine	4 (6)	1 (5)	
Family medicine	4 (6)	2 (9)	
Internal medicine	56 (88)	19 (86)	
Fellowship plans	38 (59)	12 (55)	0.69
Foreign medical graduate	16 (25)	6 (27)	0.83
Alpha Omega Alpha member	5 (8)	2 (9)	0.85
Mentor support named in application	48 (75)	12 (55)	0.19
Substance use (SU) usually/always a topic in residency program	32 (50)	10 (45)	0.71
Training in SU treatment is encouraged in residency program	50 (78)	15 (68)	0.35
Has informed, competent faculty member source of SU information	54 (86)*	21 (100)†	0.07

*(n=63)
†(n=21)

RESULTS

Eighty-six chief resident (CR) applicants to the CRIT program (2003–2005) were assessed; 64 attended CRIT (intervention group), and 22 did not (control group). Of the control group, six had been accepted to the program. The baseline characteristics of the intervention and control groups were similar (Table 1). Applicants represented 56 different residency programs (87% internal medicine, 7% family medicine, and 6% emergency medicine) from 23 states. All intervention and control CRs completed the baseline assessment; 6-month follow-up was 100% (64/64) and 86% (19/22) for the intervention and control groups, respectively. All of the intervention CRs completed the pre- and post-CRIT multiple-choice knowledge exams. All intervention CRs developed a SU teaching project, and 98% (63/64) completed the 11-month follow-up questionnaire.

Substance Use Clinical Knowledge, Skills, and Practice (Table 2)

The intervention CRs showed significant improvement ($p<0.001$) at 6-month follow-up compared to baseline assessment in self-reported knowledge on SU neurobiology, screening, readiness to change assessment, referral options, pharmacotherapy, and relapse. Similarly, they had increased confidence ($p<0.05$) in diagnosing SU problems and more frequently used a SU screening tool with new patients, counseled drug and alcohol using patients, and referred drug-dependent patients to treatment. The control CRs improved in some outcomes from baseline to follow-up. When comparing the intervention group and the control group medians at 6-month follow-up, the intervention group displayed statistically significant ($p<0.05$) improvement in self-reported knowledge of SU neurobiology, screening, referral options, pharmacotherapy, and relapse, as well as confidence in diagnosing SU problems and frequency of using a SU screening tool with new patients. The mean pre- and post-CRIT multiple-choice knowledge exam scores for the intervention CRs were 67% and 78% answers correct, respectively ($p<0.001$).

Substance Use Teaching Skills and Practice (Table 2)

The intervention CRs showed significant improvement ($p<0.001$) at 6-month follow-up compared to baseline in confidence in incorporating SU into teaching, making presentations on SU issues, and incorporating SU information into residents' curriculum. The frequency of covering alcohol and drug abuse in teaching similarly improved. Differences between the intervention group and the control group medians at 6-month follow-up were also significant ($p<0.05$) for those outcomes, with the exception of frequency of covering drug abuse. At baseline, both groups reported high levels of feeling responsible for teaching about SU. While the intervention CRs had no

Table 2 . Chief residents' clinical knowledge, skills and practice and teaching skills and practice related to substance use (SU), at baseline and 6-month follow-up

	Intervention mean (SD)		Control mean (SD)		Difference between group follow-up medians[b] (range)
Characteristic (on a scale of 1–5)	Baseline n=64	Follow-up[a] n=64	Baseline n=22	Follow-up[a] n=19	
Clinical knowledge, skills, and practice					
How knowledgeable are you about:					
Neurobiology of addiction?	2.2 (0.8)	3.3 (0.7)***	2.0 (0.6)	2.4 (0.7)*	1 (−2,4)*
Screening for substance use?	3.1 (0.8) †	4.2 (0.7)***	2.9 (0.8)	3.2 (0.9)	1 (−2,3)**
Readiness to change assessment?	2.9 (1.0) †	4.2 (0.9)***	2.5 (0.7)	3.4 (0.9)**	1 (−2,4)
Available referral options?	2.7 (0.9)	3.8 (0.8)***	2.6 (0.7)	3.1 (0.8)	1 (−2,4)*
Pharmacotherapy for addiction?	2.9 (1.0)	4.0 (0.8)***	2.7 (0.7)	3.0 (0.7)	1 (−2,3)**
Relapse?	2.1 (0.8)	3.4 (0.9)***	2.0 (0.7)	2.5 (0.9)*	1 (−1,4)*
How confident are you in diagnosing SU problems?	3.1 (0.8)	4.0 (0.8)***	3.3 (1.0)	3.4 (0.8)	1 (−2,3)**
How often do you:					
Use a SU screening tool with new patients?	3.3 (0.9)	4.0 (0.9)***	3.0 (1.1)	3.1 (0.9)	1 (−1,3)**
Counsel drug abusing patients about drug problems?	3.9 (0.8)	4.2 (0.7)**	3.7 (0.8)	4.0 (0.7)	0 (−2,3)
Counsel drinkers about alcohol problems?	3.9 (0.7)	4.2 (0.7)**	3.9 (0.7)	4.0 (0.7)	0 (−2,2)
Refer drug-dependent patients to treatment?	3.8 (0.9)	4.0 (0.9)*	3.6 (0.8)	3.8 (0.9)	0 (−2,2)
Refer alcohol-dependent patients to treatment?	3.8 (0.8)	4.1 (0.8)	3.7 (0.9)	3.9 (0.8)	0 (−2,2)
Teaching skills and practice					
How responsible do you feel for teaching about SU?	4.2 (0.9)	4.4 (0.7)	4.3 (0.9)	3.6 (1.0)*	0 (−4,4)*
How confident are you in:					
Incorporating SU into your teaching?	2.9 (1.0)	4.0 (0.8)***	3.1 (1.2)	3.1 (0.9)	1 (−2,4)**
Making presentations on SU issues?	2.7 (1.0)	4.0 (0.8)***	2.7 (1.3)	2.6 (0.8)	1 (−3,4)**
Incorporating SU information into resident's curriculum?	3.1 (1.1)	4.1 (0.8)***	3.1 (1.2)	3.0 (1.0)	1 (−2,4)**
How often do you:					
Cover alcohol abuse?	2.5 (0.8)	3.2 (0.7)***	2.1 (0.6)	2.3 (0.9)	0 (−3,3)*
Cover drug abuse?	2.4 (0.8)	3.0 (0.7)***	2.0 (0.6)	2.3 (1.0)	0 (−3,1)

*p<0.05, **p<0.01, ***p<0.001 †n=60

[a]*The test of significance compared the baseline mean to the follow-up mean within each group*

[b]*The test of significance compared the intervention group follow-up median to the control group follow-up median*

change in their reported levels of feeling responsible to teach about SU at follow-up, the control CRs had a significant decrease (p<0.05). Almost all (97%) of intervention CRs reported being "more" or "much more" likely to incorporate SU content into their teaching as a result of their CRIT training. When reporting their preparedness to teach about SU and other chronic conditions (i.e., congestive heart failure, dementia), the intervention group showed significant improvement (p<0.05) at follow-up in their preparedness to teach about alcohol abuse as compared to the control group (Fig. 1).

Substance Use Teaching Project

All intervention CRs developed a SU teaching project. At 11-month follow-up, 86% (56/64) reported that their teaching project had an impact on their residency program curriculum. Examples of the SU teaching projects included new or expanded SU curriculum (n=21), new SU lecture or teaching activity (n=33), and SU screening protocols (n=16). Two CRs published an assessment of the impact of their SU teaching project, one included taking residents and medical students to an Alcoholics Anonymous meeting [42] and the other the successful implementation of an "Educational Half Day" focusing on SU.[44] CRs identified cooperative residency program staff and support from a mentor as the top two facilitators in helping to achieve their SU teaching project, while the top two barriers were time constraints and competing priorities.

DISCUSSION

The Chief Resident Immersion Training (CRIT) program in addiction medicine effectively transferred evidence-based SU knowledge and practice to 64 CRs in generalist disciplines and more importantly, enhanced the SU curriculum in 47 residency programs. Training CRs, who have a primary responsibility for educating medical trainees, appears to be one important pragmatic strategy to address the compelling need for better physician training in the identification and management of patients with SU-related problems.

Educating trainees about SU poses extraordinary challenges. Until a few decades ago, there was little research to guide the practice of addiction medicine by generalists. Yet in 1990, the Institute of Medicine stated that most patients with SU issues should receive care in a primary care setting.[45] A decade later, Fiellin and colleagues described the basic core competencies in SU education for physicians and recommended increased efforts to provide physicians with a level of SU training commensurate with the economic and clinical

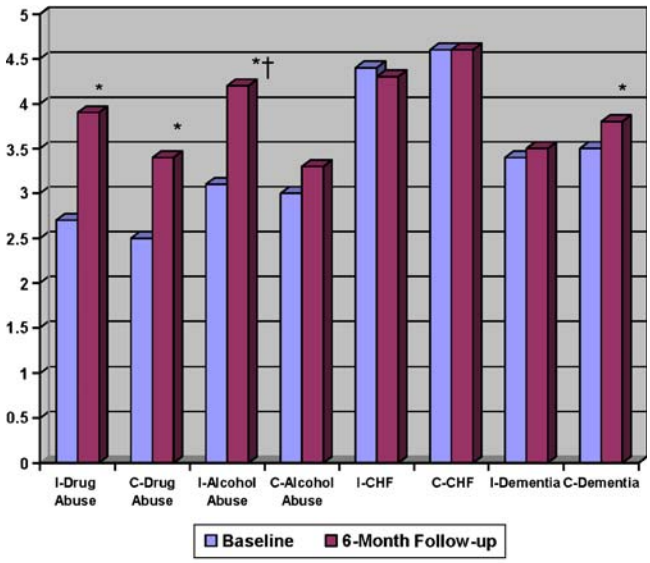


Figure 1. Changes in chief resident's preparedness to teach substance use compared to other chronic conditions at baseline and 6-month follow-up by group {n=86 (64 intervention (I), 22 control (C)) at baseline, and n=83 (64 intervention (I), 19 control (C)) at 6-month follow-up}. *p<0.05, comparing baseline to follow-up within group; †p<0.05, comparing intervention to control group at 6-month follow-up. I = Intervention, C = Control.

impact of SU disorders.[1] In 2008, medical education has yet to provide adequate training in SU, resulting in a dearth of generalist faculty with an interest and expertise in SU to serve as teachers, role models, or mentors. Recently, the National Institute on Drug Abuse has joined forces with the American Medical Association to improve SU training in medical education.[21]

Renner described three critical elements to successfully train physicians in addiction medicine: (1) an adequate knowledge base, (2) a positive attitude toward the patient and the benefits of treatment, and (3) a sense of responsibility for the clinical problem.[46] The CRIT program exposed CRs to up-to-date SU knowledge and evidence-based practices. The immersion experience allowed CRs to focus on the CRIT curriculum without the inevitable distractions of their jobs at home. The faculty modeled different teaching methodologies with an emphasis on interactive learning. Meeting with individuals in recovery demonstrated the reality of treatment, its successes, and challenges. In addition, the physician faculty, all generalists with special expertise in addiction and medical education, served as role models of physicians who had successfully integrated academic, clinical, and research careers in addictions and medical education. The SU teaching project helped foster the transfer of SU knowledge and skills to the CRs' home institutions. CRs were asked to identify a mentor at their institution as this support system was viewed as key to successful local implementation of teaching projects. The finding that intervention CRs had greater increases in self-perceived preparedness to teach about alcohol abuse compared to no changes in teaching about another chronic medical condition, such as dementia, which remained persistently low, highlights the fact that such improvements do not occur spontaneously.

Targeting CRs was a tactical decision in order to magnify the impact of the immersion program. Other successful interventions have increased the knowledge and skills of generalist residents,[26,32] but have not demonstrated the potential to shift the values and culture of a residency program to include more SU training. In addition to being pivotal educators in the clinical training of students and residents, CRs have additional strengths. They are chosen for the high quality of their medical knowledge, clinical skills, and organizational abilities. As role models, CRs have the opportunity to impact the knowledge, skills, and attitudes of trainees. Surprisingly, there have been relatively few efforts to advance medical training by capitalizing on the role of CRs as influential instructors, leaders, and role-models. One example is a CRIT program for the care of older adults to teach geriatric medicine skills to a single institution's chief residents.[47] Engaging these key young medical leaders early in their careers when they are open to new ideas is a strategy with enormous educational potential.

The CRIT program evaluation has several limitations worth considering. The control group was smaller than the intervention group, and neither group was randomly assigned. If the control group had been larger, it is possible that some of the outcomes that were significant in the intervention group would have also been significant in the control group; however, the absolute change between baseline and follow-up was almost always smaller in the control group. The non-randomized nature of the study could have led to confounding. For example, because enrollment in the intervention group was based on selection by the CRIT program directors and 73% of the control group was not accepted, intervention CRs might have been more qualified than the controls. However, both groups were similar in the baseline variables measured. Another limitation is the overall small sample size, which makes it difficult to identify differences between groups. Nevertheless, many results did reach statistical significance. Next, it is possible that due to the self-reported nature of the data, some of the findings may be attributed to social desirability bias. Social desirability may have been more of an issue for the intervention group. To mitigate this bias, CRs returned their questionnaires to an independent evaluator, and were told that faculty and staff would only see de-identified aggregate data. While we found improvements in clinical knowledge, skills, and practice among the CR attendees, we were unable, by study design, to detect if these improvements continued downstream to their trainees. However, in two cases we are aware of CRs conducting their own assessment of impact of their teaching project.[42,44] Future research on the CRIT model should consider a more in-depth investigation of impact on the CRs' trainees. Finally, the conclusions of this study may not be generalizable to all generalist CRs, as the CRIT program enrolled CRs who self-selected for interest in this training and had a disproportionately high representation from internal medicine.

In summary, immersion programs directed at CRs incorporating a variety of teaching modalities and the explicit development of a teaching project can help transmit substance use knowledge and teaching expertise to medical trainees. Chief Resident Immersion Training (CRIT) impacted physicians who play a critical role in medical trainee education but are not often trained in substance use themselves. CRs are an untapped resource for changing medical education and practice about substance use disorders, and this CRIT model holds

potential to address other high-priority medical content areas as well.

Acknowledgements: *Preliminary study results were presented at the 27th annual meeting of the Society of General Internal Medicine in Chicago, IL, in May 2004, the 28th annual meeting of the Association for Medical Education and Research in Substance Abuse in Washington, DC, in November 2004 and the 67th annual meeting of the College on Problems of Drug Dependence in Orlando, FL, in June 2005. The authors thank the following for their contributions to the manuscript: Jacqueline Ashba, MA, MPH, and Don Allensworth-Davies, MSc, at the Data Coordinating Center, Boston University School of Public Health, Boston, MA (data management); Thomas W. Clark, formerly at Health and Addictions Research, Inc. (evaluation); Wendy Budwey, MSW (program management); the CARE Program Advisory Committee (program oversight); Nella P. Mupier, MPH (manuscript preparation); and Belle Brett, EdD, for thoughtful comments on this manuscript. Program support was provided by the National Institute on Drug Abuse R25 DA13582. Dr. Samet receives support from NIAAA: K24-AA015674.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *Daniel P. Alford, MD, MPH; Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston University School of Medicine, 801 Massachusetts Avenue, 2nd floor, Boston, MA 02118-2393, USA (e-mail: dan.alford@bmc.org).*

REFERENCES

1. **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 Abus. 2002;23:207–12.
2. **Michaud CM, Murray CJ, Bloom BR.** Burden of disease—implications for future research. JAMA. 2001;285:535–9.
3. NIAAA. Estimated economic costs of alcohol abuse in the United States, 1992 and 1998. Available at: http://www.niaaa.nih.gov/Resources/DatabaseResources/QuickFacts/EconomicData/cost8.htm. Accessed August 25, 2008.
4. Office of National Drug Control Policy. The economic cost of drug abuse in the United States, 1992–2000. Available at: http://www.whitehousedrugpolicy.gov/publications/economic_costs/. Accessed August 25, 2008.
5. SAMHSA. Results from the 2006 national survey of drug use and health: national findings. Available at: http://www.drugabusestatistics.samhsa.gov/NSDUH/2k6NSDUH/2k6results.cfm. Accessed August 25, 2008.
6. **Friedmann PD, Saitz R, Samet JH.** Management of adults recovering from alcohol or other drug problems: relapse prevention in primary care. JAMA. 1998;279:1227–31.
7. **Lewis DC.** The role of internal medicine in addiction medicine. J Addict Dis. 1996;15:1–17.
8. **McLellan AT, Lewis DC, O'Brien CP, Kleber HD.** Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. JAMA. 2000;284:1689–95.
9. **Saitz R, Horton NJ, Larson MJ, Winter M, Samet JH.** Primary medical care and reductions in addiction severity: a prospective cohort study. Addiction. 2005;100:70–8.
10. **Samet JH, Friedmann P, Saitz R.** Benefits of linking primary medical care and substance abuse services: patient, provider, and societal perspectives. Arch Intern Med. 2001;161:85–91.
11. **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.
12. **Kuehn BM.** Despite benefit, physicians slow to offer brief advice on harmful alcohol use. JAMA. 2008;299:751–3.
13. **Vastag B.** Addiction poorly understood by clinicians: experts say attitudes, lack of knowledge hinder treatment. JAMA. 2003;290:1299–303.
14. **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;17:373–6.
15. **Stimmel B, Cohen D, Colliver J, Swartz M.** An Assessment of house staff's knowledge of alcohol and substance abuse utilizing standardized patients. Subst Abus. 2000;21:1–7.
16. Institute of Medicine. Improving the quality of health care for mental and substance-use conditions. Washington, DC: The National Academies Press; 2006.
17. **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;69:362–9.
18. **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;76:410–8.
19. **Foley ME, Garland E, Stimmel B, Merino R.** Innovative clinical addiction research training track in preventive medicine. Subst Abus. 2000;21:111–9.
20. **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;48:725–9.
21. **Kuehn BM.** Centers to weave addiction treatment into medical education. JAMA. 2007;297:1763.
22. **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.
23. Project Mainstream: improving substance abuse education for health professionals. Available at: http://www.projectmainstream.net. Accessed March 11, 2008.
24. **Klamen DL, Miller NS.** Integration in education for addiction medicine. J Psychoactive Drugs. 1997;29:263–8.
25. **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 Med Educ. 2008;8:11.
26. **Parish SJ, Ramaswamy M, Stein MR, Kachur EK, Arnsten JH.** Teaching about substance abuse with objective structured clinical exams. J Gen Intern Med. 2006;21:453–9.
27. **Samet JH, Rollnick S, Barnes H.** Beyond CAGE. A brief clinical approach after detection of substance abuse. Arch Intern Med. 1996;156:2287–93.
28. **Alford DP, Compton P, Samet JH.** Acute pain management for patients receiving maintenance methadone or buprenorphine therapy. Ann Intern Med. 2006;144:127–34.
29. **Isaacson JH, Fleming M, Kraus M, Kahn R, Mundt M.** A national survey of training in substance use disorders in residency programs. J Stud Alcohol. 2000;61:912–5.
30. **Abrams-Weintraub T, Saitz R, Samet JH.** Education of preventive medicine residents: alcohol, tobacco, and other drug abuse. Am J Prev Med. 2003;24:101–5.
31. **Karam-Hage M, Nerenberg L, Brower KJ.** Modifying residents' professional attitudes about substance abuse treatment and training. Am J Addict. 2001;10:40–7.
32. **Chossis I, Lane C, Gache P, et al.** Effect of training on primary care residents' performance in brief alcohol intervention: a randomized controlled trial. J Gen Intern Med. 2007;22:1144–9.
33. **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. 1989;261:3115–20.
34. **Marcus MT, Brown RL, Straussner SL, et al.** Creating change agents: a national substance abuse education project. Subst Abus. 2005;26:5–15.
35. **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;20:655–64.
36. **Doughty RA, Williams PD, Seashore CN.** Chief resident training. Developing leadership skills for future medical leaders. Am J Dis Child. 1991;145:639–42.
37. **Levine SA, Caruso LB, Vanderschmidt H, Silliman RA, Barry PP.** Faculty development in geriatrics for clinician educators: a unique model for skills acquisition and academic achievement. J Am Geriatr Soc. 2005;53:516–21.

38. **Steel K.** The medical chief residency in university hospitals. Ann Intern Med. 1972;76:541–4.
39. **Alpert JJ, Levenson SM, Osman CJ, James S.** Does being a chief resident predict leadership in pediatric careers? Pediatrics. 2000;105:984–8.
40. **Mygdal WK, Monteiro M, Hitchcock M, Featherston W, Conard S.** Outcomes of the first family practice chief resident leadership conference. Fam Med. 1991;23:308–10.
41. **Bigby J, Barnes HN.** Evaluation of a faculty development program in substance abuse education. J Gen Intern Med. 1993;8:301–5.
42. **Rose AJ, Stein MR, Arnsten JH, Saitz R.** Teaching internal medicine resident physicians about alcoholics anonymous: a pilot study of an educational intervention. Subst Abus. 2006;27:5–11.
43. **D'Onofrio G, Nadel ES, Degutis LC, et al.** Improving emergency medicine residents' approach to patients with alcohol problems: a controlled educational trial. Ann Emerg Med. 2002;40:50–62.
44. **Nigwekar SU, Morse DS.** Educational half day: an innovative way to incorporate substance abuse curriculum into residency training. Subst Abus. 2006;27:1–3.
45. Institute of Medicine. Broadening the base of treatment for alcohol problems. Washington DC: The National Academies Press; 1990.
46. **Renner JA Jr.** How to train residents to identify and treat dual diagnosis patients. Biol Psychiatry. 2004;56:810–6.
47. **Levine S, Chao S, Brett B, et al.** Chief resident immersion training in the care of older adults: an innovative inter-specialty education and leadership intervention. J Am Geriatr Soc. 2008;56:1140–5.

APPENDIX A: EXAMPLE SCHEDULE

Clinical teaching in addiction medicine. A Chief Resident Immersion Training (CRIT) program

Day 1		
12:00–3:30	Arrival and hotel check-in	
3:30–5:00	Pre-CRIT multiple-choice knowledge exam	
5:00–6:30	Reception and dinner	
	Welcome, opening remarks	
6:15–7:00	Ice breaker	
	"What was your most memorable adult learning experience? Why?"	
	Review of adult learning principles	
7:00–7:15	Introduction to substance use teaching project	
7:15–8:15	Teaching project discussion: What is a doable chief resident project?	
Day 2		
7:40–8:45	Breakfast	
8:45–9:00	Introductions	
9:00–10:00	NIDA keynote address: The science of addiction	Dr. Condon/ Miner (NIDA)
10:15–10:35	Principles of addiction	Dr. Samet
10:35–11:05	Screening	Dr. Barnes
11:05–12:30	Assessment and brief intervention	Dr. Samet
12:30–3:00	Lunch and break	
3:00–3:30	Alcohol: Inpatient management	Dr. Saitz
3:30–4:00	Opioids: Inpatient management	Dr. Alford
4:00–4:30	Running small groups	Dr. Jackson
5:00–6:30	Skills practice session I (three groups)	All faculty
7:00–8:30	Dinner/social	
Day 3		
7:40–9:00	Breakfast (one on one: teaching project discussion)	All faculty
9:00–9:30	Alcohol: Outpatient management	Dr. Saitz
9:30–10:00	Opioids: Outpatient management	Dr. Alford
10:00–10:30	Prescription drug abuse	Dr. Alford
11:00–12:30	Skills practice session II (three groups)	All faculty
12:30–3:00	Lunch and break	
3:00–4:00	Giving feedback/reluctant learner	Dr. Jackson
4:10–5:10	Five concurrent workshops	
	HIV and substance use	Dr. Samet
	Clinical epidemiology rounds	Dr. Saitz
	Marijuana: Myths and realities	Dr. Barnes
	Tool kit for teachers	Dr. Jackson
	SU psychosocial services	Dr Amodeo
5:30–6:30	Dinner	
6:00–6:30	Orientation: 12-Step meetings	Dr. Amodeo
6:40–8:30	12-Step meetings	
6:30–7:30	One on one: Teaching project discussion	All faculty
Day 4		
6:15–8:30	12-Step meetings	
8:30–9:30	Breakfast (one on one: Teaching project discussion)	All faculty
9:40–10:00	Debrief: 12-Step meetings	Dr. Amodeo
10:00–10:30	Stimulants and sedatives	Dr. Alford
10:50–11:50	Skills practice session III (three groups)	All faculty
11:50–12:15	Orientation: Luncheon with guests in recovery	Dr. Amodeo
12:15–1:15	Luncheon with guests in recovery	
1:30–2:00	Debrief: Luncheon with guests in recovery	Dr. Amodeo
2:00–2:35	Medical complications of substance use	Dr. Samet
2:55–3:55	Four selected teaching project presentations	Four participants
3:55–4:20	Questions and answers	Faculty
4:20–5:00	Incorporating substance use into curriculum	Dr. Barnes
5:00–5:30	Post-CRIT multiple-choice knowledge exam	
5:45–7:00	Dinner: Certificates of completion	
7:00	Adjourn	

APPENDIX B: SU TEACHING PROJECT WORKSHEET

CRIT Program Substance Use Teaching Project Plan

Name_____Institution_____Date_____
Project Title _____
Goal_____

Objectives:	
Target Audience and Setting: • Who? • Where? • When?	
Resources: • Who, what is available to assist with your goal and objectives?	
Barriers: • Who, what may impede your goal and objectives?	
Time Period: • Target date for completion of goal and objectives?	
Action Steps: • What steps will enable you to achieve each objective?	

Addictive Behaviors 34 (2009) 636–640

Contents lists available at ScienceDirect

Addictive Behaviors

Does readiness to change predict subsequent alcohol consumption in medical inpatients with unhealthy alcohol use?

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

[a] Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, United States
[b] Clinical Epidemiology Center, Institute of Social and Preventive Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland
[c] Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States
[d] Department of Psychology, Boston University, Boston, MA, United States
[e] Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, United States
[f] Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, United States
[g] Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

ARTICLE INFO

Keywords:
Unhealthy alcohol use
Readiness to change
Medical inpatients
Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES)

ABSTRACT

We studied whether readiness to change predicts alcohol consumption (drinks per day) 3 months later in 267 medical inpatients with unhealthy alcohol use. We used 3 readiness to change measures: a 1 to 10 visual analog scale (VAS) and two factors of the Stages of Change Readiness and Treatment Eagerness Scale: Perception of Problems (PP) and Taking Action (TA). Subjects with the highest level of VAS-measured readiness consumed significantly fewer drinks 3 months later [Incidence rate ratio (IRR) and 95% confidence interval (CI): 0.57 (0.36, 0.91) highest vs. lowest tertile]. Greater PP was associated with more drinking [IRR (95%CI): 1.94 (1.02, 3.68) third vs. lowest quartile]. Greater TA scores were associated with less drinking [IRR (95%CI): 0.42 (0.23, 0.78) highest vs. lowest quartile]. Perception of Problems' association with more drinking may reflect severity rather than an aspect of readiness associated with ability to change; high levels of Taking Action appear to predict less drinking. Although assessing readiness to change may have clinical utility, assessing the patient's planned actions may have more predictive value for future improvement in alcohol consumption.

1. Introduction

Consequences resulting from excessive alcohol consumption are responsible for considerable disease burden (Rehm et al., 2003). Consequently, interventions aimed at reducing excessive alcohol consumption are viewed as beneficial from a public health point of view. In 2004, the US Services Preventive Task Force recommended the use of brief counseling intervention in primary care, where its efficacy has been confirmed; this practice is among the most effective and cost-effective of preventive care services (Bertholet, Daeppen, Wietlisbach, Fleming, & Burnand, 2005; Solberg, Maciosek, & Edwards, 2008). Assessment of patients' readiness-to change provides a self-report index of patient motivation to alter their drinking patterns that may be used to tailor advice and counseling to patients and determine treatment dispositions. Indeed, health care providers have been encouraged to see increases in readiness-to-change as a desirable intermediate goal on the path to behavior change (Samet, Rollnick, & Barnes, 1996).

The focus on readiness is based in large part on the assumption that there is a clear association between readiness-to-change and outcome (e.g., decreases in drinking). However, there is conflicting evidence regarding the relationship between readiness to change and outcome (Forsberg, Ekman, Halldin, & Ronnberg, 2004; Isenhart, 1997; Reed et al., 2005; Rollnick, 1998). These equivocal findings may have a number of explanations. First, a family of concepts is included under the term "readiness." These concepts may include importance of change, problem recognition, confidence, and actions reflecting a commitment to change. These concepts, especially importance of change (sometimes related to or understood as problem or consequence recognition) or confidence in ability to change (also known as self-efficacy), could operate differently. In particular, confidence seems to predict better outcomes (Maisto, Conigliaro et al., 1999). Initiation of behavior change appears to be associated with the expectancy to cope successfully (Demmel, Beck, Richter, & Reker, 2004). However, other indices of motivation to change, such as Problem recognition, do not appear to be associated with better outcomes. Miller and Tonigan developed a questionnaire (the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES)) aimed at capturing stages of change as described by Prochaska and DiClemente. After examining the psychometric properties of the questionnaire, it appeared that items about negation of the problem (precontemplation) and recognition of the problem (determination) formed a single factor.

* Corresponding author. Alcohol Treatment Center, Lausanne University Hospital, 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).

0306-4603/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.addbeh.2009.03.034

This factor was named "Recognition" (Miller & Tonigan, 1996). Problem recognition is linked to one's capacity to attribute the cause of a health or social problem to alcohol and to the existence of such a problem. As such, problem recognition is considered to be an important component of readiness-to-change drinking (Nye, Agostinelli, & Smith, 1999). However, in addition to suggesting greater awareness of problematic alcohol use patterns, problem recognition may also indicate higher levels of alcohol consequences and thus serve as a marker of alcohol use disorder severity (Maisto, Conigliaro et al., 1999; Williams, Horton, Samet, & Saitz, 2007).

The differing associations between readiness measures and outcomes may also be explained by patient populations and by the assessment instrument. For example, the development of the SOCRATES questionnaire was designed to assess readiness among a treatment seeking population. The structure and validity of these measures may be quite different among non-treatment seeking problem drinking samples (Maisto, Conigliaro et al., 1999).The role of readiness to change is of particular interest in patients with unhealthy alcohol use (i.e., the spectrum from risky consumption to alcohol dependence) identified by screening in general health settings, such as hospitals. In this circumstance, patients are not necessarily help-seeking, unlike patients in specialty treatment. Such a population is less homogeneous than a treatment seeking population, which may be responsible for differences in the potential associations between readiness to change and drinking.

Therefore we studied whether readiness to change predicts subsequent alcohol consumption in medical inpatients with unhealthy alcohol use. We studied this association using three different measures of readiness to change based on two instruments—a visual analog scale (VAS) for the simple question "how ready are you to change your drinking habits?" and two factors from the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES), level of perception of the drinking problem and taking action towards change/commitment to making a change. The use of two different instruments covering a more global readiness concept and more specific constructs, is of interest since we expect to capture various aspects of readiness. There is currently no gold standard in readiness to change measurement and so the use of two different instruments will also give additional information on their respective predictive values. We hypothesized that a high level of taking action towards change/commitment to change would be associated with less drinking, and that greater problem perception, reflecting severity, would be associated with more drinking, based on what has been observed in a primary care population (Williams et al., 2007).

2. Methods

We studied a prospective cohort of medical inpatients at an urban academic hospital who were drinking risky amounts (>14 drinks/wk or ≥5 drinks/occasion for men, >11 drinks/wk or ≥4 drinks per occasion for women and persons aged 66 and over). The general medical inpatient service we studied was internal medicine and it did not include intensive care unit beds. Subjects were participants in a randomized trial of brief intervention to reduce alcohol use (Saitz et al., 2007) and were recruited from the inpatient medical service of an urban teaching hospital. Research associates approached all patients aged 18 or older whose physicians did not decline patient contact. Individuals fluent in English or Spanish who gave consent were asked to complete a screening interview. Eligibility criteria included the following: currently drinking risky amounts, 2 contacts to assist with follow-up, no plans to move from the area for the next year, and a Mini-Mental State Examination score of ≥21 (Smith, Horton, Saitz, & Samet, 2006). Eligible subjects were randomized to receive usual care or a brief intervention to reduce alcohol use. The study population was used as a cohort in the present analyses. Assessments took place before group allocation.

2.1. Assessments

Demographics were assessed at study entry, as well as medical diagnoses by medical record review to identify those that were alcohol related (see Table 1), and alcohol use disorder diagnosis based on the Diagnostic and Statistical Manual on Mental Disorders, 4th edition and determined by the Composite International Diagnostic Interview (CIDI) Alcohol Module (Robins et al., 1988; WHO, 1996). More details on assessment and enrollment were previously published (Saitz et al., 2007).

At study entry in the hospital and 3 months later, alcohol consumption was assessed using a validated calendar method (Timeline Followback) (Sobell & Sobell, 1995). Readiness to change was assessed with a 1–10 visual analog scale (VAS) and with the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). The VAS measure was a response to: "How ready are you to change your drinking habits?" The VAS has not been extensively validated but is attractive for clinical use in busy settings because of its brevity (LaBrie, Quinlan, Schiffman, & Earleywine, 2005; Williams et al., 2007). The SOCRATES is a 19 item questionnaire developed to assess readiness to change alcohol use (Miller & Tonigan, 1996). Based on a factor analysis in this sample (Bertholet, Dukes, Hortton, Palfai, Pedley, & Saitz, 2009), we used a 2 factor structure for analyses on 16 items: 1—"Perception of Problems" (PP), and 2—"Taking Action" (TA). These two factors had good internal consistency (Cronbach's alpha 0.94 and 0.88, respectively). PP represents problem awareness and recognition of the need for additional help to address the drinking problem and TA denotes the concrete steps a person is taking or has already taken towards a decrease in drinking and commitment to change.

Table 1
Baseline characteristics of the 267 medical inpatients with unhealthy alcohol use ($n = 267$).

Demographics	
Women, no. (%)	81 (30.3)
Age, mean (SD)	45.0 (10.5)
Race/Ethnicity:	
Black, no. (%)	129 (48.3)
White, no. (%)	96 (36.0)
Hispanic, no. (%)	23 (8.6)
Other, no. (%)	19 (7.1)
Alcohol related diagnosis at hospital admission, no. (%)‡	129 (48.3)
Alcohol diagnosis (past year)†	
No diagnosis, no. (%) (risky drinking)	46 (17.2)
Alcohol abuse, no. (%)	13 (4.9)
Alcohol dependence, no. (%)	208 (77.9)
Alcohol consumption (past 30 days)	
Drinks per day, mean (SD), median	6.9 (9.0) 4.0
# of days with binge drinking, no. (%), median	12.8 (10.7) 9
Drug use (last 30 days)	
Heroin or cocaine use, no. (%)	68 (25.5)
Marijuana use, no. (%)	81 (30.6)
Readiness to change measures	
Visual analog scale*, mean (SD), median (IQR)	6.9 (3.5), 8(5, 10)
SOCRATES**:	
Perception of Problems, mean (SD), median (IQR)	35.6 (10.8), 39 (28, 44)
Taking Action, mean (SD), median (IQR)	21.2 (5.8), 22 (18, 26)

‡ Includes any of the following: acute alcoholic cirrhosis, alcoholic cardiomyopathy, alcoholic gastritis, alcoholic hepatitis, alcohol intoxication, alcoholic liver damage, alcoholic fatty liver, alcoholic pellagra, alcoholic polyneuropathy, alcohol withdrawal, alcohol withdrawal convulsion, alcohol withdrawal delirium, alcohol withdrawal hallucinosis, other alcoholic psychosis, alcoholic amnesic syndrome, other alcoholic dementia, alcoholic pancreatitis, or other diagnoses thought to be alcohol-attributable by the investigator (for example "holiday heart", alcoholic ketoacidosis, alcohol related rhabdomyolysis).
† Determined with the Composite International Diagnostic Interview (CIDI) Alcohol Module.
* How ready are you to change your drinking habits?" 1 to10.
** Perception of Problems, possible score: 10–50; Taking Action (commitment to change), possible score 6–30.
IQR: interquartile range (25th, 75th percentile).

2.2. Analysis

The primary outcome was the average number of standard drinks per day (past 30 days) assessed at 3 months. The main predictors of interest were the three readiness to change measures (VAS readiness, SOCRATES PP and TA). To avoid assumptions of a linear relation between readiness to change measures and outcome, each independent variable was categorized into quartiles, with the exception of VAS where division into quartiles was not feasible due to the distribution of the data and tertiles were used instead. The distribution of drinks per day at 3 months, a count variable, was skewed, with a considerable number of zeros and a long tail, so the use of models assuming normality was not adequate. Therefore we used overdispersed Poisson regression models to assess the effect of readiness to change on subsequent drinking (Horton, Kim, & Saitz, 2007; McCullagh & Nelder, 1989). The Pearson chi-square correction was used to account for overdispersion in the data. Separate models were fit for VAS readiness and each of the two SOCRATES factors, and controlled for drinking at study entry (drinks per day, last 30 days), randomization group (despite the fact that there was no effect of intervention group in the parent trial on drinking outcome, both groups having reduced their drinking), age, gender, any heroin or cocaine use (last 30 days), marijuana use (last 30 days) and presence of an alcohol related medical diagnosis at hospital admission. The magnitude of association between measures of readiness to change and drinking were quantified using incidence rate ratios (IRRs). The IRR is a ratio of drinks per day for the exposed group of interest (i.e. level of readiness) versus the reference group (lowest level of readiness). The null value of no association for the IRR is equal to 1. An IRR greater (less) than 1 indicates that readiness is associated with more (less) drinking. We calculated 95% confidence intervals for the IRR. 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).

3. Results

Of the 7824 individuals approached for the study, 2011 were excluded (1127 for language barriers, 392 because they were too confused, 492 because they had time conflict or declined). Of the 5813 screened, 4775 did not have unhealthy alcohol use, and for 52 amounts were not determined, leaving 986 patients with unhealthy alcohol use. Of these, 462 were ineligible (230 were not able to provide 1 or 2 contacts for follow-up purposes, 94 refused to complete the screening, 52 were moving in the next months, and 86 had a Mini Mental State Evaluation of less than 21) and 183 declined further participation. Therefore 341 subjects were included in the study (65% of the 524 eligible subjects). Eligible subjects who enrolled were more likely to be Black (45% vs 31%) and to drink greater amounts of alcohol (24 vs 18 drinks per week) compared to eligible subjects who refused participation. Of the 341 remaining eligible subjects, we excluded 7 because although they had unhealthy alcohol use according to the screening questions, results of the calendar assessment method were that they had not consumed risky amounts recently. An additional 67 of the 334 subjects (20.1%) were lost to follow-up and were excluded from the present analyses (Fig. 1). Those analyzed were not different from those lost to follow-up ($p>0.05$) with respect to alcohol consumption, readiness to change measures (VAS readiness, SOCRATES factors), age, gender, living with a partner, employment, homelessness, or drug use.

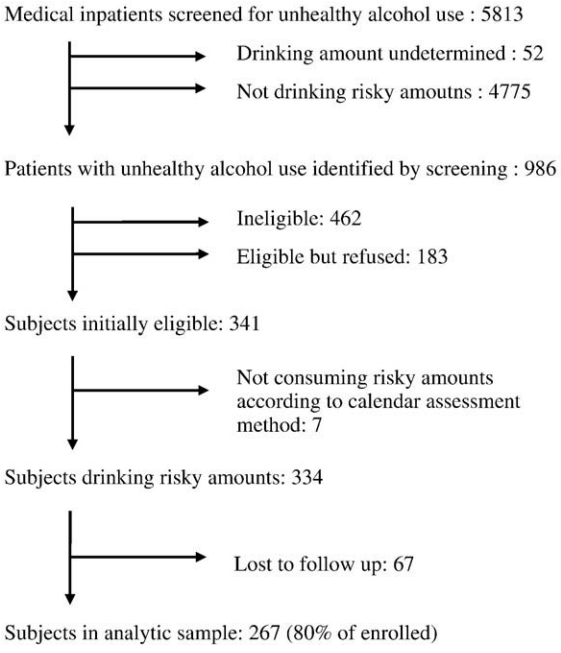


Fig. 1. Selection of subjects with unhealthy alcohol use for the prospective cohort study.

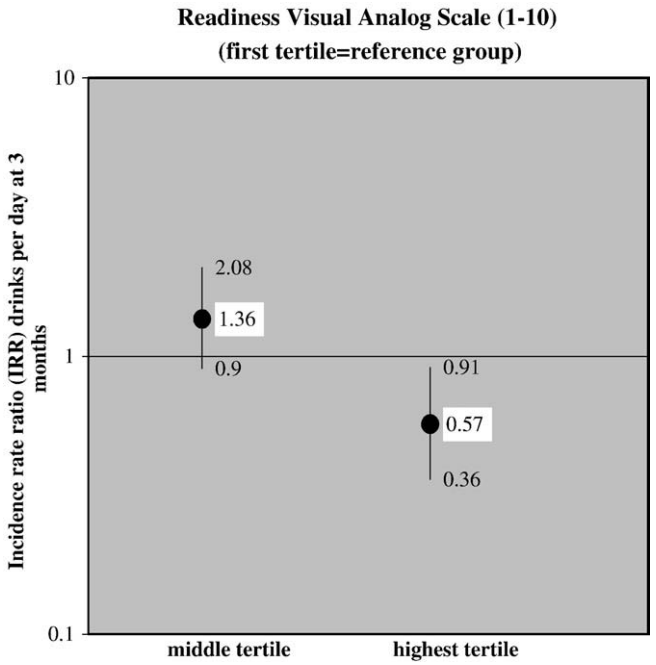


Fig. 2. Association between readiness to change (assessed with a visual analog scale) and drinks per day at 3 months†. †: Controlled for drinking at study entry (drinks per day, last 30 days), randomization group, age, gender, heroin or cocaine use (last 30 days), marijuana use (last 30 days) and presence of an alcohol related medical diagnosis at hospital admission. Lines represent 95% confidence intervals. Tertiles were used instead of quartiles because of the distribution of data (see text).

Subject's median (25th and 75th percentiles) readiness to change drinking at study entry on a visual analog scale (VAS) that ranged from 1 to 10 was 8 (5, 10). Median (25th and 75th percentiles) SOCRATES Perception of Problems (PP) score was 39 (28, 44) on a scale from 10 to 50. Median (25th and 75th percentiles) SOCRATES Taking Action (TA) score was 22 (18, 26) on a scale from 6 to 30. Other subjects' characteristics are presented in Table 1.

3.1. Association between readiness and consumption

In adjusted analyses, subjects with VAS-measured readiness in the highest tertile at study entry, compared to those in the lowest tertile, drank significantly *fewer* drinks per day 3 months later (incidence rate ratio [IRR] 0.57, 95% confidence interval [CI] 0.36, 0.91); the middle tertile was not significantly associated with consumption (IRR [95%CI] 1.36 [0.90, 2.08]) (Fig. 2).

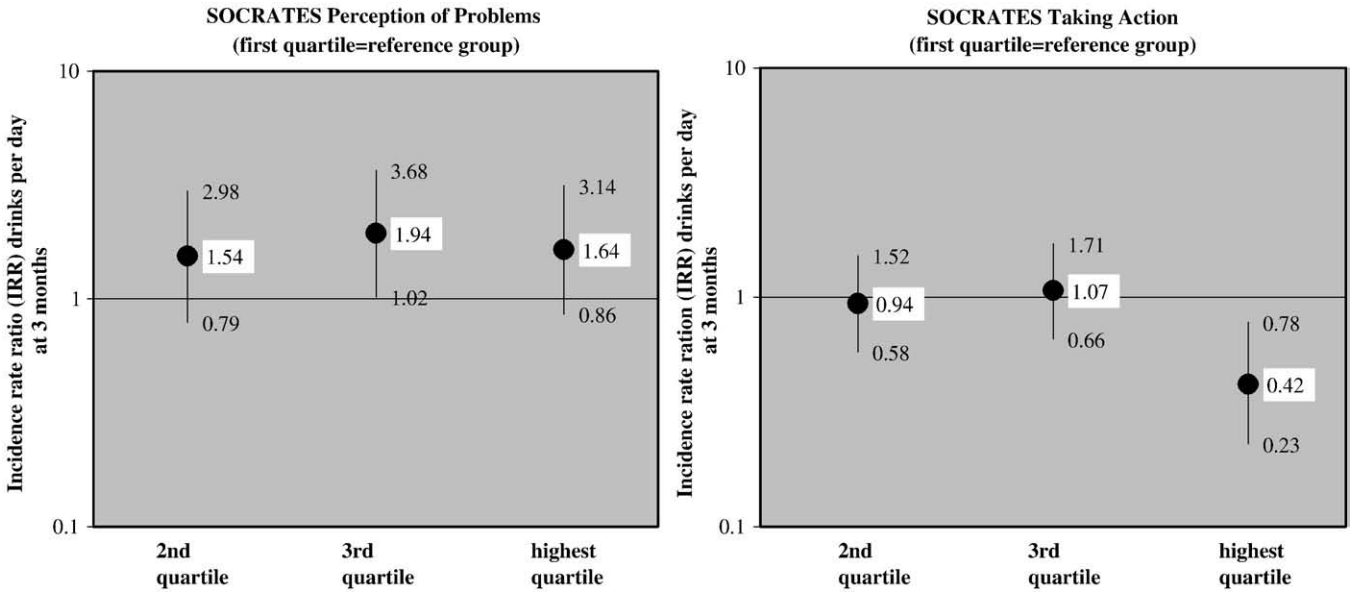


Fig. 3. Associations between readiness to change (assessed with the SOCRATES factor 1: Perception of Problems and factor 2: Taking Action) and drinks per day at 3 months†. †: Controlled for drinking at study entry (drinks per day, last 30 days), randomization group, age, gender, heroin or cocaine use (last 30 days), marijuana use (last 30 days) and presence of an alcohol related medical diagnosis at hospital admission. SOCRATES = Stages of Change Readiness and Treatment Eagerness Scale. Lines represent 95% confidence intervals.

Subjects with Perception of Problems (PP) scores in the 3rd quartile, compared with those in the lowest quartile, drank significantly more at 3 months (IRR [95% CI] 1.94 [1.02, 3.68]); drinking was also higher but associations were not significant for the highest and second quartiles (IRRs [95% CIs] for 2nd quartile: 1.54 [0.79, 2.98], and for highest quartile: 1.64 [0.86, 3.14]) (Fig. 3). Subjects with Taking Action (TA) scores in the highest quartile, compared with those in the lowest quartile, drank significantly less at 3 months (IRR [95% CI] 0.42 [0.23, 0.78]); associations were not significant for the second and third quartiles (IRRs [95% CIs] for 2nd quartile: 0.94 [0.58, 1.52], and for 3rd quartile: 1.07 [0.66, 1.71]) (Fig. 3).

4. Discussion

We explored whether readiness to change predicts subsequent alcohol consumption three months after a medical hospitalization among adults with unhealthy alcohol use. The predictive value of readiness to change appears to depend on the measures used, and on the various concepts included under the rubric of "readiness." Components of readiness to change measured by the SOCRATES questionnaire ("Perception of Problems" and "Taking Action" or commitment to change) appear to operate differently. Higher levels of Perception of Problems appeared to be associated with more, not less, drinking 3 months later. Clinically meaningful differences in consumption at 3 months were observed for each Perception of Problems quartile (compared to the lowest level), however the result was statistically significant only for the 3rd quartile. These results, consistent with prior literature, suggest that this measure may capture severity rather than an aspect of readiness associated with the ability to change in the future (Maisto, Conigliaro et al., 1999). In contrast, the highest level of Taking Action was associated with less drinking. Results suggested a threshold level for Taking Action; only the highest level was associated with a decrease in drinking. The highest level of Taking Action was the strongest predictor of decrease in drinking 3 months later of all the measures examined.

The implications of our results for the SOCRATES are that greater Perception of Problems is not a predictor of decreased drinking since it may be a measure of severity. In contrast to the denial stereotype, patients in this sample of mainly non-treatment seeking alcohol dependent adults appeared to recognize that their drinking patterns were problematic, which is consistent with other studies where higher readiness to change was associated with more consequences (Williams et al., 2006). Our results also underscore the importance of commitment to change, and change-related actions (i.e. concrete steps a person takes towards a decrease in drinking); a high level of Taking Action was predictive of less drinking, consistent with studies that have emphasized the impact of commitment to change on behavior change (Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003).

The second measure used, the visual analog scale, was associated with less drinking at the highest level. Of note, the middle tertile appeared to be associated with *more* drinking, although the results were not statistically significant. It is possible that different concepts (e.g. severity, and commitment) may be captured with this single scale, or that there may be simply threshold effects. Thus, outside its clinical utility to elicit discussion about alcohol use during a patient–physician interaction, the VAS should probably be used with caution as a predictive tool.

We are unaware of similar studies of the predictive value of readiness measures in medical inpatients with unhealthy alcohol use. Nonetheless, these results are comparable to those observed in other populations regarding the importance of commitment to change and planned action. Readiness is predictive of change when related to action in people with alcohol dependence recruited in psychiatric hospitals (Demmel et al., 2004). As such, counseling that supports self-efficacy and action plans may be particularly useful for medical inpatients with unhealthy alcohol use (Maisto, Carey, & Bradizza, 1999). Our results also add to the evidence linking problem recognition to severity. Individuals with more severe problems related to their alcohol consumption are more likely to have high level of recognition (Williams et al., 2007, 2006).

These findings should be considered in the context of limitations of the present study. First, our subjects agreed to participate in a study in which they could receive alcohol counseling. This sample could have been predisposed to change. In addition, participants drank more than those who refused to enroll in the study. However, enrolled subjects and refusers did not differ on VAS-measured readiness to change. Given the approach to recruitment in this study, these results are likely generalizable to individuals who agree to talk about their alcohol consumption after screening. Second, although recommended for use based on its face validity, the visual analog scale for readiness

has not been extensively validated. Third, the follow-up period was 3 months. This could be considered short. We believe it likely that any effects of readiness on drinking would be present in the short-term. Whether these effects would persist (or appear anew) in long term follow-up is an empirical question for future study.

Secondary analysis of randomized trial data can raise methodological challenges. But unlike secondary analyses in other cohorts, an intervention is well-specified and its receipt and effects are known. In this sample, the intervention delivered was not effective: as demonstrated in the parent study, both groups (intervention and usual care) reduced their drinking without statistically significant differences between groups (Saitz et al., 2007). In addition we controlled all analyses for randomization group. Furthermore, the follow-up rate was 80%, which is good for this type of population and did not differ by group. Finally, prospective data collection and follow-up in this randomized trial is a strength because of high quality assessments and a detailed description of the study population.

Our study also had notable strengths. We used a large sample of mostly dependent medical inpatients, a population that is of great interest since they are at higher risk of morbidity and mortality, represent the vast majority of medical inpatients with unhealthy alcohol use (Freyer-Adam et al., 2008; Saitz, Freedner, Palfai, Horton, & Samet, 2006), and are generally not seeking alcohol treatment despite experiencing alcohol consequences. Data collection was prospective and loss to follow-up was relatively small. We used a factor structure for the SOCRATES questionnaire that was determined in medical inpatients. Since this questionnaire has been shown to have different factor structure across settings, it was a strength to use a structure determined for this sample (Demmel et al., 2004; Figlie, Dunn, & Laranjeira, 2004; Maisto, Chung, Cornelius, & Martin, 2003; Maisto, Conigliaro et al., 1999; Miller & Tonigan, 1996). By using various measures of readiness to change, we were able to test different aspects of readiness. Also, by avoiding the assumption of a linear relation between readiness to change measures and drinking outcome in our analyses, we were able to identify a more refined relation between readiness and outcome.

In conclusion, readiness to change does predict subsequent alcohol consumption in medical inpatients with unhealthy alcohol use, however, readiness appears to be less useful as a monolithic construct. Readiness appears to be comprised of a number of more specific constructs. These findings may explain, at least in part, inconsistencies in the literature regarding the predictive ability of "readiness" and the role and importance of "readiness" for behavior change. In medical inpatients, the readiness construct "Perception of Problems" (or problem recognition) appears to be a measure of severity rather than as a predictor of change. Nevertheless, it may still be useful to assess problem recognition as a first step to set goals and to plan actions to reduce drinking or to attend specialty treatment. Visual analog scale measures of readiness have some ability to predict behavior change however this ability may operate only after some threshold. Since a high level of taking action/commitment to change appeared to predict less drinking, physicians should be encouraged to enhance their patient's self-efficacy and help them set goals and action plans. For greatest clinical and predictive utility, the component concepts of readiness should be separated when they are measured and used in research and practice. Additional studies of these separate constructs may help determine their optimal use.

References

Amrhein, P. C., Miller, W. R., Yahne, C. E., Palmer, M., & Fulcher, L. (2003). Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of Consulting and Clinical Psychology*, *71*(5), 862–878.

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. *Archives of Internal Medicine*, *165*(9), 986–995.

Bertholet, N., Dukes, K., Horton, N. J.., Palfai, T. P., Pedley, A., & Saitz, R. (2009). Factor structure of the Socrates questionnaire in hospitalized medical patients. *Addictive Behaviors*, *34*(6–7), 568–572, doi:10.1016/j.addbeh.2009.03.013.

Demmel, R., Beck, B., Richter, D., & Reker, T. (2004). Readiness to change in a clinical sample of problem drinkers: Relation to alcohol use, self-efficacy, and treatment outcome. *European Addiction Research*, *10*(3), 133–138.

Figlie, N. B., Dunn, J., & Laranjeira, R. (2004). Factor structure of the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES) in alcohol dependent outpatients. *European Addiction Research*, *26*(2), 91–99.

Forsberg, L., Ekman, S., Halldin, J., & Ronnberg, S. (2004). The readiness to change questionnaire: Reliability and validity of a Swedish version and a comparison of scoring methods. *British Journal of Health Psychology*, *9*(Pt 3), 335–346.

Freyer-Adam, J., Coder, B., Baumeister, S. E., Bischof, G., Riedel, J., Paatsch, K., et al. (2008). Brief alcohol intervention for general hospital inpatients: A randomized controlled trial. *Drug and Alcohol Dependence*, *93*(3), 233–243.

Horton, N. J., Kim, E., & Saitz, R. (2007). A cautionary note regarding count models of alcohol consumption in randomized controlled trials. *BMC Medical Research Methodology*, *7*, 9.

Isenhart, C. E. (1997). Pretreatment readiness for change in male alcohol dependent subjects: Predictors of one-year follow-up status. *Journal of Studies on Alcohol*, *58*(4), 351–357.

LaBrie, J. W., Quinlan, T., Schiffman, J. E., & Earleywine, M. E. (2005). Performance of alcohol and safer sex change rulers compared with readiness to change questionnaires. *Psychology of Addictive Behaviors*, *19*(1), 112–115.

Maisto, S. A., Carey, K. B., & Bradizza, C. M. (1999). Social learning theory. In K. E. Leonard & H. T. Blane (Eds.), *Psychological theories of drinking and alcoholism* (pp. 106–163). New York: Guilford Press.

Maisto, S. A., Conigliaro, J., McNeil, M., Kraemer, K., O'Connor, M., & Kelley, M. E. (1999). Factor structure of the SOCRATES in a sample of primary care patients. *Addictive Behaviors*, *24*(6), 879–892.

Maisto, S. A., Chung, T. A., Cornelius, J. R., & Martin, C. S. (2003). Factor structure of the SOCRATES in a clinical sample of adolescents. *Psychology of Addictive Behaviors*, *17*(2), 98–107.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London, England: Chapman and Hall.

Miller, W. R., & Tonigan, J. S. (1996). Assessing drinker's motivation for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). *Psychology of Addictive Behaviors*, *10*, 81–89.

Nye, E. C., Agostinelli, G., & Smith, J. E. (1999). Enhancing alcohol problem recognition: a self-regulation model for the effects of self-focusing and normative information. *Journal of Studies on Alcohol*, *60*(5), 685–693.

Reed Jr., D. N., Wolf, B., Barber, K. R., Kotlowski, R., Montanez, M., Saxe, A., et al. (2005). The stages of change questionnaire as a predictor of trauma patients most likely to decrease alcohol use. *Journal of the American College of Surgeons*, *200*(2), 179–185.

Rehm, J., Room, R., Monteiro, M., Gmel, G., Graham, K., Rehn, N., et al. (2003). Alcohol as a risk factor for global burden of disease. *European Addiction Research*, *9*(4), 157–164.

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., et al. (1988). The Composite International Diagnostic Interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry*, *45*(12), 1069–1077.

Rollnick, S. (1998). Readiness, importance and confidence: critical conditions of change in treatment. In W. R. Miller & N. Heather (Eds.), *Treating addictive behaviour* (2nd ed.). New York: Plenum.

Saitz, R., Freedner, N., Palfai, T. P., Horton, N. J., & Samet, J. H. (2006). The severity of unhealthy alcohol use in hospitalized medical patients. The spectrum is narrow. *Journal of General Internal Medicine*, *21*(4), 381–385.

Saitz, R., Palfai, T. P., Cheng, D. M., Horton, N. J., Freedner, N., Dukes, K., et al. (2007). Brief intervention for medical inpatients with unhealthy alcohol use: a randomized, controlled trial. *Annals of Internal Medicine*, *146*(3), 167–176.

Samet, J. H., Rollnick, S., & Barnes, H. (1996). Beyond CAGE. A brief clinical approach after detection of substance abuse. *European Addiction Research*, *156*(20), 2287–2293.

Smith, K. L., Horton, N. J., Saitz, R., & Samet, J. H. (2006). Use of mini-mental state examination in recruitment for substance abuse research studies. *Drug and Alcohol Dependence*, *82*, 231–237.

Sobell, L. C., & Sobell, M. B. (1995). *Alcohol timeline followback (TLFB) users' manual.* Toronto, Canada: Addiction Research Foundation.

Solberg, L. I., Maciosek, M. V., & Edwards, N. M. (2008). Primary care intervention to reduce alcohol misuse ranking its health impact and cost effectiveness. *American Journal of Preventive Medicine*, *34*(2), 143–152.

WHO (1996). *Composite International Diagnostic Interview (CIDI)(Core Version 2.0).* Geneva, Swtizerland: World Health Organization.

Williams, E. C., Horton, N. J., Samet, J. H., & Saitz, R. (2007). Do brief measures of readiness to change predict alcohol consumption and consequences in primary care patients with unhealthy alcohol use? *Alcoholism, Clinical and Experimental Research*, *31*(3), 428–435.

Williams, E. C., Kivlahan, D. R., Saitz, R., Merrill, J. O., Achtmeyer, C. E., McCormick, K. A., et al. (2006). Readiness to change in primary care patients who screened positive for alcohol misuse. *Annals of Family Medicine*, *4*(3), 213–220.

Addictive Behaviors 34 (2009) 568–572

Contents lists available at ScienceDirect

Addictive Behaviors

Factor structure of the SOCRATES questionnaire in hospitalized medical patients

Nicolas Bertholet [a,b,*], Kim Dukes [c], Nicholas J. Horton [d], Tibor P. Palfai [e,f], Alison Pedley [c], Richard Saitz [a,f]

[a] Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA
[b] Clinical Epidemiology Center, Institute of Social and Preventive Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland
[c] DM-STAT Inc., Malden, MA, USA
[d] Department of Mathematics and Statistics, Smith College, Northampton, MA, USA
[e] Department of Psychology, Boston University, Boston, MA, USA
[f] Youth Alcohol Prevention Center and Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

ARTICLE INFO

Keywords:
Stages of Change Readiness and Treatment Eagerness Scale
Factor structure
Medical inpatients
Unhealthy alcohol use

ABSTRACT

The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES), a 19-item instrument developed to assess readiness to change alcohol use among individuals presenting for specialized alcohol treatment, has been used in various populations and settings. Its factor structure and concurrent validity has been described for specialized alcohol treatment settings and primary care. The purpose of this study was to determine the factor structure and concurrent validity of the SOCRATES among medical inpatients with unhealthy alcohol use not seeking help for specialized alcohol treatment. The subjects were 337 medical inpatients with unhealthy alcohol use, identified during their hospital stay. Most of them had alcohol dependence (76%). We performed an Alpha Factor Analysis (AFA) and Principal Component Analysis (PCA) of the 19 SOCRATES items, and forced 3 factors and 2 components, in order to replicate findings from Miller and Tonigan (Miller, W. R., & Tonigan, J. S., (1996). Assessing drinkers' motivations for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). Psychology of Addictive Behavior, 10, 81–89.) and Maisto et al. (Maisto, S. A., Conigliaro, J., McNeil, M., Kraemer, K., O'Connor, M., & Kelley, M. E., (1999). Factor structure of the SOCRATES in a sample of primary care patients. Addictive Behavior, 24(6), 879–892.). Our analysis supported the view that the 2 component solution proposed by Maisto et al. (Maisto, S.A., Conigliaro, J., McNeil, M., Kraemer, K., O'Connor, M., & Kelley, M.E., (1999). Factor structure of the SOCRATES in a sample of primary care patients. Addictive Behavior, 24(6), 879–892.) is more appropriate for our data than the 3 factor solution proposed by Miller and Tonigan (Miller, W. R., & Tonigan, J. S., (1996). Assessing drinkers' motivations for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). Psychology of Addictive Behavior, 10, 81–89.). The first component measured Perception of Problems and was more strongly correlated with severity of alcohol-related consequences, presence of alcohol dependence, and alcohol consumption levels (average number of drinks per day and total number of binge drinking days over the past 30 days) compared to the second component measuring Taking Action. Our findings support the view that the SOCRATES is comprised of two important readiness constructs in general medical patients identified by screening.

1. Introduction

Brief motivational counseling interventions have efficacy for people with nondependent unhealthy alcohol use (Bertholet, Daeppen, Wietlisbach, Fleming, & Burnand, 2005; Bien, Miller, & Tonigan, 1993; Dunn, Deroo, & Rivara, 2001; Saitz, 2005). Motivational interviewing is an extended intervention that has efficacy for a number of health behaviors, including alcohol dependence (Carbonari & DiClemente, 2000; Project MATCH Research Group, 1997). Readiness-to-change and motivation are frequently viewed as intermediate outcomes and have been seen as mediators and potential predictors of change (Demmel, Beck, Richter, & Reker, 2004; Heather, Rollnick, & Bell, 1993; Maisto et al., 1999; Williams, Horton, Samet, & Saitz, 2007). As a result, the assessment of motivation to change has been of great interest to researchers and clinicians alike (Fiellin, Reid, & O'Connor, 2000; Miller & Rollnick, 1991).

The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES) was developed and designed to measure stages of readiness to change alcohol use (Miller & Tonigan, 1996). It was first intended as a self-administered questionnaire to categorize individuals into one of four stages of change (pre-contemplation, contemplation, determination, and action (Prochaska & DiClemente, 1984). After several iterations of the SOCRATES, Miller and Tonigan validated

* Corresponding author. Centre de Traitement en Alcoologie (CTA), Département Universitaire de Médecine et Santé Communautaires (DUMSC), 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).

0306-4603/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.addbeh.2009.03.013

a shorter (19-item) version of the SOCRATES in 1996 and reported on the factor structure in a population of participants with alcohol dependence in a multi-site clinical treatment trial [9]. These participants were in specialty settings: either in outpatient treatment programs or seen for aftercare following residential or day treatment. The authors identified 3 independent factors (using exploratory alpha factor analysis with a varimax rotation): Ambivalence, Recognition, and Taking Steps. The 19-item version of the SOCRATES is widely used and non-English versions of the scale have been validated among treatment seeking populations (Demmel et al., 2004; Figlie, Dunn, & Laranjeira, 2005). Each item response is based on a 5-point Likert scale (i.e., 1 = strongly disagree, 2 = disagree, 3 = undecided or unsure, 4 = agree, 5 = strongly agree). The 19-item SOCRATES is included in Table 2; more information is available online at http://casaa.unm.edu/inst/SOCRATESv8.pdf.

Because most of the research on the psychometric properties of the SOCRATES has been done in specialized settings in patients with alcohol dependence, there has been concern about whether the questionnaire's factor structure is applicable to non-treatment seeking patients in other settings.

To address the question of the applicability of the factor structure in other populations, Maisto et al. (1999) investigated the factor structure of the SOCRATES in a population of opportunistically screened primary care patients. Using exploratory component analysis and confirmatory factor analysis, they concluded that a two factor solution (retaining 15 items) was more parsimonious than the three factor solution. Specifically, the first factor contained 9 of the Ambivalence and Recognition items (named AMREC) and the second factor comprised 6 of the Taking Steps items (and was named Taking Steps).

Several publications have reported either a 2 or 3-factor solution for the SOCRATES (Burrow-Sanchez & Lundberg, 2007; Demmel et al., 2004; Figlie et al., 2005). Given inconsistencies in the literature among different populations, we investigated the most appropriate factor structure solution based on a sample hospitalized for medical illness in a general hospital (i.e. not seeking or receiving specialty alcohol treatment) who were identified opportunistically by screening for the spectrum of unhealthy alcohol use (i.e. from use of risky amounts through dependence). This is to our knowledge the first study to investigate the appropriate factor structure solution of the SOCRATES in this population. We performed both an exploratory analysis replicating techniques utilized by and compared our results to those of Miller and Tonigan (1996), and Maisto et al. (1999).

2. Methods

2.1. Participants

The participants were enrolled in a randomized trial of a brief motivational intervention for unhealthy alcohol use (Saitz et al., 2007). They were recruited while on the inpatient internal medicine service of a large, urban, academic medical center hospital. Eligibility criteria included: 18 or more years old, fluent in English or Spanish, currently (past month) drinking risky amounts (defined as more than 14 standard drinks per week or 5 or more drinks per occasion for men 18 to 65 years of age, and more than 11 standard drinks per week or 4 or more drinks per occasion for women and people over age 65), availability of 2 contacts to assist with follow-up, no plans to relocate in the next 2 years, and a Mini-Mental State Examination score of at least 21. Subjects provided written informed consent and completed the SOCRATES at the time of enrollment.

Study assessments (collected prior to randomization) were administered by trained research associates. Questions regarding alcohol consumption and consequences, medical and mental health, and health care utilization as well as other domains were obtained. Subjects completed the Short Inventory of Problems, a questionnaire assessing alcohol-related consequences (Miller, Tonigan, & Longabaugh, 1995), and the Alcohol Use Disorders Identification Test (Saunders, Aasland, Babor, de la Fuente, & Grant, 1993). The presence of an alcohol use disorder diagnosis was determined with the Composite International Diagnostic Interview (CIDI) Alcohol Module (WHO, 1996). Alcohol consumption was assessed with a validated calendar method (30-day Timeline Followback) (Sobell & Sobell, 1995). The study was approved by the Boston University Medical Center Institutional Review Board and all subjects provided written informed consent. Details of the methodology of the randomized controlled trial are reported elsewhere (Saitz et al., 2007).

2.2. Analyses

The first step of the analysis was to ensure that the underlying assumptions of the factor analytic model were appropriate and to investigate the distributional properties of each of the 19 items comprising the SOCRATES. In order to replicate findings from Miller and Tonigan (1996) and Maisto et al. (1999), we performed an Alpha Factor Analysis (AFA) and Principal Component Analysis (PCA) of the 19 SOCRATES items, using orthogonal (varimax) rotation and forced the 3 factors and 2 components, respectively. When evaluating the factor and component structures, we retained items with component or factor loadings ≥ 0.4 and with factorial complexity of one (Nunally & Bernstein, 1994). In order to assess the potential impact of the inadvertent omission of item 16 of the SOCRATES by Maisto et al., we also performed analyses removing this item to allow comparison with their findings. As a confirmatory FA technique we computed the coefficient of congruence to compare matrix structures between datasets or studies (Cureton & D'Agostino, 1993). Internal consistency using Cronbach's alpha coefficients was assessed utilizing items that loaded on our factors or components. Cluster-based scores were created for each factor or component and concurrent validity was assessed for demographic, drinking and drug use measures using Pearson and Spearman correlation coefficients. All computations were performed using SAS Version 9.1. Unless otherwise specified, all statistical tests and/or confidence intervals were performed at $\alpha = 0.05$ (2-sided).

3. Results

3.1. Participants

Of 986 medical inpatients who reported at-risk drinking amounts during screening, 341 enrolled. Subjects enrolled were more likely to be African American (45% vs 31%) and drank larger quantities of alcohol (median 24 vs 18 drinks per week) compared to eligible subjects who refused participation, but did not significantly differ on readiness to change measured with a 1 to 10 visual analog scale. Of enrolled subjects, 337 completed the SOCRATES and comprise our analytic sample. Baseline characteristics are presented in Table 1.

3.2. Determination of the most appropriate factor structure solution in the study sample

3.2.1. Principal Component Analysis (PCA)

The PCA yielded 2 components based upon Kaiser's Rule (i.e., eigenvalue >1). The first three eigenvalues were 8.97, 2.65, and 0.91. The orthogonal component structure and associated loadings from the PCA with a varimax rotation are presented in Table 2. Sixteen of the 19 items comprising the SOCRATES were retained based on the criteria of having a component loading ≥ 0.4 and a factorial complexity of one. The first component, accounting for 37% of the item response variance, was comprised of 10 items: 3 that were originally described as Ambivalence, 6 as Recognition, and 1 as Taking Steps. Six items originally described as Taking Steps composed the second component, accounting for an additional 24% of the item response variation. The remaining

Table 1
Subjects' baseline characteristics[a] ($n = 337$).

Demographics	
Women	99 (29.4%)
Age, mean (SD) median	44.4 (10.7) 45.0
Race/ethnicity	
Black	153 (45.4%)
White	132 (39.2%)
Hispanic	29 (8.6%)
Other	23 (6.8%)
Alcohol diagnosis (past year)[b]	
Alcohol abuse	15 (4.5%)
Alcohol dependence	257 (76.3%)
No diagnosis	65 (19.3%)
Alcohol consumption (past 30 days) mean (SD) median	
Drinks per day	6.8 (9.0) 3.7
Days with heavy episodic drinking	12.8 (10.7) 9.0
Drug use (last 30 days)	
Heroin or cocaine use	88 (25.8%)

Alcohol consumption was assessed over the past 30 days with the Timeline Followback. Heavy episodic drinking was defined as drinking more than 5 drinks per occasion for men and more than 4 drinks per occasion for women and persons >65 years.

[a] Number and percentage are presented for categorical variables, Mean (SD) and median are presented for continuous variables.

[b] Determined with the Composite International Diagnostic Interview (CIDI) Alcohol Module.

items split a load between two components (items 1 and 19) or did not load on either component (item 11).

3.3. Comparison with previously published factor structure solutions

3.3.1. Maisto et al.

To compare our data and findings with those of Maisto et al. (1999), we repeated the PCA excluding item 16 (Table 3). In these analyses, the PCA using Kaiser's Rule yielded 2 factors (first three eigenvalues: 8.57, 2.62, 0.87). The first component is comprised of 11 items (using scale names from Miller and Tonigan): 2 Ambivalence, 7 Recognition and 2 Taking Steps, while the second component contains 8 items: 1 Recognition and 7 Taking Steps. Fifteen of the 18 SOCRATES items included in this procedure had a component loading ≥ 0.4 and a factorial complexity of one, and as with the prior analysis, two of the items (1 and 19) split a load between two components and item 11 did not load on either component. Analyses performed with a non-orthogonal rotation without item 16 yielded similar results.

3.3.2. Miller and Tonigan

The hypothesized underlying three factor structure, based on Miller and Tonigan's (1996) published work, is displayed in Table 3. This structure is comprised of 4 Ambivalence items (items 2, 6, 11 and 16), 7 Recognition items (items 1, 3, 7, 10, 12, 15 and 17), and 8 Taking Steps items (items 4, 5, 8, 9, 13, 14, 18 and 19). Utilizing Miller's approach with our data, the AFA yielded 2 factors using Kaiser's rule. However, for comparison, we assessed a forced 3 factor solution. The first three eigenvalues were 14.09, 3.81 and 0.80. The orthogonal factor structure and associated loadings from the AFA with a varimax rotation are presented in Table 3. Three of the 4 hypothesized Ambivalence items, all 7 of the hypothesized Recognition items, and 2 of the 8 hypothesized Taking Steps items loaded on the first factor. Seven of the 8 hypothesized Taking Steps items and 1 of the hypothesized Recognition items loaded on the second factor. No items loaded on the third factor. Analyses performed utilizing a non-orthogonal solution yielded similar results (not shown).

3.3.3. Confirmatory factor analytic techniques

To compare our component and factor analytic structure with that described by Miller and Tonigan (1996) and Maisto et al. (1999), we estimated the coefficients of congruence. Coefficients of congruence (CC) range from −1 to 1, with greater absolute coefficients indicating increased concordance between structures (Cureton & D'Agostino, 1993). Comparing the 3 factor structure solution from our data to the structure reported by Miller and Tonigan (1996), the coefficients of congruence implied strong concordance between both of the Recognition factors and the Taking Steps factors ($CC = 0.885$ and 0.963, respectively) and a weak dissimiliarity between the Ambivalence factors ($CC = -0.527$, due to no items from our data loading on the third factor). For the 2 component structure compared to the solution published by Maisto et al., the coefficients of congruence were 0.988 (for comparison with Maisto et al.'s AMREC factor) and 0.985 (for comparison with their Taking Steps factor), for the first and second components from our data, implying excellent concordance. The 2 component solution proposed by Maisto et al. is therefore more appropriate for our data in comparison to the 3 factor solution. A confirmatory factor analysis (also omitting item 16) for the 2 and 3 factor models also provided better fit for the 2 factor solution (not shown).

3..4. Concurrent validity

To assess concurrent validity we developed cluster-based scores based on the two component PCA solution and assessed the associations between these scores and important measures. Results are reported in Table 4. Both of the cluster-based scores are correlated with the presence of alcohol dependence, the presence of alcohol-related problems, and Alcohol Use Disorder Identification Test (AUDIT) score. Component 1 consistently has stronger correlations with each of these alcohol-related measures. In addition, component 1 is significantly correlated with alcohol consumption levels (average number of drinks

Table 2
Determination of the most appropriate component structure solution in the study sample (exploratory PCA).

Item	Question	Component 1	Component 2
1	I really want to make changes in my drinking	0.49	0.46
2	Sometimes I wonder if I am an alcoholic	**0.73**	0.00
3	If I don't change my drinking soon, my problems are going to get worse	**0.78**	0.24
4	I have already started to make some changes in my drinking	0.14	**0.80**
5	I was drinking too much at one time, but I've managed to change my drinking	0.02	**0.58**
6	Sometimes I wonder if my drinking is hurting other people	**0.69**	0.22
7	I am a problem drinker	**0.87**	0.09
8	I am not just thinking about changing my drinking, I am already doing something about it	0.26	**0.80**
9	I have already changed my drinking, and I am looking for ways to keep from slipping back into my old pattern	0.16	**0.80**
10	I have a serious problem with drinking	**0.85**	0.19
11	Sometimes I wonder if I am in control of my drinking	0.39	0.26
12	My drinking is causing a lot of harm	**0.79**	0.22
13	I am actively doing things now to cut down or stop drinking	0.15	**0.82**
14	I want help to keep from going back to the drinking problems that I had before	**0.72**	0.38
15	I know that I have a drinking problem	**0.88**	0.22
16	There are times when I wonder if I drink too much	**0.64**	0.21
17	I am an alcoholic	**0.85**	0.12
18	I am working hard to change my drinking	0.33	**0.79**
19	I have made changes in my drinking and I want some help to keep from going back to the way I used to drink	0.51	0.55
Percentage of variance explained		37%	24%
Cronbach's alpha (standardized)		0.94	0.88

Bolded values indicate that the item had a component loading of 0.4 or greater on one and only one component.

Table 3
Comparison with previously published component and factor structure solutions.

Item	Present study		Maisto et al.		Present study			Miller and Tonigan		
	C1	C2	C1 (AMREC)	C2 (TS)	F1	F2	F3	F1 (R)	F2 (TS)	F3 (A)
1. I really want to make changes in my drinking	**0.49**	**0.46**	[a]	[a]	**0.48**	**0.42**	−0.05	**0.38**	0.16	0.04
2. Sometimes I wonder if I am an alcoholic	**0.72**	0.00	**0.58**	0.26	**0.67**	0.05	−0.07	−0.07	−0.07	**0.58**
3. If I don't change my drinking soon, my problems are going to get worse	**0.78**	0.24	**0.82**	0.14	**0.77**	0.25	−0.02	**0.6**	0.12	0.05
4. I have already started to make some changes in my drinking	0.14	**0.8**	0.15	**0.81**	0.15	**0.79**	−0.13	0.15	**0.73**	0
5. I was drinking too much at one time, but I've managed to change my drinking	0.01	**0.58**	0.02	**0.65**	0.06	**0.48**	−0.1	−0.24	**0.4**	0.16
6. Sometimes I wonder if my drinking is hurting other people	**0.68**	0.23	**0.58**	0.29	**0.67**	0.25	−0.24	0.3	0.07	**0.31**
7. I am a problem drinker	**0.88**	0.09	**0.84**	0.12	**0.85**	0.1	0.05	**0.61**	0.03	0.04
8. I am not just thinking about changing my drinking, I am already doing something about it	0.26	**0.8**	0.18	**0.85**	0.25	**0.77**	0.1	0.25	**0.69**	−0.06
9. I have already changed my drinking, and I am looking for ways to keep from slipping back into my old pattern	0.16	**0.8**	0.26	**0.8**	0.15	**0.77**	0.1	0.09	**0.81**	−0.02
10. I have a serious problem with drinking	**0.86**	0.19	**0.82**	0.15	**0.83**	0.2	0.23	**0.8**	0.09	−0.02
11. Sometimes I wonder if I am in control of my drinking	0.38	0.27	**0.63**	0.22	0.37	0.25	−0.14	−0.06	0	**0.55**
12. My drinking is causing a lot of harm	**0.79**	0.23	**0.79**	0.11	**0.76**	0.24	−0.03	**0.62**	0.15	−0.01
13. I am actively doing things now to cut down or stop drinking	0.15	**0.82**	0.23	**0.8**	0.16	**0.78**	0.03	0.22	**0.76**	−0.04
14. I want help to keep from going back to the drinking problems that I had before	**0.72**	0.38	[a]	[a]	**0.69**	0.38	0.28	0.45	**0.46**	0
15. I know that I have a drinking problem	**0.87**	0.22	**0.8**	0.32	**0.86**	0.23	0.13	**0.76**	0.15	−0.03
16. There are times when I wonder if I drink too much	[b]	[b]	[b]	[b]	**0.61**	0.24	−0.24	0.06	0.04	**0.66**
17. I am an alcoholic	**0.86**	0.12	**0.77**	0.13	**0.82**	0.14	0.15	**0.68**	0.22	−0.18
18. I am working hard to change my drinking	0.32	**0.79**	0.37	**0.78**	0.32	**0.78**	0.05	0.28	**0.76**	−0.05
19. I have made changes in my drinking and I want some help to keep from going back to the way I used to drink	**0.52**	**0.55**	[a]	[a]	**0.49**	**0.53**	0.27	0.16	**0.68**	0.06
Variance explained by each component/factor	37%	25%	48%	13%	35%	22%	2%	11%	27%	7%
Cronbach standardized coeffficient alpha	0.92	0.88	0.91	0.89	0.93	0.89	0.72	0.85	0.83	0.60

Bolded values indicate that in the present study, the item had a component or factor loading of 0.4 or greater on one and only one component or factor. For the Maisto et al., and Miller & Tonigan study columns, bolded values indicate that the item was retained by those authors according to the methods described in their publications Maisto et al. (1999) and Miller and Tonigan (1996).
C: component.
F: factor.
A: Ambivalence, R: Recognition, TS: Taking Steps, AMREC: Ambivalence and Recognition (see text).
[a] Item not retained in Maisto (did not load on one and only one factor at >0.40).
[b] Because of clerical error in original Maisto et al. study, item was not included in analyses.

per day and total number of heavy drinking days over the past 30 days). Neither component is significantly correlated with drug use.

4. Discussion

We examined the factor structure of the SOCRATES questionnaire in a population of adult men and women with unhealthy alcohol use identified by opportunistic screening, who were hospitalized in a general hospital and not attending specialized alcohol treatment.

Table 4
Concurrent validity — correlation[a] of components with various clinical variables.

Variable	First component	Second component
Demographics		
Age	0.07	0.09
Gender	0.04	0.07
Alcohol measures		
Alcohol dependence diagnosis (DSM-IV)[c]	0.55[b]	0.30[b]
Alcohol-related problems (SIP score)	0.73[b]	0.24[b]
AUDIT score	0.68[b]	0.21[b]
Alcohol consumption (average drinks per day)	0.39[b]	−0.02
Number of binge drinking days (past 30 days)	0.47[b]	0.01
Drug use		
Heroin or cocaine use (past 30 days)	0.10	0.07
Marijuana use (past 30 days)	−0.02	−0.06

SIP: Short Inventory of Problems.
AUDIT: Alcohol Use Disorder Identification Test.
Alcohol consumption was assessed with the Timeline Followback.
[a] Spearman correlation coefficients are presented for dichotomous variables, and Pearson correlation coefficients are presented for continuous variables.
[b] $p<0.05$.
[c] Determined with the Composite International Diagnostic Interview (CIDI) Alcohol Module.

We found a 2 component structure. We propose that the first component be named "Perception of Problems" (PP) (consisting of 10 items) and the second component be named "Taking Action" (TA) (consisting of 6 items). PP includes 3 items originally classified by Miller and Tonigan as Ambivalence, 6 as Recognition and one as Taking Steps. This component reflects the cognitive dimension of acceptance and recognition of alcohol problems. Item 14, "I want help to keep from going back to the drinking problems that I had before," originally classified as Taking Steps, is part of PP. This could reflect that the acceptance of needing help is more a recognition of an underlying problem than an action statement. PP appears to reflect both the perception of problems related to alcohol drinking and a need for help. TA consists of 6 items originally described as Taking Steps, and appears to report actions that individuals are already doing in order to address their drinking problem. The desire to get help appears to be separate from taking actions to change drinking behavior.

In the PCA, 3 of the 19 items had component loadings ≥0.4 with a factorial complexity of one. Item 19 ("I have made changes in my drinking and I want some help to keep from going back to the way I used to drink"), originally classified as Taking Steps, is a composite question made of two statements, one on changes already made in drinking and the other on the desire to get help. It loaded ≥0.4 on both components, consistent with what the 2 factors appear to capture. Item 1 loaded ≥0.4 on both components, and item 11 did not load ≥0.4 on any component.

The assessment of concurrent validity indicates that PP (the first component) is correlated with alcohol consumption level (drinks per day and heavy drinking episodes). PP has stronger associations with the presence of alcohol-related consequences than TA (the second component). This is consistent with the interpretation that PP reflects perception of alcohol problems, but suggests also that PP could reflect the severity of the problems related to alcohol use.

The exploratory analysis results were reinforced by the comparative analysis. In comparative and confirmatory analyses, the most appropriate structure in our data was similar to that found by Maisto et al. and less similar to that found by Miller and Tonigan.

Our data support the evidence that the factor structure of the SOCRATES questionnaire may be dependent upon the population and the therapeutic setting in which the questionnaire is administered. The use of a 3 factor solution seems to be appropriate in specialized addiction and psychiatric settings, especially with alcohol dependent patients, even if this remains questionable since Figlie and colleagues demonstrated a 2 factor solution in a mixed population of patients from specialized setting and from a gastroenterology clinic (Demmel et al., 2004; Figlie et al., 2005; Miller & Tonigan, 1996). On the other hand, the use of a 2 factor solution seems more appropriate for patients screened opportunistically in general health settings such as primary care clinics, community samples or hospitals (Maisto et al., 1999; Burrow-Sanchez & Lundberg, 2007). Among adolescents and young adults, published data are inconsistent in favor of one or the other structures (Maisto, Chung, Cornelius, & Martin, 2003; Vik, Culbertson, & Sellers, 2000). Contrary to the ambivalence and recognition constructs that the originally described Ambivalence and Recognition factors intended to capture, the concept captured in the factor called "Taking Steps" (measure of actions taken towards change, or change-related actions) in both the Maisto et al. and Miller and Tonigan studies is consistent across populations and settings.

The main strength of our study was the examination of the SOCRATES in a large sample of medical inpatients, adding to the literature on readiness to change in patients who are identified opportunistically, and are not seeking treatment (in contrast to studies in specialty addiction treatment settings). This is the first study to our knowledge investigating the factor structure in this population. Nevertheless, the generalizability of these results should be limited to hospitalized medical patients. The results are particularly applicable to those who agreed to participate in a clinical trial where they could receive alcohol counseling. It is possible that the subjects included in our sample were more motivated to change than were those who refused to participate. However, individuals who agreed to participate had similar readiness scores on a 1 to 10 visual analog scale compared to eligible subjects who refused participation.

In conclusion, our findings support the likelihood that the SOCRATES can assess and measure two important motivational constructs in patients identified by screening, who are not necessarily seeking nor receiving specialty alcohol treatment. One of these constructs, change-related actions, was consistently found across settings and populations. The first component identified in our sample (PP) reflects perception of problems and need for help, and the second taking action or change-related actions (TA). Nevertheless, identification of these two readiness-to-change constructs is of interest primarily as potential predictors of change or determinants of behavior change. The predictive validity of the 2 components and their relationship with behavior change need to be further explored. Since only about 5% of individuals with alcohol dependence seek and receive treatment, having tools that help researchers to better study the 95% who do not seek help is important and relevant, particularly when it is assumed that seeking treatment is related to motivation and problem recognition.

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. *Archives of Internal Medicine*, *165*(9), 986—995.

Bien, T. H., Miller, W. R., & Tonigan, J. S. (1993). Brief interventions for alcohol problems: A review. *Addiction*, *88*(3), 315—335.

Burrow-Sanchez, J. J., & Lundberg, K. J. (2007). Readiness to change in adults waiting for publicly funded substance abuse treatment. *Addictive Behaviors*, *32*(1), 199—204.

Carbonari, J. P., & DiClemente, C. C. (2000). Using transtheoretical model profiles to differentiate levels of alcohol abstinence success. *Journal of Consulting and Clinical Psychology*, *68*(5), 810—817.

Cureton, E. E., & D'Agostino, R. B. (1993). *Factor analysis: An applied approach*, (1st ed.): Lawrence Erlbaum.

Demmel, R., Beck, B., Richter, D., & Reker, T. (2004). Readiness to change in a clinical sample of problem drinkers: Relation to alcohol use, self-efficacy, and treatment outcome. *European Addiction Research*, *10*(3), 133—138.

Dunn, C., Deroo, L., & Rivara, F. P. (2001). The use of brief interventions adapted from motivational interviewing across behavioral domains: A systematic review. *Addiction*, *96*(12), 1725—1742.

Fiellin, D. A., Reid, M. C., & O'Connor, P. G. (2000). New therapies for alcohol problems: Application to primary care. *American Journal of Medicine*, *108*(3), 227—237.

Figlie, N. B., Dunn, J., & Laranjeira, R. (2005). Motivation for change in alcohol dependent outpatients from Brazil. *Addictive Behavior*, *30*(1), 159—165.

Group, P.M.R. (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*, *58*(1), 7—29.

Heather, N., Rollnick, S., & Bell, A. (1993). Predictive validity of the readiness to change questionnaire. *Addiction*, *88*(12), 1667—1677.

Maisto, S. A., Chung, T. A., Cornelius, J. R., & Martin, C. S. (2003). Factor structure of the SOCRATES in a clinical sample of adolescents. *Psychology of Addictive Behavior*, *17*(2), 98—107.

Maisto, S. A., Conigliaro, J., McNeil, M., Kraemer, K., O'Connor, M., & Kelley, M. E. (1999). Factor structure of the SOCRATES in a sample of primary care patients. *Addictive Behavior*, *24*(6), 879—892.

Miller, W. R., & Rollnick, S. (1991). Motivational interviewing. *Preparing people to change addictive behavior* New York: Guilford Press.

Miller, W. R., & Tonigan, J. S. (1996). Assessing drinkers' motivations for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). *Psychology of Addictive Behavior*, *10*, 81—89.

Miller, W. R., Tonigan, J., & Longabaugh, R. (1995). The drinker inventory of consequences (DrInC). *An instrument for assessing adverse consequences of alcohol abuse*, *vol. 4.* (pp.) Bethesda: NIH Test Manual.

Nunally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*, (3rd ed.): McGraw-Hill.

Prochaska, J. O., & DiClemente, C. C. (1984). *The transtheoretical approach.* Irwin, Illinois: Crossing Traditional Foundations of Change.

Saitz, R. (2005). Clinical practice. Unhealthy alcohol use. *New England Journal of Medicine*, *352*(6), 596—607.

Saitz, R., Palfai, T. P., Cheng, D. M., Horton, N. J., Freedner, N., Dukes, K., et al. (2007). Brief intervention for medical inpatients with unhealthy alcohol use: a randomized, controlled trial. *Annals of Internal Medicine*, *146*(3), 167—176.

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption—II. *Addiction*, *88*(6), 791—804.

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

Vik, P. W., Culbertson, K. A., & Sellers, K. (2000). Readiness to change drinking among heavy-drinking college students. *Journal of Studies on Alcohol*, *61*(5), 674—680.

WHO (1996). *Composite International Diagnostic Interview (CIDI) (Core Version 2.0).* Geneva, Swtizerland: World Health Organization.

Williams, E. C., Horton, N. J., Samet, J. H., & Saitz, R. (2007). Do brief measures of readiness to change predict alcohol consumption and consequences in primary care patients with unhealthy alcohol use? *Alcohol: Clinical and Experimental Research*, *31*(3), 428—435.

BMC Public Health

Research article

Improvements in readiness to change and drinking in primary care patients with unhealthy alcohol use: a prospective study

Nicolas Bertholet*[1,2], Nicholas J Horton[3] and Richard Saitz[1,4,5]

Address: [1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts, USA, [2]Clinical Epidemiology Center, Institute of Social and Preventive Medicine, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland, [3]Department of Mathematics and Statistics, Smith College, Northampton, Massachusetts, USA, [4]Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, Massachusetts, USA and [5]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA

Email: Nicolas Bertholet* - nberthol@bu.edu; Nicholas J Horton - nhorton@email.smith.edu; Richard Saitz - rsaitz@bu.edu

* Corresponding author

Published: 9 April 2009

BMC Public Health 2009, **9**:101 doi:10.1186/1471-2458-9-101

Received: 17 September 2008
Accepted: 9 April 2009

This article is available from: http://www.biomedcentral.com/1471-2458/9/101

Abstract

Background: The course of alcohol consumption and cognitive dimensions of behavior change (readiness to change, importance of changing and confidence in ability to change) in primary care patients are not well described. The objective of the study was to determine changes in readiness, importance and confidence after a primary care visit, and 6-month improvements in both drinking and cognitive dimensions of behavior change, in patients with unhealthy alcohol use.

Methods: Prospective cohort study of patients with unhealthy alcohol use visiting primary care physicians, with repeated assessments of readiness, importance, and confidence (visual analogue scale (VAS), score range 1–10 points). Improvements 6 months later were defined as no unhealthy alcohol use or any increase in readiness, importance, or confidence. Regression models accounted for clustering by physician and adjusted for demographics, alcohol consumption and related problems, and discussion with the physician about alcohol.

Results: From before to immediately after the primary care physician visit, patients (n = 173) had increases in readiness (mean +1.0 point), importance (+0.2), and confidence (+0.5) (all $p < 0.002$). In adjusted models, discussion with the physician about alcohol was associated with increased readiness (+0.8, $p = 0.04$). At 6 months, many participants had improvements in drinking or readiness (62%), drinking or importance (58%), or drinking or confidence (56%).

Conclusion: Readiness, importance and confidence improve in many patients with unhealthy alcohol use immediately after a primary care visit. Six months after a visit, most patients have improvements in either drinking or these cognitive dimensions of behavior change.

Background

Unhealthy alcohol use (the spectrum from at-risk drinking amounts through alcohol dependence) and its consequences represent a major burden of disease in the general population [1,2]. Among those with unhealthy alcohol use, brief intervention (BI) and motivational interviewing have demonstrated evidence of efficacy [3-5]. In primary care, BI is recommended by national practice guidelines (US Preventive Services Task Force, 2004), and, as part of BI, clinicians are encouraged to assess motivation and

readiness to change, and to help patients increase readiness [6]. These changes in readiness are seen as short term goals on the way to decreased consumption [7,8].

Processes of change have been conceptualized in various ways; in the Transtheoretical model, Prochaska and DiClemente described the progression of individuals through stages of change [9]. Others have pointed out the role of importance of change for the patient [10], while Bandura emphasized the role of self efficacy, or confidence in ability to change [11]. Importance and confidence are considered components of readiness or readiness-related factors. For the purpose of this article we will refer to readiness, importance and confidence as 3 behavior change constructs, cognitive dimensions of behavior change, that comprise "readiness to change."

These three constructs can facilitate clinical conversations about health behaviors, and may even have predictive value for later behavior change [12-14]. They can be assessed with visual analog scales (VAS), which, because of their brevity, can be widely used. Longer assessments such as the Readiness to Change Questionnaire (RTCQ) (developed to assess readiness per se, not importance or confidence specifically) allow classification according to stage of readiness to change in addition to the computation of a continuous score [15].

The predictive value of cognitive dimensions of behavior change, however, has been mixed. Demmel et al. demonstrated that in alcohol dependent inpatients, readiness accounted for 9.4% of the variance in outcome [12]. Self-efficacy also appears to be a predictor of abstinence [16]. In other studies, however, readiness was not associated with subsequent consumption, and was predictive of more, not fewer consequences (or at least greater recognition of consequences) [17,18].

Thus, how behavior change constructs relate to later behavior change is not well understood. Furthermore, how these constructs change over time and in response to brief counseling, particularly in general healthcare settings, is also not well known. The efficacy for BI in primary care for unhealthy alcohol use is modest, and interventions with different theoretical rationales (e.g. MATCH and COMBINE studies) lead to similar results [10,19,20], calling into question the roles played by behavior change versus other constructs. These concerns are also encountered more broadly in psychotherapy research, and were named by Rosenzweig in 1936 after Alice in Wonderland as the "Dodo Bird effect" [21]. Rosenzweig proposed that "common" factors were responsible for the efficacy of psychotherapy. He used Lewis Carroll's Alice in Wonderland's Dodo bird conclusion "everybody has won and all must have prizes." In most trials of alcohol BI, control (and intervention) groups decrease consumption over time (one way to understand natural history in these settings). This improvement has a number of possible explanations: regression to the mean, effects of assessments or contact with study staff, natural history, and selection of individuals more prone to change (since agreement to participate might indicate some desire to change) [22]. A better understanding of behavior change constructs, the natural history of behavior change, and how the constructs relate to outcome might therefore inform the design of more efficacious interventions that achieve and sustain changes in drinking [23]. Therefore we studied improvements and predictors of improvements in 3 behavior change constructs (readiness, importance, confidence) and subsequent drinking after a primary care visit in a prospective cohort of patients.

Methods

We studied a prospective cohort of adults with unhealthy alcohol use visiting an urban academic primary care practice who participated in a randomized trial of the impact of providing (or not providing) primary care physicians with patients' alcohol screening results [24]; physicians were the unit of randomization. Additional detail regarding the clinical trial has been published [24]. Patients were screened if they had a visit with an included physician. A trained staff researcher approached and interviewed enrolled subjects before and after their visits with a physician (February 1998–August 1999) in the waiting room. Patients were told they were being asked (initial screening) questions for research purposes. Eligible patients were told they were being asked to participate because they were primary care patients who reported drinking alcohol, and that the study might help physicians learn how to identify alcohol use by patients. Six months later, subjects were interviewed by telephone, after which they received compensation in the form of a voucher worth ten U.S. dollars. The screening was done using the CAGE questionnaire and the 3 questions published in the 1995 NIAAA guide to assess quantity and frequency of drinking [25,26]. The CAGE and the 3 alcohol questions were completed by paper and pencil in the waiting room or by help of a research assistant when needed. The baseline assessment was done by face to face interview. Inclusion for the present study was based on drinking as assessed using the Timeline Followback method [27]. Subjects were asked how many drinks they had each day for the past 30 days using a calendar and chronologic cues (e.g. weekends, holidays, significant events during the time period) based on published instructions for this assessment.

For the present study, eligible patients drank risky amounts (> 14 U.S. standard drinks [12 g each]/week or > 4 drinks/occasion for men and > 7 drinks/week or > 3 drinks/occasion for women in the past month), were flu-

ent in English or Spanish, and had data available on 3 behavior change constructs of interest. Subjects were interviewed before, immediately after the visit and 6 months later. At each time point, subjects completed 3 visual analog scales (VASs), one for each of 3 behavior constructs ("how ready are you to change your drinking habits", "how important is it for you right now to change your drinking", "if you decide to change your drinking habits, how confident are you that you would succeed", with a score of 1 being "not ready/not important to change/not confident to succeed" and 10 "ready/very important to change/very confident to succeed") [17,28]. Subjects also completed the RTCQ, a validated instrument that assesses readiness that has satisfactory test-retest reliability and consists of 12 statements, each one evaluated by the participant on a 5 point Likert scale [29].

For the present study, we used a continuous RTCQ score (-24 to +24) that has good reliability (alpha = 0.85–0.86)[30,31]. The readiness item was included in the screener (paper and pencil). Confidence and importance items and the RTCQ were included in the previsit interview (face to face interview). After the visit and 6 months later, subjects completed the 30-day Timeline Followback, a validated calendar method considered a reference standard for assessing alcohol consumption,[32] and the Short Inventory of Problems (SIP) to assess alcohol-related problems [33]. Immediately after the visit, subjects were asked by research assistants if they had had any discussion with their physician about alcohol consumption as well as the content of such a discussion (i.e. "Did the doctor give you advice about your drinking habits," "Did the doctor talk about drinking?", "Did the doctor tell you how many drinks would be safe for you to drink?", "Did the doctor recommend that you cut down/quit drinking, go to Alcoholics Anonymous/treatment?"). Research assistants also assessed demographics, social support ("Do you currently have a partner?"), and illicit drug use over the past month ("In the last 30 days have you used marijuana/cocaine/heroin/other illegal drug?"). The study was approved by the Boston University Medical Center Institutional Review Board and a Certificate of Confidentiality was obtained from the U.S. government.

Analyses were performed using SAS software 9.1.3 (Cary, North Carolina). P values less than 0.05 were considered to be statistically significant. All analyses controlled for clustering of subjects within physician and physician randomization group, using Generalized Estimating Equations (GEE) regression models with an exchangeable working correlation and empirical variance estimator. Of note, in the original trial, the intervention (i.e. randomization group) was not associated with improvements in drinking risky amounts.

We assessed the outcomes changes in readiness, importance, and confidence between pre- and immediate post-visit assessments. Differences were computed by subtracting the pre-visit score from the immediate post-visit score. We used unadjusted models (but accounting for clustering) and models that adjusted for predictors to assess the changes between pre- and immediate post-visit assessments in each of the behavior change constructs.

Because a change in behavior could take place during the 6-month follow-up period after the pre-visit assessment, the state of readiness to change drinking at 6 months could become irrelevant (e.g. readiness to change in a subject who no longer consumed alcohol). Therefore, to assess change outcomes, we created 3 dichotomous variables representing "improvement," each defined as either no longer drinking risky amounts or, if still drinking risky amounts, having improvement (difference > 0) in readiness to change drinking (i.e. improvement in drinking *or* readiness), importance of changing drinking (i.e. improvement in drinking *or* importance), or confidence in ability to change drinking (i.e. improvement in drinking *or* confidence). Since readiness, importance and confidence are often considered intermediate outcomes, it is of clinical relevance to use an outcome taking into account increases in these readiness to change constructs [34,35]. We did not analyze these behavior change constructs in those who had already changed, and we did not have data on readiness/importance/confidence to sustain changes in drinking (e.g. not drinking risky amounts), and therefore could not analyze these constructs as outcomes.

Predictors of interest were: demographics (age, sex, race/ethnicity), social support (currently having a partner), alcohol consumption at study entry (drinks per day), SIP score, illicit drug use (marijuana, heroin, cocaine, other), and discussion with the physician about alcohol consumption during the visit.

All analyses were repeated with the RTCQ continuous score as the outcome, in order to corroborate the results obtained with the readiness VAS.

Results

Of 4143 patients approached, 182 did not complete the screener. Of 487 who reported drinking risky amounts in the past month, 235 refused participation in the study and 18 had no time before the visit to complete the pre-visit assessment. Of the 234 remaining enrolled patients, our analytic sample was restricted to the 173 subjects drinking risky amounts who had available data on "readiness to change" at all 3 assessments (Table 1); they saw one of 36 physicians. We tested potential differences between included subjects who completed and did not complete the follow up: There were no significant differences (p <

Table 1: Characteristics of 173 Adults with Unhealthy Alcohol Use Visiting a Primary Care Physician

Characteristic	%(n)
Female	42% (72)
Race/ethnicity	
African-American	58% (101)
White	18% (31)
Latino	16% (28)
Other	8% (13)
Employed (last 3 month)	60%(103)
Having a partner	68%(118)
Illicit drug use, last 30 days (any drug)	34% (58)
Patient's physician randomized to receive intervention (provision of screening results)	58% (100)
Discussion with physician about alcohol consumption	55% (95)
	mean, (SD)
Age	43.10 (12.61)
Drinks per day	3.05 (4.76)
Alcohol related consequences	8.41 (10.36)
Readiness to change drinking measures (pre visit)	
Readiness (1 to 10)	5.04 (3.13)
Importance (1 to 10)	6.01 (3.56)
Confidence (1 to 10)	7.75 (2.60)
RTCQ (-24 to 24)	3.32 (6.76)

Alcohol related consequences: Short Inventory of Problems score
RTCQ: Readiness To Change Questionnaire

0.05) between study subjects (n = 173) and those enrolled but not included in the analytic sample (n = 61) on pre-visit readiness, importance, confidence, discussion with the physician about alcohol consumption, provision of screening results to the physician, drinks per day, age, race, employment status, having a partner or drug use. There was a significant difference between study subjects and those enrolled but not included in the analytic sample on gender: subjects not included (i.e. lost to follow up) were more likely to be male (74% vs 58%, p = 0.05).

Regarding generalizability, we compared included subjects to all other individuals identified by screening who were eligible. Subjects included in the analytic sample drank more (mean days drinking per week 2.98 vs 2.49, p = 0.01, mean number of drinks per typical day of consumption 4.45 vs 3.44, p < .0001). There were no significant differences (p < 0.05) between subjects included in the analytic sample and all individuals identified by screening on gender, race and readiness to change.

Pre- to immediate post-visit changes

In unadjusted analyses of visual analog scales (VASs) (score range 1 to 10 points), subjects had significant increases immediately post-visit in readiness (+1.02 points, p < 0.0001), importance (+0.16, p < 0.0001), and confidence (+0.49, p = 0.003) when compared to pre-visit assessments (Table 2). Similarly, readiness as assessed by the RTCQ also increased (mean change +0.14, p < 0.0001). All of the differences indicated increases in readiness, importance, and confidence.

In adjusted models (Table 3), a discussion with the physician about the patient's alcohol consumption was a significant positive predictor of an increase in readiness as measured by VAS (adjusted mean change +0.78 points, p = 0.04), as was not having a partner (+1.06, p = 0.006).

Table 2: Changes (from Before to) Immediately After a Primary Care Physician Visit in Patient "Readiness" to Change Drinking

Measure (range)	Mean Change	(95% CI)	Effect size (d)
Readiness (1–10)	1.02	(0.74, 1.30)	0.33
Importance (1–10)	0.16	(0.15, 0.18)	0.05
Confidence (1–10)	0.49	(0.16, 0.81)	0.19
RTCQ (-24 - +24)	0.14	(0.13, 0.15)	0.02

RTCQ: Readiness To Change Questionnaire
CI = confidence interval

Table 3: Predictors of Improvement Immediately After a Primary Care Physician Visit

Improvement in	Readiness (1–10)		Importance (1–10)		Confidence (1–10)		RTCQ (-24- +24)	
	Mean adjusted change	(95% CI)	Mean adjusted change	(95% CI)	Mean adjusted change	(95% CI)	Mean adjusted change	(95% CI)
Female	-0.37	(-1.19, 0.46)	0.21	(-0.34, 0.77)	-0.32	(-0.98, 0.34)	0.57	(-0.26, 1.41)
Age	0.01	(-0.02, 0.04)	-0.01	(-0.03, 0.02)	0.00	(-0.03, 0.03)	0.01	(-0.03, 0.04)
White	**-1.41**	(-2.31, -0.51)	-0.09	(-1.06, 0.89)	-0.36	(-0.86, 0.14)	**1.24**	(0.10, 2.39)
No Partner	**1.06**	(0.30, 1.81)	-0.05	(-0.65, 0.55)	-0.21	(-0.79, 0.37)	0.09	(-1.07, 1.25)
Drinks per day	0.02	(-0.08, 0.12)	0.04	(-0.10, 0.18)	-0.01	(-0.06, 0.05)	-0.03	(-0.18, 0.11)
Alcohol consequences (SIP score)	-0.03	(-0.06, 0.00)	-0.01	(-0.04, 0.01)	0.01	(-0.03, 0.05)	0.03	(-0.02, 0.07)
Discussion with physician	**0.78**	(0.05, 1.51)	0.16	(-0.46, 0.79)	-0.07	(-0.60, 0.46)	**1.14**	(0.04, 2.25)
Illicit drug use	-0.03	(-0.75, 0.68)	-0.51	(-1.27, 0.24)	0.07	(-0.52, 0.66)	0.87	(-0.41, 2.16)

SIP: Short Inventory of Problems
RTCQ: Readiness To Change Questionnaire
Bold entries have p-values < 0.05

Being white was a negative predictor of change in readiness (white versus other: -1.41, p = 0.002). Age, gender, alcohol consumption and alcohol-related problems were not significant predictors. Similarly for readiness measured by the RTCQ, discussion with the physician about the patient's alcohol consumption was a significant predictor of an increase (+1.14 points, p = 0.04). However, not having a partner was not significantly associated with an increase in RTCQ-measured readiness, and being white was associated with an increase (not a decrease) (+1.24, p = 0.03). None of the predictors were significantly associated with increases in importance and confidence.

Pre-visit to 6-month follow up changes

Of the 173 subjects, 62 (36%) reported not drinking risky amounts during 30 days prior to the 6-month follow-up interview. But most subjects had improvements in drinking or readiness, importance and confidence: 62% were no longer drinking risky amounts or had improved readiness, 58% were not drinking risky amounts or had improved importance, and 56% were not drinking risky amounts or had improved confidence. Similarly, 67% of subjects were not drinking risky amounts or had improved RTCQ-readiness. We assessed the proportion of subjects who had improved because of a change in drinking. Of the 98 subjects who had improved drinking or readiness, 54 (55%) were no longer drinking risky amounts and 44 (45%) were still drinking risky amounts but had improved readiness. For improvement in drinking or importance, 62 (62%) had improved drinking, 38 (38%) improved importance; for improvement in drinking or confidence, 62 (64%) improved drinking, 35 (36%) confidence. Results were similar for improvement in drinking or RTCQ-readiness (62 (54%) had improved drinking, 53 (46%) improved RTCQ score.

In adjusted analyses, few predictors of improvement were identified. Not having a partner was a significant predictor of improvement in drinking or importance (p = 0.007), and drinking or confidence (p = 0.005). Being white was a negative predictor of improvement in drinking or readiness (p = 0.007). No other predictors were significantly associated with improvements in the hypothesized direction. Unexpectedly, having a discussion with the physician (at the initial visit) was a negative predictor (p = 0.03) of improvement in drinking or importance (Table 4).

Discussion

We assessed variations in readiness, importance and confidence regarding changing drinking after a single primary care physician visit and improvements in these constructs and drinking 6 months later. After the visit, we observed significant increases in readiness, importance, and confidence. The effects were small (i.e. 1 point for readiness, 0.16 for importance, 0.49 for confidence, on a 1 to 10 scale). However, a clinically significant change in these constructs has not yet been well-defined, and the impact of changes of any magnitude is not known. Based on the transtheoretical model and motivational interviewing, clinicians are encouraged to help patients increase motivation, which in turn is expected to lead eventually to behavior change [10,35]. After a physician visit we can detect the beginning of such changes. Having a discussion about alcohol with the physician appeared to have an additional impact on readiness to change, an effect that was no longer detectable 6 months later. Like other meas-

Table 4: Predictors of Improvement 6 Months After a Primary Care Physician Visit

Improvement in drinking or...	Readiness (1–10)		Importance (1–10)		Confidence (1–10)		RTCQ (-24- +24)	
	OR	**(95% CI)**	**OR**	**(95% CI)**	**OR**	**(95% CI)**	**OR**	**(95% CI)**
Female	0.63	(0.30, 1.30)	0.60	(0.31, 1.18)	0.66	(0.32, 1.38)	0.91	(0.47, 1.73)
Age	1.02	(0.99, 1.05)	1.00	(0.97, 1.03)	1.01	(0.98, 1.04)	1.00	(0.98, 1.03)
White	**0.25**	(0.09, 0.69)	0.66	(0.29, 1.47)	0.47	(0.21, 1.03)	0.85	(0.36, 1.99)
No Partner	0.80	(0.41, 1.54)	**0.42**	(0.23, 0.79)	**0.44**	(0.25, 0.78)	0.30	(0.34, 1.99)
Drinks per day	0.98	(0.90, 1.07)	0.99	(0.93, 1.06)	0.98	(0.91, 1.06)	0.96	(0.89, 1.03)
Alcohol consequences (SIP score)	0.98	(0.95, 1.02)	1.00	(0.96, 1.04)	1.02	(0.98, 1.06)	0.99	(0.96, 1.03)
Discussion with physician	1.06	(0.58, 1.95)	**0.56**	(0.33, 0.95)	0.62	(0.27, 1.41)	0.85	(0.42, 1.73)
Illicit drug use	1.45	(0.65, 3.22)	0.71	(0.31, 1.63)	0.96	(0.42, 2.19)	1.61	(0.63, 4.09)

OR = odds ratio
SIP: Short Inventory of Problems
RTCQ: Readiness To Change Questionnaire
Bold entries have p-values < 0.05

ures of health states (e.g. blood pressure), "readiness to change" should be viewed as an instantaneous measure dependent on various internal and external influences.

In addition to these short-term changes, we studied the course of risky drinking and "readiness to change" in primary care patients. Six months after a physician visit, most subjects improved either their drinking or readiness. These improvements suggest that primary care physicians should be somewhat optimistic regarding the course of unhealthy alcohol use, with more than half of patients improving in a relatively short period of time.

We identified few predictors of changes in "readiness to change" and none that were consistent across measures or time. Not having a partner was a positive predictor of immediate changes in readiness but a negative predictor of improvement in two measures 6 months later. Being white was associated with worse readiness immediately after a physician visit and less improvement 6 months later but the finding was not confirmed in analyses with the RTCQ or the other two behavior change construct measures. Speculation regarding the mechanism for these hypothesis generating and inconsistent findings would be premature.

A discussion between the physician and the subject about alcohol predicted a positive change in readiness immediately after the visit (confirmed by the RTCQ), but paradoxically, was associated with less improvement in drinking or importance (but no change in VAS- or RTCQ-measured readiness). The fact that no association was found between discussion with the physician and drinking 6 months later may have been due to the use of ineffective counseling, but given the observed short-term effect, another explanation could be simply that this effect did not last.

Neither alcohol consumption, alcohol problems, nor illicit drug use significantly affected behavior change constructs or improvements. The fact that these markers of severity were not found to be negative predictors of improvement is of interest and should encourage physicians to address problems related to alcohol consumption even in the presence of concomitant illicit drug use, considering that most of their patients will have some improvement, independent of the severity of the alcohol problem.

A number of studies have assessed readiness to change and related constructs. In general, these studies have focused on characterizing specific populations [13,36-39] or on studying readiness as a predictor of behavior change [12,14]. Our study instead focused on how these constructs change over time. Improvement over time in untreated adults with unhealthy alcohol use and alcohol use disorders (alcohol abuse and dependence) has been previously reported [22,40,41]. Alcohol abuse and alcohol dependence seem to be (especially the latter) chronic conditions characterized by recurrent episodes of disease activity [42]. But the natural history of the spectrum of unhealthy alcohol use (risky drinking amounts through dependence) is not well described in the literature, nor is the natural history of readiness to change.

In addition to the aforementioned studies of the natural history of alcohol use disorders, studies of brief intervention for nondependent unhealthy alcohol use in primary care consistently report improvement over time in both treated and untreated individuals [43]. For example, male heavy drinkers in primary care decreased drinking over 3 years by 25 to 53% (depending on the outcome measure) in both intervention and control groups [44]. The improvement in drinking observed in our sample is in this range. Improvements such as these could be attributed to

a regression to the mean [22]. However, our study sample was not primarily composed of very heavy drinkers, and we also observed improvements in readiness, importance and confidence regarding changing drinking, which were in the opposite direction than any hypothesized regression to the mean, given the relatively high levels of readiness, importance and confidence in the study population at baseline. Assessment effects (improvements due to being asked questions about drinking and discussing answers to those questions) have also been suggested as causes of improvements in drinking [45,46]. This exposure may have in part accounted for improvements in our sample. But if asking about alcohol and discussing drinking in primary care are in fact responsible for improvements, such effects should be viewed as favorable exposures in the primary care setting, and as part of the course in these patients, rather than as methodological nuisances.

This study has some strengths. To our knowledge, this is the first study to explore changes in readiness, importance and confidence during a single primary care visit. We described rapid changes in these constructs. We were also able to describe changes in readiness, importance or confidence and drinking over a 6 month period. Subjects studied were participants in a trial but there was no experimental brief counseling intervention nor a significant treatment effect on drinking amounts.

The study also has some limitations. First, the applicability of our findings may be limited to primary care patients with unhealthy alcohol use who agree to be screened and followed in a research study, and to those with similar characteristics as in our sample (e.g. 32% reporting no alcohol problems, a third with illicit drug use). Although participants differed little from those who did not participate, participants did drink more. Similarly, subjects lost to the analytic sample differed little from those studied except on gender. The effects of these differences can be considered in the interpretation of our results. Second, findings are from secondary data analyses. Causality (of predictors) cannot be inferred, and there could be many explanations for changes in the readiness constructs. However, observational studies such as this one are likely among the best ways to study the natural course of behavior change constructs and changes in drinking, particularly prospective studies. Third, in our attempt to explore changes over time, we had to combine actual behavior (changes in alcohol consumption) with cognitions about behavior change. It is likely that these two dimensions reflect different aspects of behavior. We presented data on the dimensions separately (and combined) but performed regression analyses on the combined outcome. From a clinical perspective, improvements in either drinking or readiness (combined) seem to be most relevant. Also, the interpretation of a 6 month change in readiness on a continuous scale for someone who continues to drink risky amounts is difficult, since it not clear how one would interpret a change in some number of points. Fourth, we did not adjust the level of significance for multiple comparisons. As such one should be cautious about interpreting associations, particularly those not in the hypothesized direction. Lastly, all data were obtained from interviews and are subject to recall and social desirability biases. But interviews with trained research associates and assurances of confidentiality took place immediately after the primary care visit to maximize accurate recall and minimize bias. Nonetheless, we do not know what patients meant by discussions about alcohol, which could have been brief or extensive and may or may not have included known effective components of brief interventions.

Conclusion

Our results provide important information. First, subjects appear to change readiness, importance and confidence after a single physician visit. Second, most patients with unhealthy alcohol use will improve 6 months after a primary care visit, either on behavioral change constructs or drinking. Third, factors usually associated with worse alcohol treatment outcomes (e.g. drug use, alcohol related problems) do not seem to prevent improvement. Future research should focus on specific measures of behavior change constructs, perhaps assessed by ecological momentary assessments of these rapidly changing dimensions [47,48] and what contributes to their changing, and how and when they contribute to actual behavior change. A better understanding of these mechanisms could be used to enhance interventions for unhealthy alcohol use in primary care settings, which at present, are only modestly effective. But even if the reasons for the improvements in drinking and behavior change constructs are mostly unknown, primary care physicians should be aware of the prognosis for patients with unhealthy alcohol use, to recognize incremental steps towards change and to support their patient's efforts.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NB, NH, and RS conceived and designed the study. RS led implementation of the SIP study randomized trial. NH led and programmed the statistical analyses. NB wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Nicolas Bertholet was supported by the Swiss National Science Foundation, and Fondation Suisse de Recherche sur l'Alcool. The Screening and Intervention in Primary Care (SIP) study was supported by a Robert Wood

Johnson Foundation Generalist Faculty Physician Scholar Award to Dr. Saitz and Boston University.

References

1. Saitz R: **Clinical practice. Unhealthy alcohol use.** *N Engl J Med* 2005, **352(6):**596-607.
2. Rehm J, Room R, Monteiro M, *et al.*: **Alcohol as a risk factor for global burden of disease.** *Eur Addict Res* 2003, **9(4):**157-64.
3. 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(9):**986-95.
4. Bien TH, Miller WR, Tonigan JS: **Brief interventions for alcohol problems: a review.** *Addiction* 1993, **88(3):**315-35.
5. Dunn CW, Ries R: **Linking substance abuse services with general medical care: integrated, brief interventions with hospitalized patients.** *Am J Drug Alcohol Abuse* 1997, **23(1):**1-13.
6. Rollnick S: **Readiness, Importance, and Confidence.** In *Treating Addictive Behaviors* 2nd edition. Edited by: Miller WB, Heather N. New York: Plenum Press; 1998:49-60.
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. Barnes HN, Samet JH: **Brief interventions with substance-abusing patients.** *Med Clin North Am* 1997, **81(4):**867-79.
9. Prochaska JO, DiClemente CC, Norcross JC: **In search of how people change: Applications to addictive behaviors.** *Am Psychol* 1992, **47(9):**1102-14.
10. Miller WR, Wilbourne PL: **Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders.** *Addiction* 2002, **97(3):**265-77.
11. Bandura A: **Social Learning Theory.** Oxford, England: Prentice-Hall; 1977.
12. Demmel R, Beck B, Richter D, Reker T: **Readiness to change in a clinical sample of problem drinkers: relation to alcohol use, self-efficacy, and treatment outcome.** *Eur Addict Res* 2004, **10(3):**133-8.
13. Williams EC, Kivlahan DR, Saitz R, *et al.*: **Readiness to change in primary care patients who screened positive for alcohol misuse.** *Ann Fam Med* 2006, **4(3):**213-20.
14. Maisto SA, Conigliaro J, McNeil M, Kraemer K, O'Connor M, Kelley ME: **Factor structure of the SOCRATES in a sample of primary care patients.** *Addict Behav* 1999, **24(6):**879-92.
15. Heather N, Rollnick S, Bell A: **Predictive validity of the Readiness to Change Questionnaire.** *Addiction* 1993, **88(12):**1667-77.
16. Maisto SA, Carey KB, Bradizza CM: **Social learning theory.** In *Psychological Theories of Drinking and Alcoholism* Edited by: Leonard KE, Blane HT. New York: Guilford Press; 1999:106-63.
17. Williams EC, Horton NJ, Samet JH, Saitz R: **Do brief measures of readiness to change predict alcohol consumption and consequences in primary care patients with unhealthy alcohol use?** *Alcohol Clin Exp Res* 2007, **31(3):**428-35.
18. Forsberg L, Ekman S, Halldin J, Ronnberg S: **The readiness to change questionnaire: reliability and validity of a Swedish version and a comparison of scoring methods.** *Br J Health Psychol* 2004, **9(Pt 3):**335-46.
19. Project MATCH Research Group: **Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH posttreatment drinking outcomes.** *J Stud Alcohol* 1997, **58(1):**7-29.
20. Anton RF, O'Malley SS, Ciraulo DA, *et al.*: **Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial.** *Jama* 2006, **295(17):**2003-17.
21. Wampold BE, Mondin GW, Moody M, Stich F, Benson K, Ahn H: **A Meta-Analysis of Outcome Studies Comparing Bona Fide Psychotherapies: Empirically, "All Must Have Prizes".** *Psychological Bulletin* 1997, **122(3):**203-15.
22. Moyer A, Finney JW: **Outcomes for untreated individuals involved in randomized trials of alcohol treatment.** *J Subst Abuse Treat* 2002, **23(3):**247-52.
23. Saitz R, Svikis D, D'Onofrio G, Kraemer KL, Perl H: **Challenges applying alcohol brief intervention in diverse practice settings: populations, outcomes, and costs.** *Alcohol Clin Exp Res* 2006, **30(2):**332-8.
24. Saitz R, Horton NJ, Sullivan LM, Moskowitz MA, Samet JH: **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(5):**372-82.
25. Mayfield D, McLeod G, Hall P: **The CAGE questionnaire: validation of a new alcoholism screening instrument.** *Am J Psychiatry* 1974, **131(10):**1121-3.
26. National Institute on Alcohol Abuse and Alcoholism: **The Physicians' Guide to Helping Patients with Alcohol Problems.** Bethesda: National Institutes of Health publication; 1995. Report No.: 95–3769.
27. Sobell LC, Sobell MB: **Timeline follow-back: a technique for assessing self reported alcohol consumption.** In *Measuring Alcohol Consumption: Psychosocial and Biochemical Methods* Edited by: Litten RZAJ. Totowa, NJ: Humana; 1992:41-69.
28. Miller WR, Rollnick S: *Motivational Interviewing: Preparing People for Change* 2nd edition. New York, NY: The Guilford Press; 2002.
29. Rollnick S, Heather N, Gold R, Hall W: **Development of a short 'readiness to change' questionnaire for use in brief, opportunistic interventions among excessive drinkers.** *Br J Addict* 1992, **87(5):**743-54.
30. Budd R, Rollnick S: **The structure of the readiness to change questionnaire: a test of Prochaska & DiClemente's transtheoretical model.** *British Journal of Health Psychology* 1996, **1:**365-76.
31. McNally AM, Palfai TP: **Negative emotional expectancies and readiness to change among college student binge drinkers.** *Addict Behav* 2001, **26(5):**721-34.
32. Sobell LC, Brown J, Leo GI, Sobell MB: **The reliability of the Alcohol Timeline Followback when administered by telephone and by computer.** *Drug Alcohol Depend* 1996, **42(1):**49-54.
33. Miller WR, Tonigan J, Longabaugh R: **The Drinker Inventory of Consequences (DrInC).** In *An Instrument for Assessing Adverse Consequences of Alcohol Abuse. Test Manual* Bethesda: National Institutes of Health publication; 1995.
34. 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(3):**233-43.
35. DiClemente CC, Schlundt D, Gemmell L: **Readiness and stages of change in addiction treatment.** *Am J Addict* 2004, **13(2):**103-19.
36. Vik PW, Culbertson KA, Sellers K: **Readiness to change drinking among heavy-drinking college students.** *J Stud Alcohol* 2000, **61(5):**674-80.
37. Apodaca TR, Schermer CR: **Readiness to change alcohol use after trauma.** *J Trauma* 2003, **54(5):**990-4.
38. Bombardier CH, Ehde D, Kilmer J: **Readiness to change alcohol drinking habits after traumatic brain injury.** *Arch Phys Med Rehabil* 1997, **78(6):**592-6.
39. Rumpf HJ, Hapke U, Meyer C, John U: **Motivation to change drinking behavior: comparison of alcohol-dependent individuals in a general hospital and a general population sample.** *Gen Hosp Psychiatry* 1999, **21(5):**348-53.
40. Sobell LC, Ellingstad TP, Sobell MB: **Natural recovery from alcohol and drug problems: methodological review of the research with suggestions for future directions.** *Addiction* 2000, **95(5):**749-64.
41. Bischof G, Rumpf HJ, Hapke U, Meyer C, John U: **Types of natural recovery from alcohol dependence: a cluster analytic approach.** *Addiction* 2003, **98(12):**1737-46.
42. Venner KL, Matzger H, Forcehimes AA, *et al.*: **Course of recovery from alcoholism.** *Alcohol Clin Exp Res* 2006, **30(6):**1079-90.
43. Bertholet N, Schwan R, Duhamel O, Perney P, Daeppen JB: **Efficacité de l'intervention brève.** *Alcoologie et Addictologie* 2003, **25(45):**59S-63S.
44. Aalto M, Sillanaukee P: **Compliance rate and associated factors for entering an alcohol brief intervention treatment programme.** *Alcohol Alcohol* 2000, **35(4):**372-6.
45. Clifford PR, Maisto SA, Davis CM: **Alcohol treatment research assessment exposure subject reactivity effects: part I. Alcohol use and related consequences.** *J Stud Alcohol Drugs* 2007, **68(4):**519-28.
46. Maisto SA, Clifford PR, Davis CM: **Alcohol treatment research assessment exposure subject reactivity effects: part II. Treatment engagement and involvement.** *J Stud Alcohol Drugs* 2007, **68(4):**529-33.

47. Shiffman S, Stone AA, Hufford MR: **Ecological momentary assessment.** *Annu Rev Clin Psychol* 2008, **4:**1-32.
48. Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M: **Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers.** *Psychol Addict Behav* 2002, **16(3):**205-11.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2458/9/101/prepub

The Influence of Black Race on Treatment and Mortality for Early-Stage Breast Cancer

Jonathan P. B. Berz, MD, Katherine Johnston, MD, MA, MSc,† Bertina Backus, MPH,‡ Gheorghe Doros, PhD,§ Adam J. Rose, MD, MSc,*¶ Snaltze Pierre, MPH,‡ and Tracy A. Battaglia, MD, MPH*‖*

Background: Black Americans have higher mortality from breast cancer than white Americans. This study explores the influence of socioeconomic factors and black race on treatment and mortality for early-stage breast cancer.

Methods: A cohort of 21,848 female black and white, non-Hispanic subjects from the Massachusetts Cancer Registry diagnosed with stage I or II breast cancer between 1999–2004 was studied. Subjects with tumors larger than 5 cm were excluded. We used mixed modeling methods to assess the impact of race on guideline concordant care (GCC), defined as receipt of mastectomy or breast conserving surgery plus radiation. Cox proportional hazard regression was used to assess disease-specific mortality.

Results: Blacks were less likely to receive GCC after adjusting for age and clinical variables (OR: 0.75; 95% CI: 0.61, 0.92). Marital status and insurance were predictors of receipt of GCC. After adjustment for all covariates, there were no longer significant differences between black and white women regarding the receipt of GCC. Nevertheless, black women were more likely to die of early-stage breast cancer than white women after adjusting for clinical, treatment, socioeconomic variables, and reporting hospital (HR: 1.6; 95% CI: 1.1–2.1).

Conclusions: Socioeconomic factors are mediators of racial differences in treatment outcomes. Significant racial differences exist in disease-specific mortality for women with early-stage breast cancer. Attention to reducing socioeconomic barriers to care may influence racial differences in breast cancer treatment and mortality.

Key Words: African Americans, breast neoplasms, health inequities, survival analysis, health services

(*Med Care* 2009;47: 986–992)

Breast cancer is the second leading cause of cancer death among US women, and these deaths disproportionally affect black Americans. Recent statistics show that although blacks have a lower incidence rate of breast cancer than whites, they have the highest mortality rate of any racial or ethnic group.[1] Furthermore, despite an overall decrease in breast cancer mortality since the early 1990s as a result of improvements in both early detection and treatment, blacks have benefited the least compared with whites with respect to this decrease.[2]

Various factors have been implicated in this mortality gap including tumor biology and delays in or lack of receipt of recommended screening and therapies.[3] In addition, while it is well documented that blacks present at a later stage of disease, they continue to have higher mortality even when controlling for stage at diagnosis.[1,4] Therefore, achieving equal receipt of therapy for early-stage disease, where recommended treatment guidelines are clear and relatively uniform, may contribute to closing the mortality gap.

Receipt of either mastectomy or breast conserving surgery (BCS) with radiation has been shown to be equivalent with respect to breast cancer survival in early-stage disease, though the latter treatment may be preferred due to the less invasive nature of the procedure and the advantage of breast preservation.[5] Among those who receive BCS, radiation is an important part of treatment to decrease recurrence and improve survival.[6,7] Many studies have shown that blacks are similarly likely to receive BCS but less likely to receive radiation, compared with whites.[8–10] Only some of these studies include socioeconomic factors, such as income or insurance, and none use reporting hospital as a variable to explain differences when they do exist.[8–14]

The goal of this study was to use recent cancer registry data to understand the role that race and certain socioeconomic factors may play in treatment received and in breast

From the *Section of General Internal Medicine, Boston University Medical Center, Boston, Massachusetts; †Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts; ‡Massachusetts Department of Public Health, Boston, Massachusetts; §Boston University School of Public Health, Boston, Massachusetts; ¶Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Bedford, Massachusetts; and ‖Women's Health Unit, Section of General Internal Medicine, Boston University Medical Center, Boston, Massachusetts.

Supported by the Physician Training Award in Preventive Medicine American Cancer Society (PTAPM-97-185-04), and also supported by Career Development Award from the Veterans Affairs Health Services Research and Development Service (to A.J.R.).

The opinions expressed in this article are those of the authors and do not necessarily represent the official views of the Department of Veterans Affairs.

Reprints: Jonathan PB Berz, MD, Section of General Internal Medicine, Boston University Medical Center, Crosstown Building, 2nd Floor, 801 Massachusetts Ave., Boston, MA 02118. E-mail: jonathan.berz@bmc.org.

ISSN: 0025-7079/09/4709-0986

cancer mortality. We used the Massachusetts Cancer Registry (MCR) to explore these associations in women diagnosed with stages I and II breast cancer during the years 1999 through 2004.

METHODS

Database and Study Population

Data were obtained from the MCR, a population based cancer registry that collects information on newly diagnosed cancers among residents of the state. The criteria for case selection of female invasive breast cancer cases were defined using International Classification of Diseases for Oncology Third Edition codes (ICD-O-3). The codes used were primary site 50.0 to 50.9, and all histologies except 9590 to 9989, for the years 1999–2004. Only the first case of breast cancer for that time period was included based on the date of diagnosis. Non-Massachusetts residents were excluded.

Analysis was limited to those women with American Joint Committee on Cancer stages I, IIa, and IIb. In addition, those women who were stage II but who had a tumor size greater than 5 cm were excluded from analysis since treatment guidelines are different for this subgroup.[15] Subjects who had missing information or "unknown" listed in the variables for poverty, reporting hospital, tumor size, and radiation receipt were also excluded from analysis and accounted for approximately 5% of the total sample. Seven percent of black non-Hispanics were excluded due to missing or unknown data compared with 5% of white non-Hispanics. Those excluded were slightly younger and more likely to be stage II (mean age: 61, 51% stage II) compared with those retained (mean age: 62, 37% stage II)).

The Institutional Review Boards of Boston University Medical Center and the Massachusetts Department of Public Health approved this study.

Variables

Demographic

The main independent variable was race, classified as black non-Hispanic or white non-Hispanic, which will be referred to as black or white, respectively. We excluded all other racial and ethnic categories because the primary aim of the study was to examine differences due to black race, and because other categories had numbers too small for meaningful analysis. Age was grouped into 5 categories (<50 years, 50–59 years, 60–69 years, 70–79 years, and >79 years). Marital status was classified as married or "all other," which included those not married and those whose marital status was unknown.

Tumor Characteristics

Stage was classified as stage I or II. For those with no pathologic stage recorded, clinical stage was used. In the years 1999–2003, 3025 subjects (2948 whites, or 16.5% of whites, and 77 blacks, or 14% of blacks) were clinically staged because there was no pathologic stage information. Of these, 2394 were staged as I and 631 were staged as II. Beginning in 2004, collaborative staging was implemented to address the issue of discrepancies in staging guidelines among the 3 major staging systems in the United States. Therefore, 2004 data does not have staging variables that discern whether a subject was clinically or pathologically staged. Tumor grade was recorded as 1 through 4 or unknown. Tumor size was categorized into <2 cm and 2 to 5 cm.

Insurance and Income

Insurance was categorized as uninsured, Medicaid only, Medicare with or without other coverage, commercial and "all other." In the "all other" category, the large majority (86%) were insured but had an unknown type of insurance, whereas the rest were either unknown with respect to having insurance at all (13%) or had military insurance (1%). Census tract information was collected and served as a proxy for socioeconomic status. Using methods described by the Public Health Disparities Geocoding Project, we linked each subject's census tract of residence with the percentage of the population in that census tract living below the federal poverty line according to the 2000 census.[16] Subjects were categorized according to quintile of census tract poverty level.

Reporting Hospital

To account for intraclass correlation by site of care, we adjusted for each of the 75 hospitals that reported the cancer case for each subject. Each reporting facility was assigned a code by the MCR to keep its identity private.

Treatment Outcomes

The main treatment outcome was dichotomous based on receipt of guideline concordant care (GCC), defined as receipt of mastectomy or BCS with radiation.[15] BCS was defined as partial mastectomy, partial mastectomy with nipple resection, lumpectomy or excisional biopsy, segmental mastectomy, wedge resection, quadrantectomy, or tylectomy. Mastectomy was defined as total simple mastectomy, simple mastectomy, modified radical mastectomy, or radical mastectomy. The MCR collects information on initiation of radiation therapy but not on whether treatment was completed. We did not examine receipt of hormonal therapy because such information, coupled with hormone receptor status, was not uniformly available in the database.

Mortality Outcomes

Breast cancer mortality was determined by linking state vital records with the National Death Index and by using ICD-10 codes.[17] Analysis was limited to subjects diagnosed before December 31, 2003, as the MCR had not linked state vital records with the National Death Index for the year 2004 at the time of analysis. Those who died of causes other than breast cancer or those not recorded as having died were excluded.

Analysis

The χ^2 test was used to test for differences in the distribution of categorical variables (demographic, socioeconomic, and clinicopathologic) among black and white

women, and the Student *t* test was used to compare distributions in continuous variables (age).

Logistic regression was used to assess the associations between outcomes and the main independent variable, patient race, while controlling for possible confounders. Mixed modeling methods were used due to multilevel nature of the independent variables. Different models were considered for receipt of GCC to assess for the degree to which the socioeconomic variables, marital status, income, and insurance contributed to receipt of the outcome. First, age and year were added into the model, then clinical variables, followed by marital status, insurance, income by census tract, and reporting hospital.

For survival analyses, Cox proportional hazard models were developed after testing for proportional hazard assumptions using graphical methods.[18] This procedure gave evidence that the assumption for the radiation received variable did not hold. This issue was addressed by running stratified analysis. All analyses were conducted using SAS (SAS Institute, NC), version 9.1.

RESULTS

Subjects included 21,155 white and 693 black women with stage I and II breast cancer. Patient and tumor characteristics are described in Table 1. Compared with whites, blacks were younger at diagnosis (62 vs. 56 years old), less likely to be married (56% vs. 34%), and more likely to have Medicaid (2% vs. 16%), to be uninsured (1% vs. 5%), or to be living in the lowest income census tract. Blacks were also more likely to be diagnosed at stage II compared with stage I, to have a larger tumor size, as well as a higher tumor grade. However, with respect to type of surgery received, blacks and whites had similar rates of receipt of BCS, mastectomy, and no surgery in unadjusted analyses.

Table 2 shows findings from models predicting receipt of GCC. The models sequentially adjust for stage, age, and year (partially adjusted model), plus marital status, census tract income, insurance, and reporting hospital (fully adjusted). In the partially adjusted model, blacks were significantly less likely to receive GCC after adjusting for age, year of diagnosis, and stage (OR: 0.75; 95% CI: 0.61–0.92). However, in the fully adjusted model this difference was no longer present (OR: 0.94; 95% CI: 0.75–1.17). Married women were more likely to receive GCC (OR: 1.20; 95% CI: 1.10–1.30) as were those older than 60, with 1 exception: those older than 79 were much less likely to have received GCC (OR: 0.36; 95% CI: 0.31–0.41). Being uninsured, or having Medicare or Medicaid were all independent predictors of GCC. Living in a census tract with high levels of poverty did not appear to be an independent predictor of GCC.

Because of concern for collinearity between age and insurance status, stratified analyses were run and showed that Medicare had a similar effect on receipt of the outcome for those younger than 65 (HR: 0.85; 95% CI: 0.66, 1.09) as in those 65 and older (HR: 0.74; 95% CI: 0.64, 0.86). (data not tabulated) These results were similar to the effect of Medicare on GCC in the unstratified analysis (HR: 0.81; 95% CI: 0.72, 0.91). Analyses were also stratified by insurance type and showed that black race had a different effect on treatment received in the Medicare group (OR: 0.8; 95% CI: 0.5, 1.3) compared with the Medicaid group (OR: 0.98; 95% CI: 0.6,

TABLE 1. Patient Characteristics Stratified by Race for Women Diagnosed With Stage I, II Breast Carcinoma in the Massachusetts Cancer Registry, 1999–2004

Variable	White Non-Hispanic (n = 21,155)	Black Non-Hispanic (n = 693)	*P**
Age category, yr (%)			<0.0001
<50	22	36	
50–59	23	25	
60–69	21	20	
70–79	22	13	
>79	12	6	
Mean age at diagnosis (yr)	62.1	56.3	<0.0001
Year of diagnosis (%)			0.0005
1999	18	15	
2000	18	17	
2001	17	14	
2002	17	20	
2003	15	14	
2004	15	20	
Marital status (%)			<0.0001
Married	56	34	
All other	44	66	
Insurance (%)			<0.0001
Commercial	47	40	
Uninsured	1	5	
Medicaid only	2	16	
Medicare	34	26	
All other†	16	13	
Income by census tract (%)			<0.0001
Quintile 1 (least poverty)	21	3	
Quintile 2	20	6	
Quintile 3	21	5	
Quintile 4	20	12	
Quintile 5	19	73	
Stage at diagnosis (%)			<0.0001
I	64	50	
II	36	50	
Tumor size (%)			<0.0001
<2 cm	71	58	
2–5 cm	29	42	
Tumor grade (%)			<0.0001
1	20	13	
2	42	33	
3	29	48	
4	1	<1	
Unknown	8	6	
Surgery type (%)			0.54
Mastectomy	27	27	
BCS	72	71	
No surgery	1	1	

*χ^2 test, except for mean age where *t* test was used.
†Primarily (86%) unknown type of insurance.

TABLE 2. Regression Models for Receipt of Guideline Concordant Care, 1999–2004

Characteristic	Unadjusted	Partially Adjusted*	Fully Adjusted†
Race			
White non-Hispanic	—	—	—
Black non-Hispanic	0.84 (0.69, 1.03)	0.75 (0.61, 0.92)‡	0.94 (0.75, 1.17)
Marital category			
Married	1.61 (1.50, 1.73)‡		1.20 (1.10, 1.30)‡
Age category			
<50	—	—	—
50–59	1.07 (.95, 1.20)	1.06 (0.94, 1.19)	1.07 (0.95, 1.21)
60–69	1.21 (1.08, 1.37)‡	1.19 (1.06, 1.35)‡	1.30 (1.14, 1.49)‡
70–79	1.01 (.90, 1.13)	0.99 (0.88, 1.11)	1.20 (1.04, 1.39)‡
>79	0.29 (0.26, .33)‡	0.28 (0.25, 0.32)‡	0.36 (0.31, 0.41)‡
Year diagnosis			
1999	—	—	—
2000	.94 (.84, 1.06)	0.92 (0.82, 1.04)	0.91 (0.81, 1.03)
2001	1.23 (1.09, 1.39)‡	1.23 (1.09, 1.40)‡	1.25 (1.10, 1.42)‡
2002	1.07 (0.95, 1.21)	1.07 (0.95, 1.21)	1.10 (0.97, 1.24)
2003	1.41 (1.24, 1.61)‡	1.43 (1.25, 1.63)‡	1.50 (1.31, 1.72)‡
2004	1.51 (1.32, 1.72)‡	1.53 (1.34, 1.75)‡	1.55 (1.35, 1.78)‡
Stage			
I	—	—	—
II	0.91 (0.84, 0.98)‡	0.90 (0.84, 0.97)‡	0.92 (0.85, 0.99)‡
Insurance			
Commercial	—		—
Uninsured	0.73 (0.52, 1.01)		0.71 (0.51, 0.995)‡
Medicaid only	0.73 (0.58, 0.91)‡		0.78 (0.62, 0.997)‡
Medicare	0.61 (0.56, 0.66)‡		0.81 (0.72, 0.91)‡
All other	0.89 (0.77, 0.995)‡		0.94 (0.84, 1.06)
Income category			
1st quintile	—		—
2nd quintile	1.08 (0.96, 1.21)		1.12 (0.99, 1.27)
3rd quintile	1.06 (0.94, 1.19)		1.13 (0.99, 1.28)
4th quintile	0.87 (0.78, 0.98)‡		0.99 (0.87, 1.12)
5th quintile	0.79 (0.71, 0.89)‡		0.99 (0.88, 1.13)

*Race, stage, age, and year.
†Partially adjusted model plus marital status, income category by census tract, insurance type and adjustment for intra-class correlation by reporting hospital.
‡$P < 0.05$.

1.6) though neither result was statistically significant. (data not tabulated) In addition, removing insurance from the model did not change the results compared with the fully adjusted model.

For the analysis in Table 2, mixed (or random effects) modeling was compared with fixed effects modeling. The fully adjusted mixed model (displayed in Table 2) showed a point estimate closer to 1 than did the fully adjusted fixed effects model (OR: 0.87; 95% CI: 0.70, 1.08) (data not tabulated).

The age standardized breast cancer mortality rate per 100,000 person-years was 23.1 and 11.4 for blacks and whites, respectively (data not tabulated). Cox proportional hazard modeling of breast cancer mortality showed that blacks had a hazard ratio of 2.34 (95% CI: 1.74–3.15) after adjusting for age (Table 3). This disparity persisted, though attenuated, after controlling for clinical, treatment, socioeconomic variables, and reporting hospital (HR: 1.55; 95% CI: 1.13–2.13). Living in a census tract with the highest quintile of poverty was also an independent predictor of death (HR: 1.32; 95% CI: 1.06–1.65).

Sensitivity analyses were run to assess the effect that competing risks might have on the mortality results. When the event (death) was changed to include all cancer deaths instead of only breast cancer deaths, the hazard ratio for black race did not change appreciably (HR: 1.61; 95% CI: 1.21, 2.12). When the event analyzed was all noncancer deaths, the results also did not change appreciably (HR: 1.68; 95% CI: 1.13, 2.50).

DISCUSSION

We used the MCR to examine racial differences in first course of treatment and mortality for early-stage invasive

TABLE 3. Cox Regression Model Predicting Breast Cancer Specific Mortality Among Women in the Massachusetts Cancer Registry, 1999–2003. N = 18,399

Characteristic	Model 1*	Model 2†	Model 3‡
Race			
White non-Hispanic	—	—	
Black non-Hispanic	2.34 (1.74, 3.15)§	1.86 (1.38, 2.51)§	1.55 (1.13, 2.13)§
Married	—	—	0.93 (0.80, 1.09)
Age category			
<50	—	—	—
50–59	0.79 (0.63, 0.99)§	0.96 (0.77, 1.19)	0.94 (0.75, 1.18)
60–69	0.56 (0.44, 0.70)§	0.80 (0.64, 1.01)	0.82 (0.64, 1.05)
70–79	0.58 (0.47, 0.72)§	0.87 (0.70, 1.08)	0.88 (0.68, 1.14)
>79	0.76 (0.61, 0.94)§	1.10 (0.88, 1.37)	1.12 (0.85, 1.47)
Stage			
I	—	—	—
II	—	1.79 (1.44, 2.21)§	1.77 (1.43, 2.19)§
Tumor grade			
1	—	—	
2	—	1.58 (1.13, 2.21)§	1.57 (1.12, 2.19)§
3	—	3.60 (2.60, 4.98)§	3.65 (2.64, 5.05)§
4	—	3.14 (1.68,5.89)§	3.29 (1.76, 6.17)§
Tumor size			
<2 cm	—	—	—
2–5 cm	—	1.77 (1.44, 2.16)§	1.77 (1.44, 2.17)§
Insurance			
Commercial	—	—	—
Uninsured	—	—	1.11 (0.63, 1.94)
Medicaid	—	—	0.93 (0.66, 1.33)
Medicare only	—	—	0.87 (0.71, 1.07)
All other	—	—	0.80 (0.65, 0.99)§
Income category			
Quintile 1	—	—	—
Quintile 2	—	—	0.94 (0.74, 1.19)
Quintile 3	—	—	0.63 (0.75, 1.21)
Quintile 4	—	—	1.17 (0.93, 1.47)
Quintile 5	—	—	1.32 (1.06, 1.65)§

*Adjusted for age.
†Model 1 plus stage, tumor size, and grade.
‡Model 2 plus marital status, insurance, income, and adjusted for intraclass correlation by reporting hospital.

breast cancer. As has been found in other studies of breast cancer, blacks in this cohort were younger, poorer, and more likely to be at a later stage at diagnosis compared with their white counterparts.[11,14] We found certain measures of socioeconomic status, namely marital status and insurance type, to independently predict treatment. However, while race itself was not an independent predictor of GCC, blacks were nevertheless more likely to die of breast cancer during the study years, even after adjustment for covariates.

A recent analysis of the SEER-Medicare database found similar results regarding mortality. Using data from 1994–1999, Curtis et al found that blacks with stages II and III disease were at increased risk of cancer death after controlling for clinical, treatment, and socio-demographic variables that included income and type of community. The outcome, however, was not limited to breast cancer deaths.[14] Other studies have shown that blacks diagnosed with early-stage disease have higher all-cause mortality and all-cancer mortality but have not controlled for socioeconomic factors, reporting hospital, and used breast cancer specific mortality as the outcome.[4,9,11,12,19–21] Using more recent data, our study shows that black women with stages I and II breast cancer have increased disease-specific mortality that persists even after adjusting for clinical and treatment variables as well as marital status, income by census tract, type of insurance, and reporting hospital.

With respect to treatment received, other studies have had mixed results. When examining type of surgery received, our study is in line with several others in that blacks and whites were equally likely to receive BCS compared with mastectomy.[9,10,22,23] However after adjusting for age, year, and stage, we found that black women were less likely to receive GCC. Analysis (not shown) run with the outcome of radiation receipt among those receiving BCS demonstrated

similar results, suggesting that lack of receipt of this treatment may be driving these findings for GCC. Marital status, income by census tract, and insurance appear to be the primary mediators of this difference, as it was no longer present when these factors were added into the model. Indeed, as shown in the same analysis, those who were married were 20% more likely to receive GCC, and blacks were less likely to be married. Given that radiation therapy usually requires daily hospital visits for 6 weeks, it is not surprising that married women, with marital status being a proxy for social support, were more likely to receive GCC. These findings are supported by other studies which show that being married is an important predictor of not only receiving recommended therapies, but also stage at diagnosis and mortality.[24–26]

Evidence suggests that controlling for reporting hospital is important in studies of racial disparities in treatment, given that some treatments may be less rapidly adopted by certain hospitals, which may serve a disproportionate number of a given racial or ethnic group.[30] However, our study did not find that controlling for this variable significantly changed the results with respect to the primary predictor variable, race.

Our study has several limitations. The MCR does not collect information on comorbidity so we cannot exclude the possibility that lack of appropriate treatment was a function of increased comorbid illness. The use of census track for defining poverty level could be confounded by other issues that are related to neighborhood of residence, such as attitudes or knowledge about treatment. In addition, research has shown that race and ethnicity plays a role in both patient refusal of recommended treatment and in health care providers not offering treatment.[31–33] However, our data did not allow us to determine if such preferences were important. Additionally, prognostic variables such as hormone receptor status were not available for the mortality analysis. Black women may be more likely to have hormone receptor negative tumors than whites,[34] and since such tumors have a poorer prognosis, such a biologic difference may help explain the mortality disparity seen. Some studies have shown that black women are less likely to receive adjuvant hormonal treatment even after controlling for hormone receptor status.[23,35] Therefore, it is possible that decreased receipt of such recommended therapy may explain the mortality differences observed in our study. Finally, because this study analyzes Massachusetts data only, the results may not generalize to other states.

In summary, our study demonstrates that in early-stage invasive breast cancer, being a black non-Hispanic woman is an independent risk factor for disease-specific mortality, and that this disparity is not fully explained by differences in receipt of guideline concordant cancer care or available measures of socioeconomic status. In addition, we found that marital status, income, and insurance were the primary mediators of racial differences in receipt of GCC. Our findings support the notion that racial differences previously reported are a result of a complex interplay of both sociodemographic and tumor characteristics, and that attention to reducing sociodemographic barriers to care may influence receipt of recommended treatments and survival.

ACKNOWLEDGMENTS

The authors thank Rebecca Silliman, Wee Lock Ooi, Cynthia Boddie-Willis, Jane Mendez, Michael Winter, and Timothy Lash for their valuable contributions to this manuscript. The authors also thank the Centers for Disease Control and Prevention for its support under cooperative agreement 1 U58 DP000821-01 awarded to the Massachusetts Department of Public Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Massachusetts Department of Public Health.

REFERENCES

1. Jemal A, Siegel R, Ward E, et al. Cancer Statistics, 2008. *CA Cancer J Clin*. 2008;57:43–66.
2. Parker SL, Davis KJ, Wingo PA, et al. Cancer statistics by race and ethnicity. *CA Cancer J Clin*. 1998;48:31–48.
3. Polite BN, Olopade OI. Breast cancer and race: a rising tide does not lift all boats equally. *Perspect Biol Med*. 2005;48:S166–S175.
4. Joslyn SA, West MM. Racial differences in breast carcinoma survival. *Cancer*. 2000;88:114–123.
5. NIH Consensus Development Conference statement on the treatment of early-stage breast cancer. *Oncology (Williston Park)*. 1991;5:120–124.
6. Early Breast Cancer Trialists' Collaborative Group. Effects of radiotherapy and surgery in early breast cancer: an overview of the randomized trials. *N Engl J Med*. 1995;333:1444–1455.
7. Clarke M, Collins R, Darby S, et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet*. 2005;366:2087–2106.
8. Haggstrom DA, Quale C, Smith-Bindman R. Differences in the quality of breast cancer care among vulnerable populations. *Cancer*. 2005;104:2347–2358.
9. Roetzheim RG, Gonzalez EC, Ferrante JM, et al. Effects of health insurance and race on breast carcinoma treatments and outcomes. *Cancer*. 2000;89:2202–2213.
10. Michalski TA, Nattinger AB. The influence of black race and socioeconomic status on the use of breast-conserving surgery for Medicare beneficiaries. *Cancer*. 1997;79:314–319.
11. Joslyn SA. Racial differences in treatment and survival from early-stage breast carcinoma. *Cancer*. 2002;95:1759–1766.
12. Li CI, Malone KE, Daling JR, et al. Differences in breast cancer stage, treatment, and survival by race and ethnicity. *Arch Intern Med*. 2003;163:49–56.
13. Mandelblatt JS, Kerner JF, Hadley J, et al. Variations in breast carcinoma treatment in older Medicare beneficiaries: is it black or white. *Cancer*. 2002;95:1401–1414.
14. Curtis E, Quale C, Haggstrom D, et al. Racial and ethnic differences in breast cancer survival: how much is explained by screening, tumor severity, biology, treatment, comorbidities, and demographics? *Cancer*. 2008;112:171–180.
15. Network NCC. *NCCN Clinical Practice Guidelines in Oncology*. National Comprehensive Cancer Network; 2007.
16. Krieger N, Chen JT, Waterman PD, et al. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures–the public health disparities geocoding project. *Am J Public Health*. 2003;93:1655–1671.
17. Organization WH. International Statistical Classification of Disease and Related Health Problems, 10th Revision, Version for 2007. Available at: http://www.who.int/classifications/icd/en/. Accessed April, 2008.
18. Hess KR. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Stat Med*. 1995;14:1707–1723.
19. Ansell D, Whitman S, Lipton R, et al. Race, income, and survival from breast cancer at two public hospitals. *Cancer*. 1993;72:2974–2978.
20. Boyer-Chammard A, Taylor TH, Anton-Culver H. Survival differences in breast cancer among racial/ethnic groups: a population-based study. *Cancer Detect Prev*. 1999;23:463–473.

21. Bradley CJ, Given CW, Roberts C. Race, socioeconomic status, and breast cancer treatment and survival [comment]. *J Natl Cancer Inst*. 2002;94:490–496.
22. Velanovich V, Yood MU, Bawle U, et al. Racial differences in the presentation and surgical management of breast cancer. *Surgery*. 1999; 125:375–379.
23. Banerjee M, George J, Yee C, et al. Disentangling the effects of race on breast cancer treatment. *Cancer*. 2007;110:2169–2177.
24. Hershman DL, Buono D, McBride RB, et al. Surgeon characteristics and receipt of adjuvant radiotherapy in women with breast cancer. *J Natl Cancer Inst*. 2008;100:199–206.
25. Voti L, Richardson LC, Reis IM, et al. Treatment of local breast carcinoma in Florida: the role of the distance to radiation therapy facilities. *Cancer*. 2006;106:201–207.
26. Osborne C, Ostir GV, Du X, et al. The influence of marital status on the stage at diagnosis, treatment, and survival of older women with breast cancer. *Breast Cancer Res Treat*. 2005;93:41–47.
27. McBride R, Hershman D, Tsai WY, et al. Within-stage racial differences in tumor size and number of positive lymph nodes in women with breast cancer. *Cancer*. 2007;110:1201–1208.
28. Lyman GH, Giuliano AE, Somerfield MR, et al. American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. *J Clin Oncol*. 2005;23: 7703–7720.
29. Schrenk P, Rieger R, Shamiyeh A, et al. Morbidity following sentinel lymph node biopsy versus axillary lymph node dissection for patients with breast carcinoma. *Cancer*. 2000;88:608–614.
30. Edwards BK, Brown ML, Wingo PA, et al. Annual report to the nation on the status of cancer, 1975–2002, featuring population-based trends in cancer treatment. *J Natl Cancer Inst*. 2005;97:1407–1427.
31. Rathore SS, Ordin DL, Krumholz HM, et al. Race and sex differences in the refusal of cardiac catheterization among elderly patients hospitalized with acute myocardial infarction. *Am Heart J*. 2002;144:1052–1056.
32. Canto JG, Allison JJ, Kiefe CI, et al. Relation of race and sex to the use of reperfusion therapy in Medicare beneficiaries with acute myocardial infarction. *N Engl J Med*. 2000;342:1094–1100.
33. Rathore SS, Krumholz HM. Differences, disparities, and biases: clarifying racial variations in health care use. *Ann Intern Med*. 2004;141:635–638.
34. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295: 2492–2502.
35. Bickell NA, Wang JJ, Oluwole S, et al. Missed opportunities: racial disparities in adjuvant breast cancer treatment. *J Clin Oncol*. 2006;24: 1357–1362.

Patient Education and Counseling 75 (2009) 315–320

Contents lists available at ScienceDirect

Patient Education and Counseling

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

Using computer agents to explain medical documents to patients with low health literacy

Timothy W. Bickmore [a,*], Laura M. Pfeifer [a], Michael K. Paasche-Orlow [b]

[a] Northeastern University College of Computer and Information Science, Boston, MA, USA
[b] Boston University School of Medicine, Boston, MA, USA

ARTICLE INFO

Article history:
Received 15 September 2008
Received in revised form 8 February 2009
Accepted 9 February 2009

Keywords:
Health literacy
Computer agent
Relational agent
Informed consent

ABSTRACT

Objective: Patients are commonly presented with complex documents that they have difficulty understanding. The objective of this study was to design and evaluate an animated computer agent to explain research consent forms to potential research participants.
Methods: Subjects were invited to participate in a simulated consent process for a study involving a genetic repository. Explanation of the research consent form by the computer agent was compared to explanation by a human and a self-study condition in a randomized trial. Responses were compared according to level of health literacy.
Results: Participants were most satisfied with the consent process and most likely to sign the consent form when it was explained by the computer agent, regardless of health literacy level. Participants with adequate health literacy demonstrated the highest level of comprehension with the computer agent-based explanation compared to the other two conditions. However, participants with limited health literacy showed poor comprehension levels in all three conditions. Participants with limited health literacy reported several reasons, such as lack of time constraints, ability to re-ask questions, and lack of bias, for preferring the computer agent-based explanation over a human-based one.
Conclusion: Animated computer agents can perform as well as or better than humans in the administration of informed consent.
Practice implications: Animated computer agents represent a viable method for explaining health documents to patients.

1. Introduction

Face-to-face encounters with a health provider – in conjunction with written instructions – remains one of the best methods for communicating health information to patients in general, but especially those with low health literacy [1–4]. Face-to-face consultation is effective because providers can use verbal and nonverbal behaviors, such as head nods, hand gesture, eye gaze cues and facial displays to communicate factual information to patients, as well as to communicate empathy [5] and immediacy [6] to elicit patient trust. 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 agreement and understanding [7]. Finally, face-to-face interaction allows providers to dynamically assess a patient's level of understanding based on the patient's verbal and nonverbal behavior and to repeat or elaborate information as necessary [8].

However, there are several pervasive problems that limit a clinician's capacity to communicate effectively. Providers can only spend a limited amount of time with each patient [9]. Time pressures can result in patients feeling too intimidated to ask questions. Another problem is that of "fidelity": providers do not always perform in accordance with recommended guidelines, resulting in significant variation in the delivery of health information.

Given the efficacy of face-to-face consultation, a promising approach for conveying health information to patients with limited health literacy is the use of computer animated agents that simulate face-to-face conversation with a provider [10]. The benefits of using conversational agents include: use of verbal and nonverbal conversational behaviors that signify understanding and mark significance, and convey information in redundant channels of information (e.g., hand gestures, such as pointing, facial display of emotion, and eye gaze); use of of verbal and nonverbal communicative behaviors to maximize comprehension;

* Corresponding author at: Northeastern University College of Computer and Information Science, 360 Huntington Ave WVH202, Boston, MA 02115, USA. Tel.: +1 617 373 5477; fax: +1 617 373 5121.
E-mail address: bickmore@ccs.neu.edu (T.W. Bickmore).

0738-3991/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.pec.2009.02.007

use of verbal and nonverbal communicative behaviors used by providers to establish trust and rapport with their patients in order to increase satisfaction and adherence to treatment regimens [11]; adaptation of their messages to the particular needs of patients and to the immediate context of the conversation; and provision of 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. This latter point is particularly important as health providers frequently fail to elicit patients' questions, and patients with limited health literacy are even less likely than others to ask questions [12].

According to the 2004 National Assessment of Adult Literacy, fully 36% of American adults have limited health literacy skills, with even higher rates of prevalence among patients with chronic diseases, those who are older, minorities, and those who have lower levels of education [13,14]. Seminal reports about the problem of health literacy include a sharp critique of current norms for overly complex documents in health care such as informed consent [15,16]. Indeed, a significant and growing body of research has brought attention to the ethical and health impact of overly complex documents in healthcare [17,18]. Computer agents may provide a particularly effective solution for addressing this problem, by having the agents describe health documents to patients using exemplary communication techniques for patients with limited health literacy and by providing this information in a context unconstrained by time pressures.

Informed consent agreements for individuals to participate in medical research represent a particular challenge for individuals with limited health literacy to understand, since they typically encode many subtle and counter-intuitive legal and medical concepts. They are often written at a reading level that is far beyond the capacity of most subjects [19,20]. Researchers may not have the resources to ensure that participants understand all the terms of the consent agreement. Indeed, many potential research subjects sign consent forms that they do not understand [21–23].

Consequently, we modified an existing computer agent framework designed for health counseling [10,11] to provide explanation of health documents such as research informed consent forms. In this paper we describe the development of this agent, and then present a preliminary evaluation of the computer agent in a three-arm randomized trial in which the agent explains an informed consent document for participation in a genetic repository.

2. Methods

2.1. Preliminary studies. Part 1. Health document explanation by human experts

We conducted two empirical studies to characterize how human experts explain health documents to their clients in face-to-face interactions [24]. The first study was conducted with four different experts explaining two different health documents to research confederates. The second study was conducted with one expert explaining health documents to three laypersons with different levels of health literacy. Our primary focus was a micro-analysis of the nonverbal behavior exhibited by the expert in order to inform the development of a computational model of document explanation. We found that one kind of nonverbal behavior was nearly ubiquitous: the use of pointing gestures towards the document by the expert (Fig. 1). Of the 1994 expert utterances analyzed, 26% were accompanied by a hand gesture, and 90% of these involved pointing at the document.

We derived a predictive model of the occurrence and form of referential hand gestures and other nonverbal behavior used by the experts during their explanations. We found that initial mentions of part of a document were more likely to be accompanied by a pointing gesture (43% vs. 19%) and that the kind of document object referred to (page vs. section vs. word or image) was predictive of the kind of hand gesture used (e.g., using a flat hand to refer to a page vs. pointing with a finger to refer to a word). We also found that the expert in the second study omitted a significant amount of detail and used more scaffolding (description of document structure) when describing a health document to listeners with low health literacy, compared to listeners with adequate health literacy.

Fig. 1. Explanation of consent by experts.

2.2. Preliminary studies. Part 2. Adapting a computer agent for health document explanation

An existing computer agent framework designed for health counseling [10,11] was modified to provide explanation of health documents. The framework features an animated computer agent whose nonverbal behavior is synchronized with a text-to-speech engine (Fig. 2). Patient contributions to the conversation are made via a touch screen selection from a multiple choice menu of utterance options, updated at each turn of the conversation. Dialogues are scripted using a custom hierarchical transition network-based scripting language. Agent utterances can be dynamically tailored based on information about the patient, information from previous conversations, and the unfolding discourse context [10]. The animated agent has a range of nonverbal behaviors that it can use, including: hand gestures, body posture shifts, gazing at and away from the patient, raising and lowering eyebrows, head nods, different facial expressions, and variable proximity.

The framework was extended for document explanation in several ways. A set of animation system commands was added to allow document pages to be displayed by the character (Fig. 2), with page changes accompanied by a page-turning sound. A set of document pointing gestures was added so that the agent could be commanded to point anywhere in the document with either a pointing hand or an open hand. While the document is displayed, the agent can continue using its full range of head and facial behavior, with gaze-aways modified so that the agent looks at the document when not looking at the patient (in our studies of human experts, the expert gazed at the document 65% of the time and at the patient 30% of the time). We also extended our text-to-embodied-speech translation system ("BEAT" [25]) to automatically generate document pointing gestures given the verbal content of the document explanation script, based on models from our earlier studies [24].

We conducted a preliminary study of the document explanation agent in a three-arm randomized trial with 18 participants

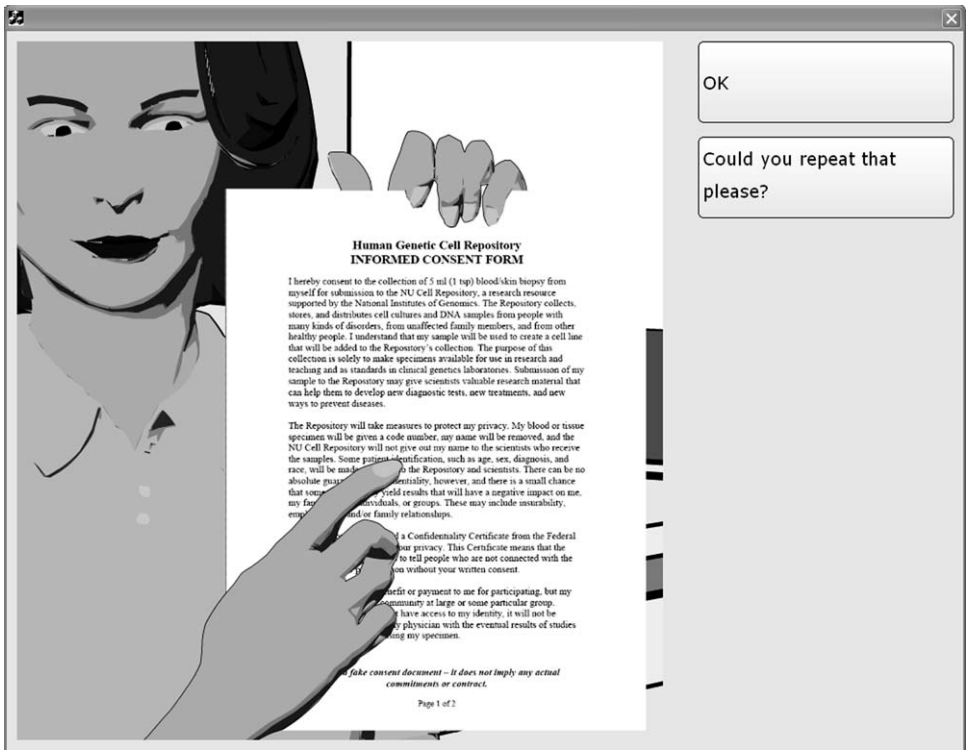


Fig. 2. Computer agent interface.

aged 19–33, in which each participant experienced two of the three conditions. We compared agent-based health document explanation with explanation by human experts and a self-study condition. While there were no significant effects of study condition on comprehension of the documents (measured by post-intervention knowledge tests), the participants who interacted with both the agent and human were significantly more satisfied with the agent (paired $t(5) = 2.7$, $p < .05$) and with the overall experience (paired $t(5) = 2.9$, $p < .05$), compared to the human [24].

2.3. Part 3. Evaluation of computer agent for explanation of research informed consent

While the preliminary study provided feedback on the promise of using agents for health document explanation, it lacked ecological validity because the participants were primarily college students who had a fairly high level of health literacy. Thus, the primary purpose of the third research activity was to repeat the pilot evaluation with a population in which limited health literacy is represented.

We conducted an evaluation study to test the efficacy of our agent-based document explanation system, compared with a standard of care control (explanation by a human) and a non-intervention control (self-study of the document in question) for individuals with adequate and inadequate health literacy. The study was a three-arm (COMPUTER AGENT vs. HUMAN vs. SELF) between-subjects randomized experimental design.

An interaction script was created to present an informed consent form for participation in a genetic repository, based on the preliminary work described above. The consent document used was taken with minor revisions from an existing National Institute of General Medical Sciences template for genetic repository research that has been used in multiple NIH funded projects [26]. The example of a study involving a genetic repository was chosen because we wanted little overlap between the simulated consent experience and the actual consent document used for participating in the current study, and we wanted material that would be largely foreign to participants to decrease the influence of prior knowledge. In each script, patients could simply advance linearly through the explanation (by selecting "OK"), ask for any utterance to be repeated ("Could you repeat that please?"), request major sections of the explanation to be repeated, or request that the entire explanation be repeated. Any number of repeats could be requested and, although the scripting language has the ability to encode rephrasings when an utterance is repeated, for the current study the agent would repeat the exact same utterance when a repeat was requested for any part of the script. The agent was deployed on a mobile cart with a touch screen attached via an articulated arm.

2.3.1. Measures

In addition to basic demographics, we assessed health literacy using the 66-word version of the Rapid Estimate of Adult Literacy in Medicine (REALM) [27]. 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 [28–31]. We also created a knowledge test for the consent document, based on the Brief Informed Consent Evaluation Protocol (BICEP) [22]. This test was administered in an "open book" fashion with the participant able to refer to a paper copy of the consent form during the test. We augmented the BICEP with scale measures of likelihood to sign the consent document, overall satisfaction with the consent process, and perceived pressure to sign the consent document. In the COMPUTER AGENT and HUMAN conditions, the number of questions or requests for clarifications asked by participants during the explanation of the consent document was also counted.

2.3.2. Participants

Twenty-nine subjects participated in the study, were recruited via fliers posted around the Northeastern University neighborhood

Table 1
Study results (mean and (S.D.)).

Measure	Condition				Literacy			Condition × literacy
	Agent ($N = 9$)	Human ($N = 9$)	Self ($N = 11$)	Main effect sig.	Adequate ($N = 16$)	Inadequate ($N = 13$)	Main effect sig.	Interaction sig.
Comprehension (% correct)	42.20 (20.33)	39.44 (12.86)	25.91 (11.36)	0.006	41.88 (17.78)	26.92 (9.69)	0.001	0.042
Satisfaction (1–7)	6.56 (1.01)	3.89 (2.47)	5.09 (1.70)	0.018	5.50 (1.90)	4.77 (2.24)	0.280	0.682
Likelihood to sign (1–7)	6.21 (1.30)	2.78 (2.39)	3.91 (2.43)	0.011	4.06 (2.52)	4.54 (2.54)	0.524	0.588
Pressure to sign (1–7)	2.11 (2.09)	2.00 (2.00)	1.55 (0.93)	0.719	1.75 (1.34)	2.00 (2.04)	0.666	0.822
Questions asked	1.12 (2.10)	1.22 (2.64)		0.967	0.89 (2.67)	1.50 (2.00)	0.559	0.207

and in a nearby apartment complex whose demographic consisted of mostly older minority adults, and were compensated for their time. Participants had to be 18 years of age or older and able to speak English. Participants were 66% female, aged 28–91 (mean 60.2). Three were categorized as 3rd grade or below, four as 4th–6th grade, six as 7th–8th grade, and the rest as high school level.

2.3.3. Procedure

The study took place either in a common room of the apartment complex or the Human–Computer Interaction laboratory at Northeastern University. After arriving, people who consented to participate filled out a demographic questionnaire and then had the REALM health literacy evaluation administered.

Following this, they were exposed to one of three treatments in which a consent document was explained to them by either the COMPUTER AGENT or a HUMAN, or were given time to read the document on their own (SELF). For the COMPUTER AGENT condition, they were given a brief training session on how to interact with the computer agent. The experimenter then gave the participant a paper copy of the consent document so they could follow along with the computer agent's explanation, and left the room. At the end of the interaction, the computer agent informed the participant that they could take as much time as they needed to review the document before signaling to the experimenter that they were ready to continue. For the HUMAN condition, a second research assistant explained the document to the study participant. Two different female instructors played this second role, and both had significant experience administering informed consent for research studies. The instructor was blind to the computer agent interaction script content and evaluation instruments, and was simply asked to explain the document in question to the participant. For the CONTROL condition, participants were handed the document and told to take as much time as they needed to read and understand it, and were then left alone in the observation room until they signaled they were ready to continue.

The research assistant then verbally administered the knowledge test in "open book" format, with the participant being able to reference their paper copy during the test. The process measures were then verbally administered and a semi-structured interview was conducted to ask participants about their impressions of the study.

3. Results

Of the 29 participants, 13 (45%) had inadequate health literacy. We conducted full-factorial ANOVAs for all measures, with study CONDITION (COMPUTER AGENT, HUMAN, SELF) and health LITERACY (ADEQUATE, INADEQUATE) as independent factors, and LSD post hoc tests when applicable. Table 1 shows descriptive statistics for the outcome measures.

There was a significant interaction between CONDITION and LITERACY on knowledge test comprehension scores, $F(2,23) = 4.41$, $p < .05$ (Fig. 3). Post hoc tests indicated that, for participants with adequate health literacy, explanations by HUMAN and COMPUTER AGENT resulted in significantly greater comprehension compared to SELF study (with no significant difference between HUMAN and COMPUTER AGENT). However, for participants with inadequate health literacy, there were no significant differences on comprehension between study conditions, and they scored significantly lower as a group compared to participants with adequate health literacy.

There was a main effect of study CONDITION on satisfaction with the consent process, $F(2,23) = 4.78$, $p < .05$, with participants being significantly more satisfied with explanations by the COMPUTER AGENT compared to the HUMAN (participants were also more satisfied with the COMPUTER AGENT compared to SELF study, with post hoc tests approaching significance, $p = .09$).

There was also a main effect of study CONDITION on self-reported likelihood to sign the consent document, $F(2,32) = 5.46$, $p < .05$, with participants significantly more likely to sign the consent form following explanation by the COMPUTER AGENT, compared to either explanation by the HUMAN or SELF study.

There were no significant differences between groups on perceived PRESSURE to sign the consent form, $F(2,23) = 0.20$, $p = .72$.

Finally, it appeared that participants with limited health literacy asked more questions of the computer agent compared to the human, while those with adequate health literacy asked more questions of the human, although this interaction was not significant, $F(1,13) = 1.76$, $p = .21$ (Fig. 4).

3.1. Qualitative results

Participant responses to semi-structured interview questions were transcribed from the videotape and common themes were identified [32].

When asked about their impressions of the computer agent, the most frequently mentioned theme (7 participants) was that the computer agent was clear, direct and easy to understand. One

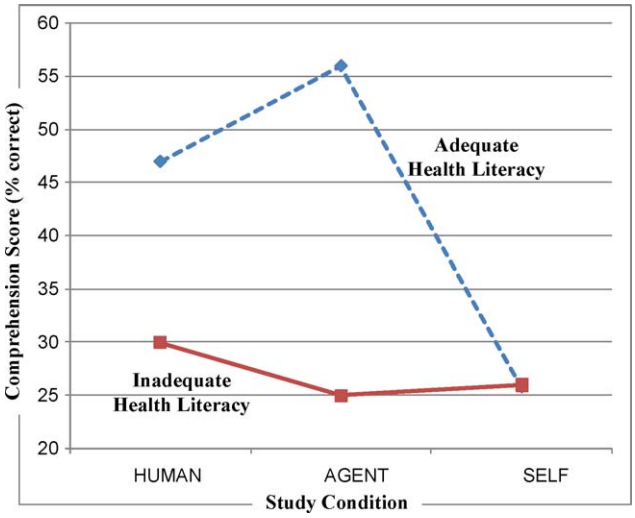


Fig. 3. Comprehension of informed consent.

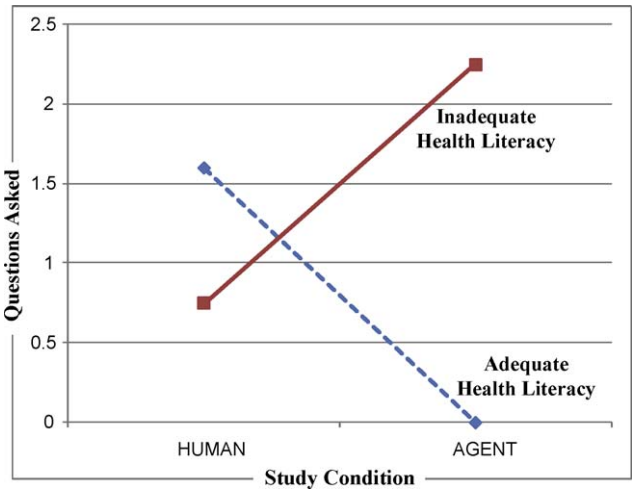


Fig. 4. Questions asked of the computer agent or human.

participant explicitly said that this clarity was due to the computer agent's ability to point at the virtual document, with the participant following along:

> "She was very direct and very clear when she was explaining it, she was explaining it very nice and slow. And she was pointing to the areas that needed to be focused on. When she was explaining it, she was breaking it down on the paper. Where you couldn't get lost if you were concentrating on what she was saying. Because it was right there in front of you [points at computer] and it's like right here [points on paper document], and it's just she was explaining the whole thing. And I was very comfortable with it because as I was reading it, I understood what she was saying and what I was seeing in front of me." (49 year old female, adequate literacy)

The second most common impression of the computer agent (4 mentions) was that participants felt they could take as much time as they needed, and did not feel embarrassed asking the computer agent to repeat itself:

> "Elizabeth [the name of the agent] was very, uh, patient, and if she says something to you that you don't understand, she will repeat it again if you push the button. And she would take her time." (68 year old female, limited literacy).

> "For me, you know, when it's on the computer I can do it five times over if I want to. I can just hit repeat, wait I didn't understand it, I can just repeat it again. You know, but I wouldn't do that with you [a human] because if I didn't understand it I might ask you one time to repeat it, and if I still didn't understand it I wouldn't ask you to repeat it. Because I wouldn't want to seem stupid." (47 year old female, adequate literacy).

Two participants said that they liked the computer agent because she was polite and did not talk down to them:

> "She was really polite, she was really polite. That I liked. Besides the fact, more important than anything else, she looked at me and she talked to me. You know, she was talking to me as a person, as opposed to, um, looking down on me and saying 'did you understand me!' you know? And that made me feel really good." (50 year old female, adequate literacy)

Other positive comments included that the computer agent was "informative" and "correct" (2 mentions), that the computer agent was "honest" (1 mention), and that the respondents liked using the touch screen instead of a mouse (1 mention).

There were two negative comments about the computer agent. One participant mentioned that the computer agent seemed "impersonal", and another felt the computer agent was too "robotic".

When asked whether they would prefer that health documents be explained to them by a person or a computer agent, 3 of the 9 participants who interacted with the agent said they would prefer the agent, 1 said they would prefer a human, and 1 said that either would be equally acceptable (the others did not respond). For example:

> "I think she did the same as talking to an ordinary person in the hospital. Uh, except she would give you a little more information than they do. Because sometimes they only tell you a little bit, you know what I'm saying, and she explain the whole thing." (68 year old female, limited literacy).

4. Discussion and conclusion

4.1. Discussion

The computer agent did as well as or better than the human on all measures, with participants (regardless of literacy level) reporting higher levels of satisfaction with the consent process and greater likelihood to sign the consent document when it was explained by the computer agent, compared to either explanation by a human or self study. In addition, explanation by the computer agent led to the greatest comprehension of the document, but only for those participants with adequate levels of health literacy; participants with limited literacy scored poorly on comprehension in all treatment conditions.

The tendency for participants with inadequate health literacy to ask more questions of the agent may be due to their being comfortable asking a computer repeated questions without feeling "stupid" (as one participant put it). However, an alternate hypothesis is that they asked more questions of the agent because they had a more difficult time understanding it.

The low comprehension scores for participants with inadequate health literacy indicate that much work remains to make the computer agent effective for this population. One pedagogical methodology espoused for patients with limited health literacy is "teach back" in which the patient is asked to teach what they have learned back to the health educator [7]. While there are some problems implementing this in an unconstrained way within our system, it is at least possible to add comprehension checks at key places in the agent-patient conversation and to have the computer agent provide additional information or review if it appears the patient is having problems.

Limitations of our study include the generalizability of our findings, especially given the very small convenience sample used. The research assistants who explained the consent forms to participants may not be representative of most researchers who perform this function. There are also ecological validity issues with our study settings, although we would expect that in a rushed clinical environment the agent may outperform a typical research assistant by an even wider margin than we observed.

4.2. Conclusion

Our future work is focused on several extensions to the system and more extensive evaluation. We plan to add audio prompts to the user interface so that patients who are unable to read the text of their conversational responses can still use the system. We are also developing a framework that will allow health document

templates to be instantiated and explained, so that, for example, consent form "boilerplates" can be instantiated with the details of a research study, and the computer agent would be able to explain the document to a patient without further scripting or programming. We have also developed the capability for the agent to keep track of specific issues and questions that it could not resolve for the patient, and output these at the end of the session for follow up by a human research assistant or clinician. We also plan to explore the integration of the conversational agent with other multimedia content, such as video clips, to further explain complex topics such as randomization, or numerical concepts like rates—ideas that can be hard to convey verbally. Finally, we plan to replicate the evaluation study in a clinic or hospital environment, where we would expect that the advantages of the computer agent-based approach would be even greater given the time pressures that most human providers are under.

4.3. Practice implications

This work suggests that animated computer agents can perform as well as people in explaining health documents to patients. For the administration of informed consent in particular, it is possible to construct computer agents that result in at least as much understanding of the consent form, satisfaction of the process, and study participation rates compared to the administration of informed consent by human research assistants. Time and cost savings for research studies or medical procedures requiring informed consent could be significant when large number of patients are involved. The use of this technology may also lead to more ethical treatment of patients through a more controlled administration of informed consent and automated comprehension tests.

Conflict of interest

The authors have no conflicts of interest that could influence this work.

Acknowledgements

Thanks to Lindsey Hollister and Maggie McElduff for their assistance in conducting the study.

Role of funding: This work was supported by a grant from the NIH National Heart Lung and Blood Institute (HL081307-01). The sponsors had no involvement in the study design, collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the paper for publication.

References

[1] Clinite J, Kabat H. Improving patient compliance. J Am Pharm Assoc 1976;16:74–6.
[2] Madden E. Evaluation of outpatient pharmacy patient counseling. J Am Pharm Assoc 1973;13:437–43.
[3] Morris L, Halperin J. Effects of written drug information on patient knowledge and compliance: a literature review. Am J Public Health 1979;69:47–52.
[4] Qualls C, Harris J, Rogers W. Cognitive-linguistic aging: considerations for home health care environments. In: Rogers W, Fisk A, editors. Human factors interventions for the health care of older adults. Mahwah, NJ: Lawrence Erlbaum; 2002. p. 47–67.
[5] Frankel R. Emotion and the physician–patient relationship. Motiv Emotion 1995;19:163–73.
[6] Richmond V, McCroskey J. Immediacy. In: Nonverbal behavior in interpersonal relations. Boston: Allyn & Bacon; 1995. p. 195–217.
[7] Doak C, Doak L, Root J. Teaching patients with low literacy skills, 2nd ed., Philadelphia, PA: JB Lippincott; 1996.
[8] Clark HH, Brennan SE. Grounding in communication. In: Resnick LB, Levine JM, Teasley SD, editors. Perspectives on socially shared cognition. Washington: American Psychological Association; 1997. p. 127–49.
[9] Davidoff F. Time. Ann Intern Med 1997;127:483–5.
[10] Bickmore T, Gruber A, Picard R. Establishing the computer–patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005;59:21–30.
[11] Bickmore T, Picard R. Establishing and maintaining long-term human–computer relationships. ACM T Comput Hum Interact 2005;12:293–327.
[12] Katz MG, Jacobson TA, Veledar E, Kripalani S. Patient literacy and question-asking behavior during the medical encounter: a mixed-methods analysis. J Gen Intern Med 2007;22:782–6.
[13] 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.
[14] National Assessments of Adult Literacy; 2004 (accessed at http://nces.ed.gov/naal/).
[15] Nielsen-Bohlman L, Panzer AM, Hamlin B, Kindig DA. Institute of Medicine. Health literacy: a prescription to end confusion. Committee on Health Literacy, Board on Neuroscience and Behavioral Health. Washington, DC: National Academies Press; 2004.
[16] Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs AMA. Health literacy: report of the Council on Scientific Affairs. J Am Med Assoc 1999;281:552–7.
[17] Paasche-Orlow M, Greene SM, Wagner EH. How health care systems can begin to address the challenge of limited literacy. J Gen Intern Med 2006;21:884–7.
[18] Volandes A, Paasche-Orlow M. Health literacy, health inequality and a just healthcare system. Am J Bioeth 2007;7:5–10.
[19] Ogloff JRP, Otto RK. Are research participants truly informed? Readability of informed consent forms used in research. Ethics Behav 1991;1:239.
[20] Paasche-Orlow MK, Taylor HA, Brancati FL. Readability Standards for Informed-Consent Forms as Compared with Actual Readability. New England J Med 2003;348:721–6.
[21] Wogalter MS, Howe JE, Sifuentes AH, Luginbuhl J. On the adequacy of legal documents: factors that influence informed consent. Ergonomics 1999;42:593–613.
[22] Sugarman J, Lavori PW, Boeger M, Cain C, Edsond R, Morrison V, et al. Evaluating the quality of informed consent. Clin Trials 2005;2:34.
[23] Sudore R, Landefeld C, Williams B, Barnes D, Lindquist K, Schillinger D. Use of a modified informed consent process among vulnerable patients: a descriptive study. J Gen Intern Med 2006;21:867–73.
[24] Bickmore T, Pfeifer L, Yin L. The role of gesture in document explanation by embodied conversational agents. Int J Semantic Comput 2008;2:47–70.
[25] Cassell J, Vilhjálmsson H, Bickmore T. BEAT: The Behavior Expression Animation Toolkit. In: SIGGRAPH' 01; 2001.p. 477–86.
[26] Informed Consent form template for genetic repository research (accessed at http://ccr.coriell.org/Sections/Support/NIGMS/Model.aspx?PgId=216).
[27] Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid Estimate of Adult Literacy in Medicine: a shortened screening instrument. Fam Med 1993;25:391–5.
[28] Lindau S, Basu A, Leitsch S. Health literacy as a predictor of follow-up after an abnormal Pap smear: a prospective study. J Gen Intern Med 2006;21:829–34.
[29] Mancuso C, Rincon M. Impact of health literacy on longitudinal asthma outcomes. J Gen Intern Med 2006;21:813–7.
[30] Sudore RL, Yaffe K, Satterfield S, Harris TB, Mehta KM, Simonsick EM, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study. J Gen Intern Med 2006;21:806–12.
[31] Lincoln A, Paasche-Orlow MK, Cheng DM, Lloyd-Travaglini C, Caruso C, Saitz R, et al. Impact of health literacy on depressive symptoms and mental health-related: quality of life among adults with addiction. J Gen Intern Med 2006;21:818–22.
[32] Taylor SJ, Bogdan R. Introduction to qualitative research methods. John Wiley & Sons; 1998.

Expert Panel on Weight Loss Surgery: Executive Report Update

George L. Blackburn[1], Matthew M. Hutter[2], Alan M. Harvey[3], Caroline M. Apovian[4], Hannah R.W. Boulton[5], Susan Cummings[6], John A. Fallon[7], Isaac Greenberg[8], Michael E. Jiser[9], Daniel B. Jones[1], Stephanie B. Jones[10], Lee M. Kaplan[11], John J. Kelly[12], Rayford S. Kruger Jr[13], David B. Lautz[14], Carine M. Lenders[4], Robert LoNigro[15], Helen Luce[16], Anne McNamara[17], Ann T. Mulligan[18], Michael K. Paasche-Orlow[19], Frank M. Perna[20], Janey S.A. Pratt[2], Stancel M. Riley Jr[21], Malcolm K. Robinson[14], John R. Romanelli[22], Edward Saltzman[23], Roman Schumann[24], Scott A. Shikora[25], Roger L. Snow[26,27], Stephanie Sogg[28], Mary A. Sullivan[29], Michael Tarnoff[25], Christopher C. Thompson[30], Christina C. Wee[31], Nancy Ridley[32], John Auerbach[32], Frank B. Hu[32], Leslie Kirle[32], Rita B. Buckley[32] and Catherine L. Annas[32]

Rapid shifts in the demographics and techniques of weight loss surgery (WLS) have led to new issues, new data, new concerns, and new challenges. In 2004, this journal published comprehensive evidence-based guidelines on WLS. In this issue, we've updated those guidelines to assure patient safety in this fast-changing field. WLS involves a uniquely vulnerable population in need of specialized resources and ongoing multidisciplinary care. Timely best-practice updates are required to identify new risks, develop strategies to address them, and optimize treatment. Findings in these reports are based on a comprehensive review of the most current literature on WLS; they directly link patient safety to methods for setting evidence-based guidelines developed from peer-reviewed scientific publications. Among other outcomes, these reports show that WLS reduces chronic disease risk factors, improves health, and confers a survival benefit on those who undergo it. The literature also shows that laparoscopy has displaced open surgery as the predominant approach; that government agencies and insurers only reimburse procedures performed at accredited WLS centers; that best practice care requires close collaboration between members of a multidisciplinary team; and that new and existing facilities require wide-ranging changes to accommodate growing numbers of severely obese patients. More than 100 specialists from across the state of Massachusetts and across the many disciplines involved in WLS came together to develop these new standards. We expect them to have far-reaching effects of the development of health care policy and the practice of WLS.

Obesity (2009) **17,** 842–862. doi:10.1038/oby.2008.578

[1]Department of Surgery, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA; [2]Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA; [3]Department of Anesthesiology, Mercy Medical Center/Catholic Health East, Springfield, Massachusetts, USA; [4]Department of Medicine, Boston Medical Center, Boston, Massachusetts, USA; [5]Department of Nursing, South Shore Hospital, Weymouth, Massachusetts, USA; [6]Massachusetts General Hospital Weight Center, Massachusetts General Hospital, Boston, Massachusetts, USA; [7]Blue Cross Blue Shield of Massachusetts, Boston, Massachusetts, USA; [8]Obesity Consult Center, Tufts-New England Medical Center, Boston, Massachusetts, USA; [9]Department of Surgery, Saints Medical Center, Lowell, Massachusetts, USA; [10]Department of Anesthesiology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA; [11]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; [12]Department of Surgery, University of Massachusetts Memorial Medical Center, Worcester, Massachusetts, USA; [13]Department of Surgery, Tobey Hospital, Wareham, Massachusetts, USA; [14]Department of Surgery, Brigham and Women's Hospital, Boston, Massachusetts, USA; [15]Care Management, Tufts Health Plan, Boston, Massachusetts, USA; [16]Consumer Representative, Boston, Massachusetts, USA; [17]Department of Nursing and Surgery, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA; [18]Department of Nursing, Newton-Wellesley Hospital, Newton, Massachusetts, USA; [19]Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA; [20]Health Promotion Research Branch Divison of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland, USA; [21]Patient Care Assessment Division, Massachusetts Board of Registration in Medicine, Boston, Massachusetts, USA; [22]Department of Surgery, Baystate Medical Center, Springfield, Massachusetts, USA; [23]Department of Medicine, Tufts-New England Medical Center, Boston, Massachusetts, USA; [24]Department of Anesthesiology, Tufts-New England Medical Center, Boston, Massachusetts, USA; [25]Department of Surgery, Tufts-New England Medical Center, Boston, Massachusetts, USA; [26]University of Massachusetts Medical School, Worcester, Massachusetts, USA; [27]MassHealth, Boston, Massachusetts, USA; [28]Behavioral Medicine Services and MGH Weight Center, Massachusetts General Hospital, Boston, Massachusetts, USA; [29]Division of Medical Specialties, Lahey Clinic and Massachusetts Coalition for the Prevention of Medical Errors, Burlington, Massachusetts, USA; [30]Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; [31]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA; [32]Betsy Lehman Center for Patient Safety and Medical Error Reduction, Commonwealth of Massachusetts, Boston, Massachusetts, USA. Correspondence: George L. Blackburn (gblackbu@bidmc.harvard.edu)

Received 26 June 2007; accepted 6 September 2007; published online 19 February 2009. doi:10.1038/oby.2008.578

INTRODUCTION

Foreword

Sharp increases in the prevalence of severe obesity (BMI >40 and BMI >50) have continued to fuel demand for weight loss surgery (WLS) (**Figure 1**). In 2004, the Betsy Lehman Center for Patient Safety and Medical Error Reduction (Lehman Center) formed an Expert Panel to assess WLS procedures, identify issues related to patient safety, and develop evidence-based best practice recommendations to address those issues.

The resulting document, published as a supplement in *Obesity* in 2005, set the standard for WLS across the state and well beyond it. The Agency for Healthcare Research and Quality abstracted the report for broad use, and the American College of Surgeons adopted it as the blueprint for its Bariatric Surgery Network Center Accreditation Program. Its recommendations influenced health care policy and medical practice at home and abroad.

Since 2004, the literature on WLS has expanded rapidly. New data have been published; new procedures have been developed; and new issues have been brought to our attention. In Massachusetts, weight loss operations increased from over 2,700 in Fiscal Year 2003 to nearly 3,500 in Fiscal Year 2006 (**Figure 2**). We saw a shift from open to laparoscopic operations, and changes in reimbursement policies.

The safety of WLS continues to be of concern. In response, the Lehman Center reconvened the Expert Panel to update the literature review and evidence-based recommendations developed in 2004. Several new members joined the 2007 Expert Panel as well its task groups. All told, there were two additional task groups, bringing the total from 9 to 11. We separated the Psychology Task Group from Multidisciplinary Evaluation and Treatment, and formed a new group, Endoscopic Interventions, to develop best practice guidelines for that emerging technology. In addition, we changed the name of the Coding and Reimbursement Task Group to Policy and Access to better reflect its focus.

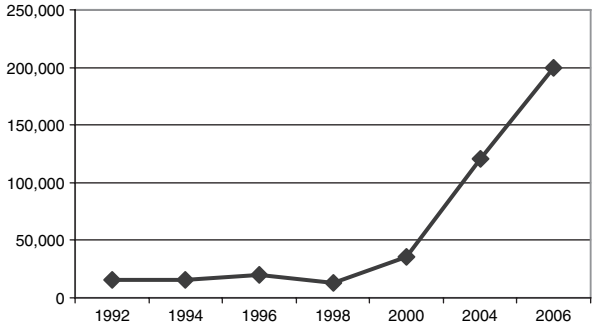


Figure 1 Estimated number of weight loss procedures performed in the United States, 1992–2006 (Adapted from refs. 20,25,36).

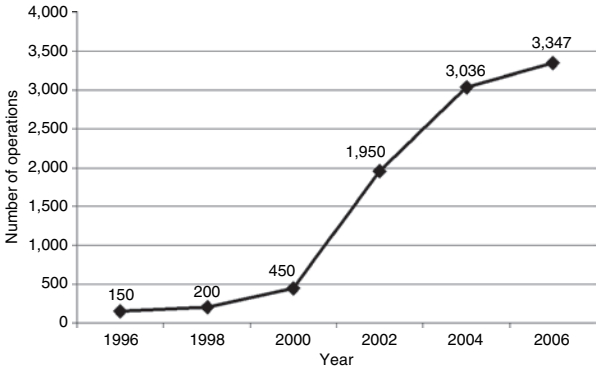


Figure 2 The number of weight loss operations performed in Massachusetts, 1996–2006 (Department of Public Health).

The charge to the 2007 Expert Panel was to update the evidence-based best practice recommendations for WLS developed 3 years ago. Toward that end, its members reviewed weight loss surgical procedures, analyzed the medical literature published since 2004, recommended specific steps to reduce medical errors and improve patient safety, developed credentialing and training standards, identified best practices, and established clinical guidelines and directions for future research.

What follows is a comprehensive evidence-based update to the original best practice recommendations. As with the first report, we hope that these guidelines will have far-reaching effects on clinical practice and health care policy, not only in the Commonwealth, but also nationwide. We hope that they will equalize access and reduce variability in performance and outcomes. Ultimately, our objective is to improve the safety of WLS in the state of Massachusetts and protect the well-being of patients who undergo it.

More than 100 individuals created this report. I express my deepest appreciation to the Expert Panel and task group members for the monumental work that went into this project. I especially thank George Blackburn, Chair, Matt Hutter, Vice Chair, Frank Hu, our clinical epidemiologist, and Rita Buckley, our librarian and medical editor, for their continued leadership and commitment to this project. Last but not least, I thank the Department of Public Health and Betsy Lehman Center staff, especially our project manager, Leslie Kirle, and Katie Annas for their diligent efforts in coordinating and facilitating the work of this project.

Preface

Overwhelming data demonstrate a reduction in known disease risk factors and improvements in health after WLS (1–3). Recent studies also indicate that WLS confers a survival advantage on patients who undergo it compared with community controls (1,2). Landmark findings from the Swedish Obese Subjects study show an estimated 28% reduction in the adjusted overall mortality rate in the surgical groups compared with conventionally treated controls (4).

Similar outcomes have been cited in other reports. A collaborative research project in Utah compared 7,925 gastric bypass patients with the same number of age-, gender-, and BMI-matched controls. Data showed that the rate of death from all diseases was 52% lower in the surgery group than in the control group ($P < 0.001$) (ref. 5). In a case study that compared 821 obese patients who received laparoscopic adjustable gastric banding (LAGB) with 821 controls treated with medical

therapy, Favretti *et al.* (6) found a statistically significant survival difference in favor of the surgically treated group.

Perry *et al.* (7) compared a cohort of extremely obese Medicare beneficiaries who underwent WLS to a similar cohort of extremely obese Medicare beneficiaries who did not. At the 2-year follow-up, younger (<65 years old) and older patients (≥65) in the surgical group had significantly reduced mortality compared with those in the nonsurgical group. Similarly, Sowemimo *et al.* (8) reported 50–85% mortality reductions with surgical intervention.

Decreased total mortality in the Swedish Obese Subjects study (4) surgical groups was primarily due to fewer deaths from cardiovascular disease (especially myocardial infarction) and cancer. In the Utah study (5), significant reductions in mortality were linked to fewer deaths from coronary artery disease (CAD), diabetes, and cancer. These results, which show substantial and consistent evidence of a survival advantage for severely obese patients who undergo WLS, are in line with those of earlier reports by Christou *et al.* (9) and Flum and Dellinger (10). They also confirm previous case series and epidemiologic observations on mortality after weight loss operations in more diverse populations (1,11).

But despite reductions in disease-related mortality after WLS, death rates from other causes, such as accidents and suicides, exceed those of nonsurgery patients. In Adams *et al.* (5), rates of death not caused by disease were 58% higher in the surgery group than in the control group. Reports reveal that a substantial number of severely obese persons have unrecognized presurgical mood disorders or post-traumatic stress disorder, or have been victims of childhood sexual abuse (12).

Data on the association between presurgical psychological status and postsurgical outcomes are limited (13). Although research shows an improved quality of life (QOL) after gastric bypass surgery (14–17), certain unrecognized presurgical conditions may reappear after surgery (18). Some WLS centers recommend that all patients undergo psychological evaluation, and, if necessary, treatment before surgery and psychologically related surveillance postoperatively (12,13,19). Adams *et al.* (5) note the need for further research on the optimal approach to evaluating candidates for WLS, including possible presurgical assessment, psychiatric treatment, and diligent postoperative follow-up.

We know from a substantial body of literature that WLS achieves significant and durable weight loss with minimal mortality or complications. We know that laparoscopy shortens length of stay and makes for a faster, easier recovery (20). Now reliable evidence is starting to accumulate on the survival advantage conferred by WLS on those who undergo it. The field is dynamic (21), with surgical approaches being developed and refined at a rapid pace. Yet technical performance of the operations, critical though it may be, is only one of many challenges.

WLS deals with a uniquely vulnerable population in need of specialized resources and ongoing multidisciplinary care. Timely best practice updates are critical to identify new risks, develop strategies to address them, and optimize treatment of WLS patients. As before (22), members of this panel have come together to protect patient safety and prevent medical errors with evidence-based standards of care. This update of best practice guidelines is part of our continued efforts to improve the efficacy and safety of WLS procedures.

Background

More than 33% of US adults are classified as obese based on objectively measured weight (23), and one-third of American children are either obese or at risk of becoming so (24). Between 2000 and 2005, the proportion of Americans with a BMI ≥40 increased by 50%, although those with a BMI ≥50 increased by 75% (25). Severe obesity has been growing at the fastest rate for the past 20 years (23,25).

Obesity, particularly abdominal obesity, is associated with increased risk of hypertension, diabetes, hyperlipidemia, sleep apnea, coronary heart disease, and strokes (26,27). In 1998, medical costs attributable to overweight and obesity accounted for 9.1% of total US medical expenditures, and may have reached as high as $78.5 billion ($92.6 billion in 2002 dollars) (28,29). In 2000, there were ~360,000 deaths associated with obesity (30). It has been suggested that in the 21st century, increasing rates of obesity may lead to a decline in overall life expectancy in the United States (31).

METHODS AND PROCEDURES

Update on common WLS procedures

Overview. WLS reduces caloric intake by modifying the anatomy of the gastrointestinal tract via restriction, malabsorption, or a combination of the two techniques. Ensuing changes in the gut–brain axis alter peptides that may regulate appetite and satiety (32) (e.g., ghrelin, glucagon-like peptide, and pancreatic polypeptide). Among the several competing approaches for the management of severe obesity, the general trend is toward combined restrictive–malabsorptive procedures (33). Over the past few decades, the number of weight loss surgeries performed in the United States has increased significantly (34,35). Between 1998 and 2004, weight loss operations rose by 900% to 121,055 (ref. 36). In 2006, the estimated total climbed to 200,000 (refs. 20,25).

Laparoscopic Roux-en-Y gastric bypass (LRYGB) is considered the gold standard operation for long-term weight control in United States (35,37). Rates of RYGB per 100,000 adults rose significantly from 1998 to 2002, from 7.0 to 38.6. This increase may be attributed, in part, to improved surgical techniques, better patient outcomes, and growing popularity of the procedure (38). LAGB is the second most commonly performed operation in the United States. Despite rapid growth in LRYGB and other weight loss procedures, only an estimated 1% of patients who are eligible for WLS receive it in any given year (39).

Common WLS procedures

LRYGB. Gastric bypass involves the creation of a small (20–30 ml) gastric pouch and a Roux limb (typically 75–105 cm) (34) that reroutes a portion of the alimentary tract to bypass the distal stomach and proximal small bowel (**Figure 3**). Following LRYGB, a pleiotropic endocrine response may contribute to improved glycemic control, appetite reduction, and long-term changes in body weight (40). LYRGB also has a profoundly positive impact on obesity-related comorbidities and QOL (41). Other advantages include established long-term effectiveness for sustained weight loss, reduction of comorbidities, minimal risk for long-term nutritional sequelae, and effective relief of gastroesophageal reflux disease (21). LRYGB is not without risks. Common causes of death include pulmonary embolism and anastomotic leaks. Nonfatal perioperative complications include

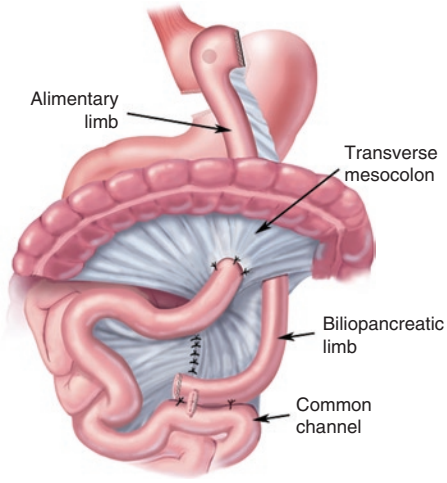


Figure 3 Roux-en-Y gastric bypass (RYGB). RYGB involves the creation of a small (<30 ml) gastric pouch and a Roux limb (typically 75–105 cm) that reroutes a portion of the alimentary tract to bypass the distal stomach and proximal small bowel. (Reprinted with permission of Atlas of Metabolic and Weight Loss Surgery, Jones *et al.* Cine-Med, 2008.) Copyright of the book and illustrations are retained by Cine-Med.

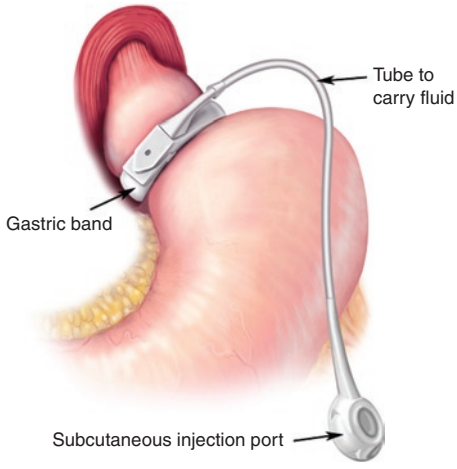


Figure 4 Adjustable gastric band (LAGB). LAGB involves the placement of a band or collar around the upper stomach 1–2 cm below the gastroesophageal junction, thereby creating an ~30 ml upper gastric pouch. The band is imbricated to prevent slippage of the stomach in a retrograde manner through the band. Degree of stomach constriction can be adjusted by modifying the amount of saline injected into a subcutaneous port, which is linked to a balloon within the band. (Reprinted with permission of Atlas of Metabolic and Weight Loss Surgery, Jones *et al.* Cine-Med, 2008.) Copyright of the book and illustrations are retained by Cine-Med.

venous thromboembolism, wound infections, small bowel obstruction, and bleeding. Postoperative gastrointestinal complications include nausea and vomiting, micronutrient deficiencies, (35) and possible weight regain (22).

LAGB. LAGB involves the placement of a band or collar around the upper stomach 1–2 cm below the gastroesophageal junction, thereby creating an ~30 ml upper gastric pouch. Degree of stomach constriction can be adjusted by modifying the amount of saline injected into a subcutaneous port, which is linked to a balloon within the band (34) (**Figure 4**). Parikh *et al.* (42) found that LAGB had fewer

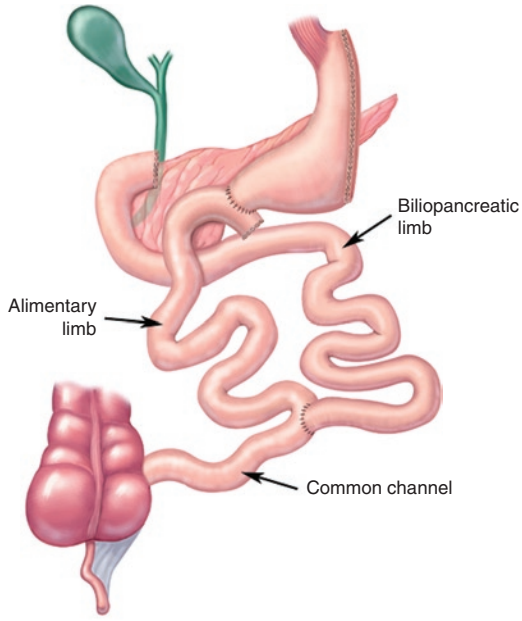


Figure 5 Biliopancreatic diversion (BPD) with duodenal switch. BPD creates malabsorption by maintaining a flow of bile and pancreatic juice through the biliopancreatic limb. The procedure is commonly performed with a duodenal switch in which a distal, common-channel length of small intestine severely limits caloric absorption. The extent of malabsorption is thought to be a function of the length of the common channel. (Reprinted with permission of Atlas of Metabolic and Weight Loss Surgery, Jones *et al.* Cine-Med, 2008.) Copyright of the book and illustrations are retained by Cine-Med.

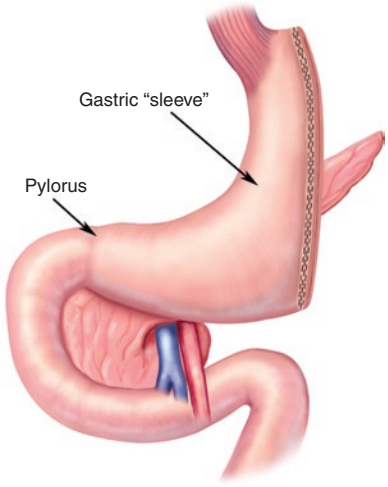


Figure 6 Sleeve gastrectomy (SG). SG consists of the restrictive component of the duodenal switch, a vertical resection of the greater curvature of the stomach creating a long tubular stomach along the lesser curvature. The pylorus and part of the antrum are preserved. (Reprinted with permission of Atlas of Metabolic and Weight Loss Surgery, Jones *et al.* Cine-Med, 2008.) Copyright of the book and illustrations are retained by Cine-Med.

and less severe complications compared with LRYGB or laparoscopic malabsorptive procedures. But other data link LAGB with intermediate and long-term complications (e.g., band erosion or slippage, failure to achieve or maintain weight loss) that require reoperation in up to 20% of patients (43,44).

Biliopancreatic diversion. Biliopancreatic diversion (BPD) creates malabsorption by maintaining a flow of bile and pancreatic juice through the biliopancreatic limb (45). The procedure is commonly performed with a duodenal switch (DS) in which a distal, common-channel length of small intestine severely limits caloric absorption (35). The extent of malabsorption is thought to be a function of the length of the common channel (34). The procedure is combined with a sleeve gastrectomy (SG) in which the greater curvature of the stomach is resected, creating a tubular section along the lesser curvature of the stomach (34) (**Figure 5**). The BPD described by Scopinaro (45) is capable of producing substantial and sustained weight loss, perhaps associated with markedly suppressed ghrelin levels (46). However, increased incidence of stomal ulceration, severe protein-energy malnutrition, diarrhea, and dumping has limited its broad acceptance (21).

Laparoscopic SG. Laparoscopic SG (LSG) is a new purely restrictive treatment for severe obesity. The technique consists of the restrictive component of the DS, a resection of the greater curvature of the stomach over a 45–50 F bougie positioned along the lesser curvature. The pylorus and part of the antrum are preserved, resulting in a lesser curvature-based "restrictive" gastric sleeve (21) (**Figure 6**). Early reports of SG have shown it to be safe and effective (47,48), with marked weight loss and significant reduction of major obesity-related comorbidities (49,50). LSG can be performed as a stand-alone operation or as a bridge to more complex WLS. Following the operation, the stomach empties its contents rapidly into the small intestine, but with little or no vomiting (characteristic of restrictive procedures) (51). There is also a significant reduction in ghrelin associated with resection of the gastric fundus, the predominant area of human ghrelin production (46,52).

Framework for evidence-based recommendations

We divided the 35-member Expert Panel into 11 task groups:

- Surgical Care (53).
- Multidisciplinary Evaluation and Treatment (54).
- Behavior and Psychological Care (55).
- Pediatric/Adolescent (56).
- Anesthetic Perioperative Care and Pain Management (57).
- Nursing Perioperative Care (58).
- Informed Consent and Patient Education (59).
- Policy and Access (Coding and Reimbursement) (60).
- Specialized Facilities and Resources (61).
- Data Collection (Registries)/Future Considerations (62).
- Endoscopic Interventions (63).

Panel members joined one or two task groups, each with an assigned coordinator. Participants were asked to update recommendations from the first Lehman Center report (22) based on the best available evidence, including randomized controlled trials, observational studies, and expert opinion. A medical librarian performed systematic literature reviews for each group. Searches were limited to English-language studies published between April 2004 and May 2007 in MEDLINE, EMBASE, and the Cochrane Library. Some groups also searched other databases (e.g., CINHAL). The process used to extract data, assess the literature, and grade evidence has been previously described (22).

Each task group prepared a critical summary of its literature review and developed updated best practice recommendations (individual studies are published in this issue of *Obesity*) based on the most current evidence. Their reports were reviewed and approved by the Expert Panel. This Executive Report, a summary of key recommendations from all the task groups, was approved by the Expert Panel at its final meeting on 19 July 2007.

RESULTS AND DISCUSSION

Summary of evidence-based recommendations

I. Surgical Care

The Surgical Care Task Group identified >135 papers; the 65 most relevant were reviewed in detail (53). These included randomized control trials, prospective and retrospective cohort studies, meta-analyses, case reports, prior systematic reviews, and expert opinion.

A. Overview

RYGB remains the predominant gold standard WLS in the United States, accounting for 93% of all such operations in 2000 (ref. 64). LAGB is the second most commonly performed procedure (65,66). RYGB is known to safely improve or reverse obesity-related comorbidities and produce significant long-term weight loss (21). Long-term data on weight loss after LAGB vary (42,67,68).

B. Types of WLS

Combination procedures. Combination procedures join a restrictive component (e.g., gastric stapling) with some form of duodenal bypass. They include RYGB, BPD, and DS.

RYGB (open and laparoscopic): Most gastric bypass operations are now done laparoscopically. LRYGB reduces pulmonary, wound, hernia-related complications, and postoperative pain (category B), but may have higher internal hernia rates than RYGB (category C). Weight loss is similar with both approaches (category B).

RYGB modifications: Long-limb RYGB and very very long-limb extend the length of the Roux limb to enhance weight loss. The procedures may increase risk of protein and micronutrient deficiencies (category C); it has yet to be determined whether they produce superior weight loss (category C).

Banded RYGB may be subject to long-term complications related to reintervention, reoperation, and QOL (categories C and D). There is insufficient evidence to make a recommendation (category D). Long-term drawbacks of mini-gastric bypass might include bile reflux and the need for revisional surgery (category C). As with banded RYGB, more data are needed to develop recommendations.

BPD and DS: BPD and DS produce effective weight loss (category B). In patients with a BMI >50, it may be superior to that achieved with RYGB (category C). However, the procedures may increase severe complications (e.g., protein and micronutrient deficiencies) (category B). They also require diligent lifelong patient follow-up (category D).

Restrictive procedures. Restrictive WLS (e.g., LAGB) has no malabsorptive or maldigestive components.

LAGB: Short-term data show promising outcomes with LAGB, but long-term studies raise questions on durability and reoperative rates (category B). We recommend monitoring of long-term data and continuation of current practice patterns, with yearly follow-up of patients (category D).

LAGB should be performed in accredited, multidisciplinary settings by experienced surgeons. They should have advanced laparoscopic skills, including those needed to revise LAGB to an alternate procedure. Barring that, WLS programs should be able to provide appropriate referrals to facilities that can provide that level of care (category D). It is safe for obesity medicine specialists, nurse practitioners,

physician assistants, residents, and bariatric nurse specialists to adjust bands under the supervision of a weight loss surgeon (category D).

LSG: Several short-term studies suggest safe and effective weight loss with LSG (categories B and C), but long-term data on safety and efficacy are needed to recommend the approach as anything other than investigational (category D). If other WLS options are ruled out for reasons of preference or safety, LSG may be considered (category D).

Vertical banded gastroplasty: Vertical Banded Gastroplasty is associated with increased peri- and postoperative complications compared with LAGB. Evidence suggests that it should not be used as a primary surgical treatment for obesity (categories A and B). However, it can be considered when alternative weight loss surgeries are not safe or possible (category D).

C. Revision of WLS

Revisional WLS can address unsatisfactory weight loss or complications after primary WLS. It may also enhance weight loss and further improve comorbidities (category B). Complications, length of stay, and mortality are higher for revisional WLS (category B), but it can be safe and effective when performed by experienced weight loss surgeons (category D).

D. Intraoperative techniques

We recommend the following as standard practice:

- testing of gastrojejunal anastomosis for leaks intraoperatively or within 48 h (category C);
- strong consideration of whether to close mesenteric defects to avoid internal hernia (category C).

E. Patient selection

Emerging issues in patient selection include treatment of those with a BMI >50 and individuals >age 60. Although procedure-specific recommendations for extremely obese patients have yet to be determined (category C), the literature suggests that combination procedures (e.g., RYGB, BPD, DS) lead to greater excess weight loss and resolution of comorbidities than restrictive procedures (e.g., LAGB) (category D).

Age may remain an independent risk factor following WLS (category C), but evidence suggests that WLS can be safe and effective in patients >60 (categories B and C). We recommend that older patients not be denied improvements in health and QOL associated with WLS (category D).

F. Facility and surgeon credentialing standards

The following are best practice updates to guidelines in our prior report (69). These recommendations are all based on category D evidence, unless otherwise noted.

Facilities

- All WLS centers should have, or be in the process of obtaining, accreditation by external review;
- they should meet WLS volume standards specified by credentialing bodies;
- centers with lower volume should be endorsed if risk-adjusted outcomes fall within benchmarks determined by credentialing body data.

Surgeon—credentialing

General requirements: All surgeons seeking WLS credentials for the first time should

- complete an accredited general surgery program and be board-certified, board-eligible, or the equivalent;
- have documented training in the fundamentals of WLS, including pre-, peri-, and postoperative care of the WLS patient.

Open privileges: Most weight loss surgeries are performed laparoscopically. Those who want only open privileges should complete the general credentialing requirements above, and

- be proctored by an experienced weight loss surgeon until proficient;
- have their first 10 cases reviewed by the chief of service and an experienced weight loss surgeon;
- count fellowship cases toward individual surgeon volume requirements.

Full privileges (open and laparoscopic): It is no longer practical to require specific and mandatory experience in open WLS prior to applying for laparoscopic privileges. Those seeking full laparoscopic privileges should complete the general requirements and a laparoscopic fellowship of 50 WLS procedures. As an alternative, they can be proctored for a minimum of 25 cases by an experienced (70) (>200 laparoscopic cases) weight loss surgeon with full privileges. In addition, surgeons should

- have their first 10 cases reviewed by the chief of staff and an experienced weight loss surgeon;
- count fellowship cases toward individual surgeon volume requirements.

Fundamentals of Laparoscopic Surgery certification is also highly recommended for newly trained laparoscopic surgeons.

Surgeon—recredentialing

- Institutions should develop in-house standards for recredentialing based on procedure-specific and risk-adjusted outcomes (benchmarks) rather than volume alone.
- An annual volume of 25 cases may be sufficient if outcomes are within accepted standards, reported to a central database, and performed at an accredited institution.
- Weight loss surgeons should complete at least 12 CME credits related to WLS or obesity every 2 years.

Procedure-specific credentialing. Rapid changes in technologies and techniques warrant disclosure of procedure-specific information to patients, and selection of those with lower risk profiles for the first 25 cases. As part of the educational process, surgeons should disclose

- the type and approximate number of procedures they perform (category D);
- alternative WLS options available (category D);
- risks, potential benefits, and program outcomes (category D).

II. Multidisciplinary Evaluation and Treatment

The Multidisciplinary Care Task Group identified over 150 abstracts related to WLS in general, and to medical, nutritional, and multidisciplinary care in particular; 112 of these studies were reviewed in detail (54).

A. Multidisciplinary care

The American Society for Bariatric Surgery recently changed its name to the American Society for Metabolic and Bariatric Surgery, reflecting growing knowledge that WLS has benefits beyond the treatment of severe obesity. This change expands the scope of multidisciplinary expertise required to provide optimal care for WLS patients. As the nature of multidisciplinary care changes, we recommend

- development of uniform minimum standards of multidisciplinary care for WLS patients (category D);
- further research on the effectiveness of general medical, surgical, anesthetic, nutritional, and psychological aspects of multidisciplinary treatment (category D).

B. Preoperative education and patient selection

Preoperative education allows for more appropriate matching of patients and procedures. It can dispel misperceptions and unrealistic expectations, and help clarify issues related to resolution of comorbid conditions, differences between surgical procedures, and required lifestyle changes after WLS (category D).

C. Operative risk

Higher BMI and medical comorbidities (e.g., obstructive sleep apnea (OSA) and coronary heart disease risk factors) increase operative risk and postoperative complications. We recommend assessment of risk factors (71) in each patient (category C).

Preoperative weight loss. Preoperative weight loss of 5–10% of initial body weight can decrease operation time and may reduce surgical risk. Patients, especially those with a BMI ≥50, should be encouraged to achieve weight loss of 5–10% of initial body weight prior to surgery (category C). Prospective randomized controlled trials are needed to determine optimal preoperative weight loss and improve supervision of preoperative weight reduction (category C).

Medical evaluation. Specific consideration should be given to WLS patients with a history of CAD or DVT/PE, those who are current smokers, and those with known or suspected abnormal liver function. *Helicobacter pylori* testing and treatment may also be useful, but more evidence is needed to determine its importance. Other risk factors include postprandial hypoglycemia, chronic renal disease, and HIV.

CAD: Patients with a history of CAD should receive preoperative assessment of cardiovascular conditions as indicated (category C). Those with stable or suspected CAD should receive perioperative β blockade unless contraindicated (category C).

Abnormal liver function: Patients with known or suspected liver disease should be evaluated to assess severity of cirrhosis and/or portal hypertension (category B). Intraoperative liver biopsy at the time of surgery may be useful for diagnosis and assessment of liver disease (category C). WLS is not recommended in patients with Child's Class C cirrhosis (category B).

DVT/PE: We recommend perioperative use of anticoagulants and sequential compression devices to reduce the risk of DVT/PE unless clinically contraindicated (category B). In patients with increased risk of DVT/PE extended prophylaxis should also be considered (category D).

Smokers: Smokers should be strongly encouraged to stop smoking prior to WLS (category B). Smoking cessation advice and treatment should be available at the institution or through the WLS program (category D).

Hypoglycemia: Patients with known or suspected hypoglycemia should be assessed by an endocrinologist prior to WLS. In that gastric bypass surgery is already being used to treat diabetes (72), purely restrictive procedures should be considered for WLS patients with a documented history of hypoglycemia (category D).

Chronic renal disease: Pre- and postoperative monitoring of renal function is recommended in patients with diabetes and hypertension (categories A and B). Patients with significant renal disease should be evaluated by a nephrologist prior to WLS (category D). Special consideration should be given to pre- and postoperative monitoring of fluid and intravascular volume status (category A).

HIV infection: Patients with HIV should be evaluated by an infectious disease specialist prior to WLS (category D). Special consideration should be given to preoperative assessment of viral loads, CD4 counts (category D), and weight gain from antiretroviral medications (category D).

D. Nutrition

Preoperative and postoperative micronutrients. WLS, especially malabsorptive procedures, can cause multiple micronutrient deficiencies. Patients should be monitored pre- and postoperatively for deficiencies in vitamin D, thiamine, calcium (including PTH), iron, vitamin B12, and folic acid, with repletion as indicated (categories A, B, and C).

E. Exercise and physical activity

WLS patients should be encouraged to increase pre- and postoperative physical activity (category D) and low-to-moderate

intensity exercise (category A). Guidance and periodic monitoring should be used to help WLS patients remain physically active (category D).

F. Pregnancy

WLS should not be performed in patients who are known to be pregnant; we strongly recommend preoperative testing for women of childbearing age (category C). Patients should be strongly counseled to not get pregnant for at least 18 months after surgery (category C).

G. Post-WLS body contouring

Post-WLS body contouring is an emerging field. The task group identified and reviewed in detail 80 relevant articles, ranging from case reports and expert opinion to prospective randomized trials.

Insurance coverage. Body contouring should generally be reserved until a patient has achieved a stable weight. This usually happens at 18 months (or more) after WLS. There are no widely accepted guidelines for insurance coverage of body contouring after substantial weight loss. We recommend third party coverage of excess skin excision, if medically indicated (category D).

Surgeon criteria. Body contouring should only be performed by board-eligible or board-certified surgeons with training and experience in the relevant procedures (category D).

III. Behavioral and Psychological Care

The Behavioral and Psychological Care Task Group identified 17 papers; the 13 most relevant were reviewed in detail (55). These included randomized controlled trials, prospective and retrospective cohort studies, meta-analyses, case reports, and prior systematic reviews.

A. Patient selection and preoperative evaluation

WLS patients are an emotionally vulnerable population. All candidates for WLS should undergo psychosocial evaluation by a credentialed expert in psychology and behavior change (category C). Evaluations should be carried out by a social worker, psychologist, or psychiatrist with a strong background in the current literature on obesity and WLS, and some experience in the pre- and postoperative assessment and care of WLS patients (category D). Though not essential, it is preferable that the evaluator be on staff or affiliated with the WLS center to facilitate communication, maintain the support network, and provide continuity of care (category D).

To address long-term complications, mental health resources should be made available to patients beyond the standard postoperative period of 6 months (category D). This recommendation can be met in a variety of ways (e.g., staff mental health professional, referral network).

Mental illness, including eating pathology, should not necessarily be a contraindication to WLS. Evaluations should determine the degree to which mental illness, including eating pathology, may jeopardize the safety or efficacy of WLS (category C). They should be used to identify patients in need of preoperative psychosocial intervention, and develop recommendations on if, how, and when to best address significant psychosocial risk factors (category C).

Psychological assessment and support have become essential components of multidisciplinary care in WLS. We recommend that organizations that provide education on obesity and WLS (e.g., North American Association for the Study of Obesity) offer continuing education units to mental health providers. This will facilitate the development of continuing education standards for mental health specialists in the fields of obesity and WLS (category D).

B. Binge eating disorder

Binge eating disorder in patients seeking WLS is clinically important, especially in the long-term. It should be taken into account in the development of treatment plans. Assessment should be done in a standardized, empirically validated way (e.g., screening with EDE-Q and follow-up with a brief, standardized interview based on DSM-IV-TR criteria) (category C). The disorder should not be considered a contraindication for WLS, but rather, a potential complication that may need to be addressed before or after surgery to ensure optimal outcome (category C).

Patients should know that eating pathology can recur after WLS, and that they may need professional help to deal with recurring patterns of binge eating. This disorder should be included in the informed consent process and as part of the WLS program's standard educational component (category C).

C. Night eating syndrome

In that there is no clear evidence that night eating syndrome has any impact on surgical outcome, the condition should not be considered a contraindication for WLS. Rather, it should be seen as a potentially complicating factor that may need to be addressed before or after surgery to ensure optimal outcome (category D).

D. Emotional eating

Data are insufficient to make recommendations on the assessment and treatment of emotional eating. As with night eating syndrome, the issue should be considered a potentially complicating factor that may need to be addressed before or after WLS to assure optimal outcome (category D).

E. Substance abuse

Findings on the prevalence of substance abuse among those seeking WLS are conflicting, and there are few studies on the subject. Evidence is insufficient to conclude that the problem is a frequent one after WLS. Further research is needed to establish the prevalence of substance abuse after WLS as well as its predictors, its relation to surgical outcome, and effective treatment approaches (category D).

F. Psychotropic medications

Data indicate significantly higher use of psychotropic medications in WLS patients compared with the general population.

Further research is needed to determine the relation between various psychotropic medications and their impact on postoperative weight loss and psychosocial adjustment (category D).

The effects of WLS on the dissolution, absorption, and clinical response to psychotropic drugs are not well understood. For this reason, we recommend close postoperative monitoring of WLS patients, especially after gastric bypass (category D).

G. Future research needs

The needs of future research are

- adequately powered and controlled prospective trials that examine the relation between psychosocial factors and surgical outcomes;
- randomized controlled trials on the effectiveness of treatments to reduce the impact of psychosocial risk factors on outcomes.

IV. Pediatric/Adolescent

The Pediatric/Adolescent WLS Task Group identified >1,085 papers; 186 of the most relevant were reviewed in detail (56).

A. Types of surgery

RYGB is considered a safe and effective option for extremely obese adolescents as long as appropriate long-term follow-up is provided (category B). The adjustable gastric band has not been approved by the FDA for use in adolescents, and therefore, should be considered investigational. Off-label use can be considered, if done in an IRB-approved study (category C).

BPD and DS procedures cannot be recommended in adolescents. Current data suggest substantial risks of protein malnutrition, bone loss, and micronutrient deficiencies. These nutritional risks are of particular concern during pregnancy. In addition, several late maternal deaths have been reported (category C).

SG should be considered investigational; existing data are not sufficient to recommend widespread and general use in adolescents (category D).

B. Comorbidities

Strong indications for WLS in adolescents include established type 2 diabetes (category B), moderate to severe OSA with AHI ≥15 (category C), severe and/or progressive NASH (category C), and pseudotumor cerebri (category C). Other indications for WLS in adolescents include mild OSA, mild NASH, hypertension, dyslipidemia, and significantly impaired QOL (categories C and D).

All adolescents with obesity should be formally assessed for depression. If found to be depressed, they should be treated prior to WLS (category B). The presence of eating disturbances is not an exclusion criterion for WLS, but adolescents with such disorders should be treated prior to surgery (category B).

C. Patient selection

When combination procedures are used in adolescents, physical maturity (completion of 95% of adult stature based on radiographic study) should be documented. In most cases, this criterion will limit surgery to children over age 12 (category D). Psychological maturity—demonstrated by understanding of the surgery, mature motivations for the operation, and compliance with preoperative therapy—should be assessed prior to WLS (category D).

BMI cutpoints in children and adolescents who meet other criteria should be ≥35 with major comorbidities (i.e., type 2 diabetes mellitus, moderate to severe sleep apnea (AHI >15), pseudotumor cerebri, or severe NASH) and ≥40 with other comorbidities (e.g., hypertension, insulin resistance, glucose intolerance, substantially impaired QOL or activities of daily living, dyslipidemia, sleep apnea with AHI ≥5) (categories B and C).

There are no data available to suggest that prolonged preoperative weight management programs are of benefit to adolescents who undergo WLS. However, children and adolescents should demonstrate the ability to comply with treatment regimens and medical monitoring before WLS. In many cases, consistent attendance in a prolonged weight management program will provide important assurance of postoperative compliance (category D).

Individuals with mental retardation vary in their capacity to demonstrate knowledge, motivation, and compliance; they should, therefore, be evaluated for WLS on a case-by-case basis. For these children, we suggest including an ethicist on the multidisciplinary evaluation team (category D).

Others who should be screened on a case-by-case basis include patients with syndromic obesity, endocrine disorders, obesity that appears to be related to the use of weight-promoting medications, and those in whom obesity cannot be controlled through medical interventions and/or carefully designed environmental and behavioral management. Very limited information is available about the outcomes of WLS for such patients (category D). Patients with uncontrolled psychosis (presence of hallucinations and delusions), bipolar disorder (extreme mood lability), or substance use disorders can be considered for WLS on a case-by-case basis after they have been in remission for 1 year (category C).

D. Team member qualifications

Although few hospitals have sufficient volume for a stand-alone pediatric surgical center, the ideal WLS team should include a minimum of four or five professionals who are colocated and have at least one preoperative face-to-face meeting to prepare a treatment plan for each patient (category D). Staff should include

- surgeon—experienced adult bariatric surgeon or pediatric surgeon with bariatric fellowship or the equivalent experience;
- pediatric specialist—internist or pediatrician with adolescent and obesity training and experience;
- registered dietician—with weight management certificate and experience in treating obesity and working with children and families;
- mental health professional—with specialty training in child, adolescent, and family treatment, and experience treating eating disorders and obesity;

- coordinator—RN, social worker, or one of the other team members who has the responsibility of coordinating each child or adolescent's care and assuring compliance and follow-up.

The ideal setting would be in an adult/pediatric hospital, with a pediatric program partnered with an adult program that has full access to pediatric specialists (category D). A comprehensive family-based evaluation should be provided to parents seeking surgery for their adolescent children (category D).

E. Risks and outcomes

Early WLS may reduce obesity-related mortality and morbidity. However, early timing must be weighed against the patient's possible psychological immaturity and the risk of decreased compliance and long-term follow-up (category C). All adolescents undergoing WLS should be included in prospective longitudinal data collection to improve the evidence base for evaluating the risks and benefits of WLS in this age group (category D).

Emphasis on compliance strategies, careful monitoring of vitamin and mineral intake, and periodic laboratory surveillance to detect deficiencies is crucial (category D). Adolescent girls are particularly vulnerable to nutritional deficiencies; this group is at substantial risk of developing iron deficiency anemia and vitamin B deficiencies during menstruation and pregnancy (category C), and should receive special attention.

Risk of pregnancy increases after WLS. All female adolescents should be informed about increased fertility following weight loss, and possible risks associated with pregnancy during the first 18 months after surgery. They should be counseled to avoid pregnancy during this period, and offered contraception (category D). In addition to risks for deficiencies of iron, calcium, and vitamin B12 after WLS, adolescents may also be at particular risk for osteopenia and thiamine deficiency (category C).

F. Informed consent

Informed assent by the adolescent should be obtained separately from the parents to avoid coercion (as in other pediatric chronic illnesses that require surgical intervention) (category D). The patient's knowledge of the risks and benefits of the procedure and the importance of postoperative follow-up should be formally evaluated to ensure true informed assent (category C). The parental permission process should include discussion of the risks of adult obesity (category C), available medical treatments (category B), surgical alternatives, and the specific risks and outcomes of the proposed WLS in the proposed institution.

V. Anesthetic Perioperative Care and Pain Management

The Anesthetic Perioperative Care and Pain Management Task Group's literature search yielded 1,788 abstracts, with 162 potentially relevant titles. Following full-text evaluation of the latter, 45 articles were reviewed in detail. Best practice recommendations integrate the latest research on obesity and collaborative multidisciplinary care (57).

A. Preoperative evaluation and preparation

Mandatory polysomnography for WLS patients has been proposed (category C). However, we recommend that it be used in selected patients as indicated. When uncertain of the indication for such testing, clinical assessment should be supplemented to include gender, waist-to-hip ratio, and neck circumference (category B). Preoperative CPAP treatment should be strongly considered for patients with a polysomnography diagnosis of moderate to severe OSA (categories B and C). We recommend smoking cessation at least 6 weeks prior to surgery (category C); the WLS program should provide active support to help patients achieve and sustain compliance (category D).

B. Intraoperative management

Induction and emergence. The ≥30° reverse Trendelenburg position prolongs the ability of severely obese patients to tolerate apnea during induction of (category A), and emergence from (category D), anesthesia. CPAP of ~10 cm H_2O may be considered during preoxygenation to prolong nonhypoxic apnea (category A). Intubating laryngeal mask airway devices provide an alternative mechanical approach to securing the airway (categories A and B), and may also improve success when attempting ventilation prior to securing the airway. Intubating laryngeal mask airway devices should be included among the alternative airway management devices immediately available in the operating room (categories A and B).

Maintenance of anesthesia. Preoperative oral administration of clonidine (an α-2 agonist) to obese patients with OSA is associated with reduced anesthetic requirements as well as reduced intra- and postoperative opioid requirements. Its use may be considered unless medically or surgically contraindicated (categories A and C).

Intraoperative oxygenation. Several methods to improve intraoperative oxygenation during WLS have been evaluated. We recommend initial treatment of intraoperative hypoxemia with recruitment maneuvers and positive end-expiratory pressure while monitoring their potential hemodynamic effects (categories A and B).

Other interventions. Postoperative nausea and vomiting in laparoscopic WLS patients is related to the volume and rate of intraoperative fluid replacement. To reduce postoperative nausea and vomiting, we recommend maintenance of euvolemia (category C).

Intraoperative drug dosing. Pharmacodynamic studies in severely obese patients have suggested optimal dosing requirements for different neuromuscular blocking agents. Cisatracurium and rocuronium should be dosed according to ideal body weight during standard induction of general anesthesia (category A). The muscle relaxant succinylcholine should be dosed at 1 mg/kg total body weight (category A). For target controlled

infusion (not yet approved in the United States), propofol dose should be calculated to more closely reflect total body weight (category C).

C. Postanesthesia care

Positive outcomes have been reported with early treatment of postoperative hypoxemia employing noninvasive positive pressure ventilatory support (NIV) in nonobese, non-OSA patients at high risk of respiratory failure. A joint decision between the surgeon, anesthesiologist, respiratory therapist, and nurse should determine NIV use on selected WLS patients (categories A, B, and C). LRYGB and LAGB have been performed safely as 23-h stay and outpatient procedures. However, patients with OSA should not be considered candidates for outpatient WLS (category C); we recommend adherence to the American Society of Anesthesiologists Practice Guidelines for the Perioperative Management of Patients with OSA (category C).

Postoperative pain management. Based on new evidence of efficacy and safety specific to WLS patients, we recommend use of opioid sparing multimodal analgesic strategies, including local anesthetic wound infiltration and nonsteroidal anti-inflammatory medications, unless contraindicated (categories A and C). Solutions for thoracic epidural pain management in OSA patients should be opioid-free to reduce the risk of respiratory depression (category C).

D. Credentialing

No evidence indicates that specific credentialing of anesthesia personnel for WLS will improve patient safety or outcomes. We recommend the selection of a board-certified anesthesiologist to coordinate intradepartmental staff education and proctoring to establish proficiency. This individual will also serve as an interdepartmental liaison to WLS programs and the multidisciplinary WLS care team (category D).

E. Medical error reduction and systems improvement

Optimal outcomes require unimpaired intra- and perioperative multidisciplinary communication among WLS caregivers (category D). Development of perioperative care pathways for patients with OSA is at an early stage (category C) and needs further refinement for WLS patients.

F. Future research needs

Research is needed in the following areas:

- the role and parameters of preoperative OSA treatment for perioperative safety outcomes in WLS;
- intra- and perioperative drug dosing, including prophylactic antibiotic tissue pharmacokinetic assessment;
- appropriate use of α-2 agonists in the perioperative care of WLS patients;
- strategies for intra- and postoperative glycemic management;
- impact of advanced monitoring of anesthetic effects on outcomes;
- evidence-based postoperative care guidelines for WLS patients with OSA;
- optimal anesthetic care for WLS patients with increased BMI, age, and quantity and severity of comorbidities;
- impact of an organized multidisciplinary care team on WLS safety outcomes;
- effect of surgical and overall care team pathways to decrease and/or treat perioperative anesthetic and surgical complications.

VI. Nursing Perioperative Care

A systematic review of MEDLINE, nursing journals, and the CINHAL database for nursing and allied health literature identified >54 papers; the most relevant were reviewed in detail. Recommendations are based on published evidence and the consensus of the Task Group members (58).

A. Planning and communication

Effective communication between all members of the health care team is paramount in the delivery of quality care. It requires sufficient time for the collection of information from patients, site verification in the operating room, timely and concise reporting of symptoms, and the "repeating back" of information exchanged between team members. To optimize communication, we recommend

- continued development of clinical pathways (category D);
- an Advanced Practice Nurse or Clinical Bariatric Nurse Specialist on staff in WLS programs (category D);
- development and fostering of good communication skills between patients and practitioners and between members of the health care team (category D);
- promotion of collaboration between nurses, physical therapists, discharge planners, social workers, nutritionists, and facilitators of support groups (category D).

B. Perioperative management

Unit-specific triage based on individual comorbidities can promote patient safety (category D). We also recommend use of the Association of Perioperative Registered Nurses Bariatric Surgery Guideline (category D) and the American Society of Anesthesiologists Practice Guidelines for the Perioperative Management of Patients with OSA (category C). Preferably, a dedicated operative team of nurses and surgical technicians should regularly assist in WLS procedures (category D).

Preventing complications. Risk of venous thromboembolic events after gastric bypass is significant. Other postoperative complications include those associated with monitoring of fluid balance, hypoxemia, anastomotic leak, tachycardia, peripheral nerve injury, and risk of skin irritation, infection, ulceration in skinfolds, and decubitus ulcers. We recommend ambulation on the day of surgery, and deep breathing/coughing (category D); careful positioning to decrease risk of peripheral nerve injury (categories C and D); and education of emergency

department staff on early and late complications in WLS patients (category D).

Perianesthesia. Obese patients present with distinct respiratory care considerations. They should be closely monitored for rapid oxyhemoglobin desaturation and respiratory depression after extubation. Facilities should reference the Association of Perioperative Registered Nurses Bariatric Surgery Guideline (category D) and educate staff on pulmonary pathophysiology in obese patients (category D).

Postoperative analgesia. The goal of postoperative pain management is to promote participation in activity, ambulation, incentive spirometry, deep breathing, and coughing. Nursing staff should consult with a pharmacist on equianalgesic agents and dosing (category D), and use multimodal, opioid-sparing strategies to keep patients comfortable (category D).

C. Patient and staff safety

WLS patients move through many areas of hospitals for tests and procedures. Facilities should review each area and its equipment to make certain that they can accommodate extremely obese patients. The weight capacity of tables, beds, stretchers, and wheelchairs should be clearly marked (categories C and D). A comprehensive ergonomics program, including lifting and transferring equipment, should be used to prevent patient handling injuries (category B). A designated nurse or back injury resource nurse should coordinate equipment selection, maintenance, staff training, and reporting (category D).

D. Outpatient postoperative nursing follow-up

Dehydration, pulmonary embolisms, and anastomotic leaks are the serious conditions most likely to occur in the early discharge phase. Later complications can include hyperinsulinemic hypoglycemia, metabolic bone disease, problems with redundant skin, nutritional deficiencies, suboptimal weight loss, issues with psychosocial adjustment, and pregnancy.

Medications and vitamin supplements should be reviewed at each postoperative outpatient visit (categories C and D). Nurses should be knowledgeable about possible late complications, know how to support patients, and be prepared to make referrals to appropriate caregivers (category D). WLS patients should be encouraged to continue treatment through ongoing WLS support groups and networks (categories A and D).

E. Credentialing

The American Society for Metabolic and Bariatric Surgery has developed national certification criteria for Clinical Bariatric Nurse Specialists. We recommend certification (category D).

F. Future research needs

Studies are needed in the following areas:

- clinical pathways for WLS, including emergency departments;
- comprehensive ergonomics programs;
- teach-to-goal educational methods for pre- and postoperative education;
- program retention tools and outcome measures;
- nursing research and involvement in pediatric WLS programs.

VII. Informed Consent and Patient Education

This Task Group's literature search identified 120 papers, 38 of which were reviewed in detail. No articles were specific to informed consent and WLS. Recommendations are extrapolated from, and supported by, existing data (59).

A. Content

Risks/complications. Informed consent should include realistic risk estimates that take into account patient factors (category C) and relevant institutional and health provider characteristics that might affect risk (e.g., experience and outcomes for specific WLS procedures) (category B). Short- and long-term risks and complications, and the potential for unknown or unforeseeable long-term risks, should be discussed (category D).

Benefits/effectiveness. Patients should receive realistic estimates of short- and long-term weight loss, including the potential for weight regain and modest benefits (category B). They should also be informed if long-term data (>5 years) are unavailable (category D).

They should be advised of the long-term health benefits of weight loss produced by WLS (category B), but also be made aware that not all pre-existing medical and psychosocial consequences of obesity (including eating disorders) will improve with WLS (category C). Candidates for WLS should be given realistic estimates for health outcomes if they decline surgical treatment (categories B and C), and be advised of known factors and interventions that might optimize benefits (category D). Informed consent and education should consider patient expectations, the value placed on different outcomes, and the risks each candidate is willing to accept. It should also address unrealistic expectations or other misconceptions patients might have (category C).

Consequences. Patients should be advised of required behavioral and dietary changes and other reasonable and foreseeable consequences of WLS that could affect health or QOL in a substantive way, e.g., gastrointestinal symptoms, cosmetic effects, nutritional restrictions (category D).

B. Alternative treatments

Patients should be advised about alternative WLS procedures and nonsurgical treatment options (e.g., medical and behavioral) (category C). They should be informed about them even if they are not available through the consenting health provider or institution (category C).

C. Patient comprehension

Each patient should have their comprehension of the risks, benefits, consequences, and alternatives to WLS evaluated

(category C). Confirmation of comprehension should be included as a protection for patients engaged in the informed consent process (category C).

D. Future research needs

Future research is needed to better identify factors that affect short- and long-term outcomes so that patients can be cited appropriate and individualized outcome information. Research should focus on important gaps in knowledge on outcomes and consequences of WLS, and the different approaches that facilitate patient understanding of, and decision making about, WLS.

VIII. Policy and Access (Coding and Reimbursement)

The Policy and Access group identified 51 publications in its literature search; the 20 most relevant were examined in detail (60). These included reviews, cost–benefit analyses, and trend and cost studies from administrative databases.

A. Policy and access

Access disparities (all category D). Public health policy should be aligned with long-term goals for the treatment of severe obesity. Barriers to WLS in populations with high prevalence of severe obesity should be identified and eliminated, and there should be uniform standards of coverage for all WLS candidates. We recommend advocacy for increased access to WLS for underserved regions and population groups; support for community-based efforts to fight health disparities; and public education about the obesity epidemic and the risks/benefits of WLS.

Childhood obesity (categories C and D). Sharp increases in childhood obesity lend urgency to the need to address the problem (category C). Policy initiatives to identify pediatric and adolescent populations most likely to benefit from surgical treatment of obesity are needed. Surgical treatment should be considered a potentially effective option for appropriately selected individuals, and there should be uniform standards of coverage for adolescent patients. We need to educate legislators, community leaders, and other stakeholders on the costs and benefits of WLS for extremely obese adolescents, and leverage opportunities for collaboration between teachers, parents, and community leaders (category D).

Insurance policies (category A, B, C, and D). Controversial issues include required documentation of prior weight loss attempts through more conservative means; access to WLS for those with a BMI of 35–40 and obesity-related comorbidities; and proof of extreme obesity for at least 5 years. We recommend

- routine examination of weight loss histories during behavioral evaluation to determine whether additional attempts at nonsurgical weight loss are advisable;
- coverage of WLS for those with a BMI of 35–40 and comorbid conditions that require ongoing treatment (e.g., CPAP, medication);
- research to characterize weight loss histories of surgical candidates, and explore the relation between dieting history and postoperative outcomes;
- ongoing collection and dissemination of data on WLS costs, risks, and benefits;
- collaborative efforts between government, industry, and other stakeholders to promote safe and effective delivery of WLS.

Cost-effectiveness issues. Obesity is linked to higher health care costs than smoking or drinking, and plays a major role in disability (category B). Accurate short- and long-term cost savings (and risk/benefits) for employers and insurance companies need to be collected and disseminated. Clinical pathways that reduce unnecessary costs to providers should also be developed (category D).

Innovation, evidence-based medicine, and cost containment. The application of standard cost-containment policies to surgical innovations may stifle new developments. We recommend the use of evidence-based medicine to both guide clinical decisions and show reasonable trends for health care cost containment (category C).

Legislation. We need to keep legislators apprised of the personal and economic costs of obesity in the communities they serve. Dissemination of evidence-based information on the risks, benefits, and cost-effectiveness of WLS can bring these issues to their attention (categories C and D).

Stigma (all category D). The highest BMI groups are the fastest growing and the most stigmatized. To address this problem, we recommend targeted education campaigns; community-level public information/education; and sensitivity training for hospital personnel. Hospitals should also acquire obese-appropriate products (e.g., gowns, chairs, commodes).

B. Coding and reimbursement

Centers for Medicare and Medicaid Services. Centers for Medicare and Medicaid Services allows national coverage for RYGB (open and laparoscopic), LAGB, and BPD with DS (open and laparoscopic). Nationally covered procedures and new 2006 CPT codes are available.

C. Potential pathways to new codes

Category III and S codes. CPT category III Codes are a temporary set of tracking codes used to identify new and emerging technologies. CPT category III codes (T codes) support data collection on new services and procedures. CPT category III codes may be converted to CPT category I codes if the FDA and CPT Editorial Panel approve the clinical efficacy of the particular service or procedure. Blue Cross/Blue Shield and other commercial payers have developed the category of S codes, which were added to HCPCS Level II to report drugs, services, and supplies. S codes are typically used in conjunction with a nonspecific CPT code.

Medicare does not recognize or reimburse for services reported under S codes, and may or may not reimburse for CPT category III codes, depending on the service or procedure.

Individual commercial insurers may or may not reimburse for S codes or CPT category III codes as medical policies and reimbursement polices are specific to each insurer.

D. Issues and recommendations

Alignment of reimbursement policies with clinical objectives. Reimbursement policies should reflect the importance of comprehensive, multidisciplinary care. There should be full coverage for medical, nutritional, and psychological preoperative evaluation as well as pre-, peri-, and postoperative care required by insurers (category D).

CPT codes for WLS and related clinical services (all category D). CPT codes for WLS should be updated to reflect current practice. New CPT category I codes should be requested and approved as evidence accumulates in favor of new procedures (e.g., vertical SG, endoscopic interventions). T codes should be considered for evolving technologies, and procedures. The use of T codes may create a pathway for reimbursement by supporting consistent data collection and development of evidence. Evidence indicating that a promising technology or new procedure leads to improved health outcomes could support conversion of category III codes to category I codes. There should be support for the development of appropriate CPT codes for each component of multidisciplinary care (e.g., exercise therapy, pre- and postoperative support groups).

Data collection, tracking, and reporting systems. There are several national data collection, tracking, and reporting databases (see Data Collection) (62) as well as proprietary systems. We recommend standardized collection, tracking, and reporting of tiered and risk-adjusted data (category D).

IX. Specialized Facilities and Resources

The Specialized Facilities and Resources Task Group identified 1,647 papers in its literature search; the 46 most relevant were reviewed in detail (61). These included randomized control trials, prospective and retrospective cohort studies, meta-analyses, case reports, prior systematic reviews, and expert opinion.

A. Personnel

All medical and support staff must be adequately trained and credentialed as specified in the following task group reports: Surgical Care (53), Anesthesia Perioperative Care and Pain Management (57), Behavioral and Psychological Care (55), and Nursing Care (58). A team of dedicated medical specialists—fully aware of the problems and sensitivities of patients with severe obesity—should be readily available, and all personnel (including ancillary and nonclinical staff) should have obesity-specific education focused on sensitivity training.

B. Equipment

All facilities performing WLS, including pediatric WLS centers, require the same equipment. We strongly recommend that WLS centers have well-defined plans for the evaluation and treatment of post-WLS surgery patients with potential complications who cannot fit into available diagnostic equipment. Recommended equipment includes the following.

Ancillary

- Wide wheelchairs, stretchers, and walkers.
- Wide BP cuffs.
- Biphasic defibrillators.
- Size-appropriate sequential compression devices.
- Emergency airway equipment.
- Wide examination tables bolted to the floor.
- Scales of appropriate size and capacity.

Operating room. Specially equipped operating room and ancillary equipment should be available to support patients with severe obesity, including

- an automated extra-wide operating table with appropriate weight capacity;
- extra-long abdominal instrument sets;
- appropriately sized retractors;
- 43–46 cm laparoscopes.

Radiology equipment. Special diagnostic and interventional equipment is required to support and accommodate WLS patients. Such equipment should include

- CT scanners with 400 lb weight capacity;
- MRI magnet with 400 lb weight capacity;
- fluoroscopic equipment with 300 lb capacity that can study patients in a standing position with high beam voltages;
- interventional facilities available 24 h a day, 7 days a week.

C. Physical plant

Size-appropriate facilities should be available in both postanesthesia and intensive care units; postoperative, dedicated in-patient floors with specially trained personnel should be available. Patient rooms and elevators must have sufficiently wide entrances. Floor-mounted commodes are recommended, but support systems can be used as an alternative. Design of new facilities that will accommodate the WLS patient must comply with the American Institute of Architects Planning and Design Guidelines for Bariatric Healthcare Facilities (73).

D. Extent of facility changes

WLS patients travel throughout hospitals for tests and procedures; there should be size-appropriate accommodations in all in-patient and outpatient points of service. These should include chairs and bathroom facilities, transferring equipment (stretchers and wheelchairs), and monitoring devices.

E. Investment

Specialized resources for WLS patients require a significant investment, the size of which depends on everything from

geography to patient population. Capital investments are preferred for renovations to existing facilities, and strongly recommended for new construction. WLS centers with lower volume or storage space problems should consider renting equipment.

F. Staff injury reduction

Health care consistently ranks among the top fields for back injuries. Well-established, agreed-upon, and well-known plans for transferring severely obese patients at all points of care can help reduce injuries. We also recommend that proper equipment, as well as training on how to use it, should be immediately available for the transfer of WLS patients. Staff should be well-educated in the use, location, and operation of available lift equipment. Portable equipment is more useful than ceiling lifts, but requires more room clearance. Trained and available on call "lift team" alternatives to equipment (as appropriate) should be considered.

G. Medical error reduction

We recommend dedicated facilities and staff to reduce risk of medical errors, including a dedicated hospital administrator to provide consistent support and oversight. All medical staff should be adequately trained and credentialed in best practice care of WLS patients (53,57,58). A team of designated medical subspecialists, fully aware of the problems and sensitivities of extremely obese patients, should be readily available, and all personnel who interact with WLS patients should attend obesity-specific education programs focused on sensitivity training.

H. Medication error reduction

Medication guidelines released by the Joint Commission Accreditation of Healthcare Organizations in 2004 (ref. 74) emphasize safety. We recommend that facilities follow these recommendations, as well as those specified in our prior report (61). We also recommend an Institutional Pharmacy and Therapeutics Committee to oversee WLS medical dosing regimens, and further research on medication use in the WLS patient.

I. Systems improvements

Clinical pathways are required by WLS accreditation programs, such as the American College of Surgeons Bariatric Surgery Center Network Accreditation Program (75). Clinical pathways specific to WLS patients should be established. These should be procedure-specific, updated frequently, and consistent with order sets. Regular meetings by the WLS team to review patient outcomes and address possible systems changes are essential, as is investment in a WLS database. The database should track patient outcomes and be compatible with the needs of the credentialing body that certifies the center. We recommend risk-adjusted outcomes to adequately evaluate performance.

X. Data Collection (Registries)/Future Considerations

This Task Group identified 212 papers and reviewed the 63 most relevant in detail. Recommendations are based on available evidence as well as consensus of opinions from Task Group and Expert Panel members (62,76).

A. Administrative and nonadministrative databases

Administrative databases have inherent problems, including unreliable coding and lack of WLS-specific data points. Clinical databases that are not WLS-specific have other shortcomings (e.g., short-term follow-up, sampling of WLS procedures), and single-institution, WLS-specific databases lack standardized definitions and appropriate quality benchmarks. Rather, we recommend collection of WLS-specific data (categories B and D) on 100% of weight loss surgeries performed (category D).

B. New developments

Longitudinal assessment of bariatric surgery. The NIH-funded Longitudinal Assessment of Bariatric Surgery consortium has developed a database of standardized information on WLS patients at six clinical centers. Data are being collected on patient characteristics, surgical procedures, medical and psychosocial outcomes, and economic factors.

Accreditation programs. The Centers for Medicare and Medicaid Services made a national decision to cover WLS, but only if performed by institutions and surgeons that are accredited by either the American College of Surgeons Bariatric Surgery Center Network or the American Society for Metabolic and Bariatric Surgery/Surgical Review Corporation Centers of Excellence program. WLS-specific, longitudinal data collection systems are a major part of each of these accreditation programs. The optimal data collection system should gather information on all WLS procedures using a longitudinal, universal database system. It should be prospective, risk adjusted, and benchmarked, with WLS-specific data points that track clinical effectiveness and complications following WLS (categories B and D).

The American College of Surgeons Bariatric Surgery Network Data Collection System, the Society of American Gastrointestinal Endoscopic Surgeons Bariatric Data Collection System, and the American Society for Metabolic and Bariatric Surgery/Surgical Review Corporation system should meet these criteria. If these systems are not compatible (i.e., cannot agree on the same definitions), an interface should be developed that makes them so (category D).

C. Areas that need more data

Risk adjustment. Risk adjustment helps control for differences in patient risk factors and case mix. Appropriate risk adjustment models should be developed and refined over time to account for these variables (categories C and D).

Determining the best data collector. Data entered into the system must be of the highest quality to ensure accurate analyses on quality of care. To avoid bias, data should be collected by audited, trained data collectors not directly involved in patient care (categories B and C). That data, in turn, should be analyzed

to see whether information collected by audited, trained non-nurse reviewers is as valid as that collected by nurse reviewers (category D).

Defining data points. High inter-rater reliability requires data points that are clinically relevant, objective, and easy to identify. Data points, definitions, and systems training programs should be developed that optimize clinical relevance and minimize subjectivity, and in so doing, maximize inter-rater reliability (categories C and D).

Quality indicators and benchmarking capabilities. Definitions of quality and benchmark indicators of progress can be difficult to develop. To advance patient safety, quality indicators and metrics should be appropriate and actionable (category D).

Outliers. Accurate determination of what constitutes an outlier, or bad performer, can have a direct effect on patient safety and access to WLS. Responsible analysis of data and careful definition of outliers is essential to improve quality of care. The means to regularly report that data to stakeholders should be determined (categories C and D). Poor performers, or high outliers, should be identified, and a mechanism for corrective action developed (category D).

Novel therapies. Safe introduction of novel technologies and assessment of the appropriateness of those procedures in new patient populations are critical for patient safety. Novel and experimental therapies, new patient populations, and expanded indications for WLS should be carefully studied through comprehensive data collection and analysis (category D). Experimental therapies should be performed with IRB approval, and data collected and audited by a data monitoring board to assess clinical effectiveness and patient safety (category D).

Cost-effectiveness and utility analyses. There is a critical need for well-designed prospective studies that evaluate the cost-effectiveness, cost utility, return on investment, and economic impact of WLS. Cost utility studies should be carried out to guide decision-making on the appropriate allocation of resources (category D).

State coalition. We propose the development of a statewide coalition to collectively gather and share data, and determine quality indicators and processes of care that could lead to best practices in WLS (categories C and D).

XI. Endoscopic Interventions

This Task Group's literature search identified 18 related articles, all of which were reviewed in detail. All of our recommendations are based on expert opinion (63).

A. Overview

Endoscopic interventions may provide valuable approaches to the management of WLS complications, and should be a high priority for development and investigation. Similarly, endoscopic interventions, endoscopically placed devices, and other minimally invasive, image-guided techniques may also provide valuable approaches to the primary management of obesity; they too should be a high priority for development and investigation (category D).

B. Experimental status

Until formally approved by appropriate regulatory bodies, novel endoscopic interventions and endoscopically placed devices should only be used in the setting of IRB-approved clinical trials (category D).

C. Credentials

Treatment with endoscopic and other image-guided interventions should be performed only by clinicians with specialized training and expertise in their effective and appropriate use (category D).

D. Clinical application

As is the standard for other medical and surgical therapies for obesity, endoscopic interventions should be studied and used only in the context of comprehensive patient evaluation and treatment that reflects the complex medical, nutritional, and behavioral contributors to obesity.

E. Risks and benefits

As new technologies become available, choice among therapeutic options for obesity should be determined by the comparative risk–benefit profiles of each modality. These considerations should be matched to the specific clinical characteristics, needs, and treatment goals of each patient (category D).

F. Data collection

To facilitate tracking of utilization, adverse events, and comparative outcomes, all patients who undergo endoscopic and other minimally invasive interventions for obesity and its complications should be entered into a standard registry. Methods of tracking should be compatible with those used for patients undergoing WLS (category D).

G. Coding and reimbursement

As new devices and minimally invasive surgical therapies for obesity and its complications are approved for clinical use, a new category of provisional billing codes should be established for these interventions. Reimbursement for novel therapies for obesity should be determined on the basis of scientific evidence of their safety and efficacy (category D).

H. Future research

Randomized, blinded, sham-controlled clinical trials should be the standard for investigation of the safety and efficacy of endoscopic interventions for the treatment of obesity and its complications (category D).

ACKNOWLEDGMENTS
This Expert Panel on Weight Loss Surgery report was prepared for the Betsy Lehman Center for Patient Safety and Medical Error reduction. The report was commissioned by N.R., Director, Betsy Lehman Center for Patient Safety and Medical Error Reduction. We thank all Expert Panel and Task Group members for their important contributions to the report. We also acknowledge Eileen McHale at the Betsy Lehman Center and Jane Guilfoyle at the Department of Public Health.

DISCLOSURE
The authors declared no conflict of interest.

APPENDIX I
To view Task Group Appendices, go to http://www.mass.gov.dph and search "weight loss surgery."

Framework and methodology for evidence-based systematic reviews of literature on weight loss surgery

The Expert Panel was charged with reviewing WLS operations, identifying potential safety issues, and recommending specific actions to reduce safety risks and improve patient outcomes. It used the methodology of evidence-based medicine to systematically search available literature on the subject, and developed a classification system from established models to grade the quality of evidence.

The systematic review involved a MEDLINE search of studies published from April 2004 to May 2007. These included prior systematic reviews on the subject, randomized controlled trials, prospective cohort studies, cross-sectional surveys, case reports, and existing guidelines on WLS procedures from national organizations. The panel based its grading classification system on those used by the US Preventive Services Task Force, the American Diabetes Association, and the National Heart, Lung, and Blood Institute (NHLBI) Obesity Education Initiative Expert Panel on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults.

Randomized controlled trials (RCTs) are considered the highest-level evidence of clinical efficacy and safety, but there are few such studies on WLS operations. The Expert Panel's recommendations are based on the best available evidence. The sections below detail the procedures and methodology used to develop recommendations.

1. Panel selection

At the request of Massachusetts Public Health Commissioner, Christine Ferguson, the Betsy Lehman Center for Patient Safety and Medical Error Reduction (Lehman Center) convened an Expert Panel to study patient-related safety issues in the state's WLS programs and procedures.

The 35-member panel included experienced weight loss surgeons, nurses, psychologists, and a nutritionist who counsels patients before and after the procedures; other physicians who care for patients with obesity (an anesthesiologist, internist, and pediatrician); a hospital patient safety officer; a health plan medical director; an ethicist; and a consumer. The panel delivered a report on its progress to the Lehman Center and the Department of Public Health in mid-July 2007.

2. Task groups

We divided the panel into 11 task groups:

- Surgical Care (53).
- Multidisciplinary Evaluation and Treatment (54).
- Behavioral and Psychological Care (55).
- Pediatric/Adolescent (56).
- Anesthetic Perioperative Care and Pain Management (57).
- Nursing Perioperative Care (58).
- Informed Consent and Patient Education (59).
- Policy and Access (Coding and Reimbursement) (60).
- Specialized Facilities and Resources (61).
- Data Collection (Registries)/Future Considerations (62).
- Endoscopic Interventions (63).

Panel members joined one or two task groups, each with an assigned coordinator. They were asked to update reports from the prior Lehman Center supplement (22).

3. Literature search

A medical librarian, aided by a clinical epidemiologist with experience in systematic reviews, carried out literature searches for each task group. Studies were included or excluded based on *a priori* criteria, i.e., written protocols that defined research questions and search parameters, including patient characteristics, study designs, surgical interventions, and outcomes.

MEDLINE searches were limited to English-language studies published from April 2004 to May 2007. (Some groups searched other databases or focused on more recent literature.) References in retrieved articles, guidelines from national organizations, and systematic reviews from the Cochrane Library were also examined. Task group coordinators, with input from the clinical epidemiologist, screened all titles and abstracts; they selected only those most relevant to the review questions.

4. Data extraction and tabulation

The panel developed a data extraction sheet and used it to cull detailed information from selected full articles after review. Key data included study design; size; patient demographics; follow-up time; dropout rate; description of the intervention; outcome measures, including adverse effects; and main conclusions. Information was tabulated in a format suitable for publication.

5. Synthesis of evidence

We primarily used narrative (or qualitative) summaries for the literature review because study designs and outcomes were too dissimilar to combine results in a formal meta-analysis. All selected studies were critically assessed for internal validity or methodological rigor. They were ranked according to levels of evidence based on study design

(Table 1). For example, well-conducted RCTs (category A) provide the strongest evidence on the effectiveness of a surgical weight loss procedure. We used expert opinion (category D) (including clinical experience, the opinions of respected authorities, reports from expert committees, and consensus of the Expert Panel) in conjunction with evidence from RCTs or observational studies to develop recommendations.

6. Developing evidence-based recommendations

Each task group prepared a critical summary of the literature (Table 2) and developed evidence-based recommendations on its assigned topic; these were presented to the full group for comments. This Executive Report of key recommendations from all groups was approved by the Expert Panel at its last meeting on 19 July 2007.

Table 1 Grading system for evidence-based recommendations

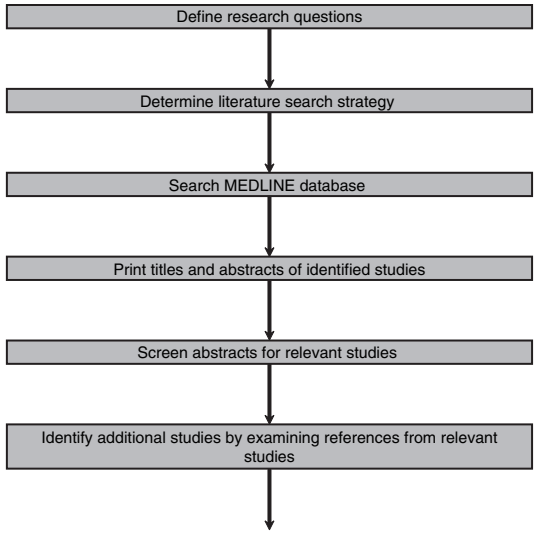
Category A	Evidence obtained from at least one well-conducted randomized clinical trial or a systematic review of all relevant RCTs
Category B	Evidence from well-conducted prospective cohort studies, registry or meta-analysis of cohort studies, or population-based case–control studies
Category C	Evidence obtained from uncontrolled or poorly controlled clinical trials, or retrospective case–control analyses, cross-sectional studies, case series, or case reports
Category D	Evidence consisting of opinion from expert panels or the clinical experience of acknowledged authorities

Adapted from the criteria used by the US Preventive Services Task Force (USPSTF) and the American Diabetes Association.

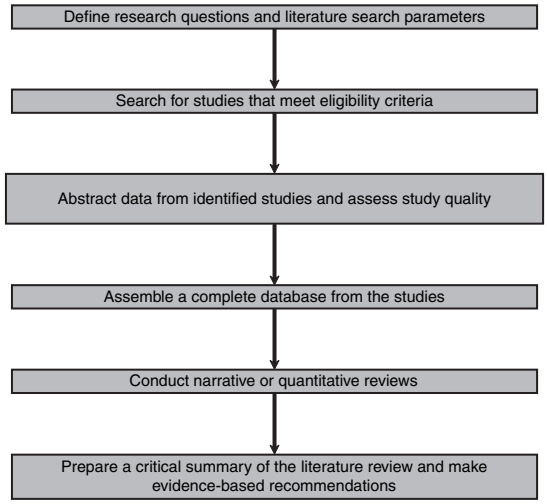
Table 2 Inclusion/exclusion criteria—example used in literature search, laparoscopic vs. open gastric bypass surgery

Inclusion criteria
English language
Published between April 2004 and May 2007
RCTs or controlled trials without randomization, cohort studies
Surgical procedures: gastric bypass, Roux-en-Y gastric bypass, open vs. laparoscopic
Minimum follow-up: 6 months
Outcomes: change in body weight, excess weight, and BMI; mortality and major morbidity
Exclusion criteria
Selection criteria not indicated
Small sample size ($n < 10$ for each intervention)
Dropout rate >50%

Literature search process



Literature review process



REFERENCES FOR THE FRAMEWORK

1. Oxford Centre for Evidence-based Medicine Levels. <http://www.musckids.com/~annibald/ebm/oxford_levels_of_evidence.pdf> (2001). Accessed 23 August 2007.
2. Clinical Guidelines on the Identification, Evaluation, and Treatment, of Overweight and Obesity in Adults. The Evidence Report: National Institutes of Health. National Heart, Lung, and Blood Institute; 1998. No. 98-4083.
3. Introduction. *Diabetes Care* 2004;27:S1–S2.
4. Agency for Healthcare Research and Quality. Current Methods of the US Preventive Services Task Force: a Review of the Process. <http://www.ahrq.gov/clinic/ajpmsuppl/harris1.htm>. Accessed 24 August 2007.
5. 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(3 Suppl):21–35.
6. Naylor CD, Guyatt GH. Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services. The Evidence-Based Medicine Working Group. *JAMA* 1996;275:554–558.
7. Barton MB, Miller T, Wolff T *et al*. How to read the new recommendations statement: methods update for the U.S. Preventive Services Task Force. *Ann Intern Med* 2007;147:123–127.
8. Guirguis-Blake J, Calonge N, Miller T *et al*. Current processes of the U.S. Preventive Services Task Force: refining evidence-based recommendation development. *Ann Intern Med* 2007;147:117–122.

APPENDIX II

Task Groups for Lehman Center Report on Weight Loss Surgery

Surgical Care

Coordinator

John Kelly, M.D., University of Massachusetts Memorial Medical Center

Co-chair

Scott Shikora, M.D., Tufts-New England Medical Center
George L. Blackburn, M.D., Ph.D., Beth Israel Deaconess Medical Center
Frederick Buckley, M.D., North Shore Medical Center
Matthew Hutter, M.D., Massachusetts General Hospital
Daniel B. Jones, M.D., M.S., Beth Israel Deaconess Medical Center
David Lautz, M.D., Brigham and Women's Hospital
Andrew B. Lederman, M.D., Berkshire Medical Center
Malcolm K. Robinson, M.D., Brigham and Women's Hospital
John Romanelli, M.D., F.A.C.S., Baystate Medical Center

Multidisciplinary Evaluation and Treatment

Coordinator

Caroline Apovian, M.D., Boston Medical Center

Co-chair

Susan Cummings, M.S., R.D., L.D.N., Massachusetts General Hospital
Wendy Anderson, R.D., L.D.N., Boston Medical Center
Loren J. Borud, M.D., Beth Israel Deaconess Medical Center
Kelly Moore, R.D., L.D.N., Beth Israel Deaconess Medical Center
Kristina Day, R.D., L.D.N., Beth Israel Deaconess Medical Center
Edward Hatchigian, M.D., Beth Israel Deaconess Medical Center
Barbara Hodges. R.D., M.P.H., L.D.N., Brigham and Women's Hospital
Mary Elizabeth Patti, M.D., Joslin Diabetes Center
Frank Perna, Ph.D., National Cancer Institute
Mark Pettus, M.D., Berkshire Medical Center
Daniel Rooks, Ph.D., Beth Israel Deaconess Medical Center
Edward Saltzman, M.D., Tufts-New England Medical Center
June Skoropowski, R.D., L.D.N., Beth Israel Deaconess Medical Center
Michael B. Tantillo, M.D., Private Practice, Brookline, MA
Phyllis Thomason, M.S., R.D., L.D.N., Faulkner Hospital

Behavioral and Psychological Care

Coordinator

Isaac Greenberg, Ph.D., Tufts-New England Medical Center

Co-chair

Stephanie Sogg, Ph.D., Massachusetts General Hospital
Frank Perna, Ph.D., National Cancer Institute

Pediatric/Adolescent

Coordinator

Janey S.A. Pratt, M.D., Massachusetts General Hospital

Co-chair

Carine Lenders, M.D., M.S., Boston Medical Center
Emily Dionne, Massachusetts General Hospital
Alison G. Hoppin, M.D., Massachusetts General Hospital
George Hsu, M.D., Tufts-New England Medical Center
Thomas Inge, M.D., Ph.D., Cincinnati Children's Hospital Medical Center
David Lawlor, M.D., Massachusetts General Hospital
Margaret Marino, Ph.D., Boston Medical Center
Alan Meyers, M.D., Boston Medical Center
Jennifer Rosenblum, M.D., Massachusetts General Hospital
Vivian Sanchez, M.D., Beth Israel Deaconess Medical Center

Anesthetic Perioperative Care and Pain Management

Coordinator

Roman Schumann, M.D., Tufts-New England Medical Center

Co-chair

Stephanie Jones, M.D., Beth Israel Deaconess Medical Center
Daniel B. Carr, M.D. (Advisor), Tufts-New England Medical Center
Kathy Connor, M.D., Newton-Wellesley Hospital
Bronwyn Cooper, M.D., University of Massachusetts Memorial Medical Center
Alan M. Harvey, M.D., M.B.A., Brigham and Women's Hospital
Michael Kaufman, M.D., Lahey Clinic
Scott Kelley, M.D., Brigham and Women's Hospital
Vilma E. Ortiz, M.D., Massachusetts General Hospital
Mark Vanden Bosch, M.D., Berkshire Medical Center

Nursing Perioperative Care

Coordinator

Ann Mulligan, R.N., Newton-Wellesley Hospital

Co-chair

Anne McNamara, R.N., Beth Israel Deaconess Medical Center
Hannah Boulton, R.N., M.S.N., South Shore Hospital
Ann Mullen, R.N., B.S.N., Newton-Wellesley Hospital
Carol Raiano, R.N., C.C.R.N., Newton-Wellesley Hospital
Linda Trainor, R.N., B.S.N., Beth Israel Deaconess Medical Center

Informed Consent and Patient Education

Coordinator

Christina C. Wee, M.D., M.P.H. Beth Israel Deaconess Medical Center

Co-chair

Michael Paasche-Orlow, M.D., M.P.H. Boston University School of Medicine
Robert Fanelli, M.D., Berkshire Medical Center
Janey Pratt, M.D., Massachusetts General Hospital
Patricia Samour, M.M.Sc., R.D., L.D.N., Beth Israel Deaconess Medical Center
Linda Trainor, R.N., B.S.N., Beth Israel Deaconess Medical Center

Policy and Access (Coding and Reimbursement)
Coordinator
Scott Shikora, M.D., Tufts-New England Medical Center
Co-chair
Rayford Kruger M.D., F.A.C.S., Tobey Hospital
George L. Blackburn, M.D., Ph.D., Beth Israel Deaconess Medical Center
John A. Fallon, M.D., M.B.A., F.A.C.P., Blue Cross Blue Shield of Massachusetts
Alan M. Harvey, M.D., M.B.A., Mercy Medical Center/ Catholic Health East
Elvira Johnson, M.S., R.D., C.D.E., L.D.N., Massachusetts Dietetics Association
Lee Kaplan, M.D., Ph.D., Massachusetts General Hospital
David Lautz, M.D., Brigham and Women's Hospital
Robert LoNigro, M.D., M.S., Tufts Health Plan
Edward C. Mun, M.D., Brigham and Women's Hospital
Malcolm K. Robinson, M.D., Brigham and Women's Hospital
Roger L. Snow, M.D., M.P.H., University of Massachusetts Medical School and MassHealth
Lee Steingisser, M.D., Blue Cross Blue Shield of Massachusetts
James Sabin, M.D., Harvard Pilgrim Health Care
Stancel M. Riley Jr., M.D., Massachusetts Board of Registration in Medicine

Specialized Facilities and Resources
Coordinator
David Lautz, M.D., Brigham and Women's Hospital
Co-chair
Michael E. Jiser, M.D., Saints Memorial Medical Center
Robert J. Cella, M.D., Berkshire Medical Center
John Kelly, M.D., University of Massachusetts Medical Center
Sheila K. Partridge, M.D., Newton-Wellesley Hospital
John Romanelli, M.D., F.A.C.S., Baystate Medical Center
John P. Ryan, R.N., Beth Israel Deaconess Medical Center
Scott Shikora, M.D., Tufts-New England Medical Center

Data Collection (Registries)/Future Considerations
Coordinator
Matthew M. Hutter, M.D., Massachusetts General Hospital
Co-Chair
Daniel B. Jones, M.D., M.S., Beth Israel Deaconess Medical Center
Robert J. Cella, M.D., Berkshire Medical Center
Stancel M. Riley Jr., M.D., Massachusetts Board of Registration in Medicine
Benjamin Schneider, M.D., Beth Israel Deaconess Medical Center
Roger L. Snow, M.D., M.P.H., University of Massachusetts Medical School and MassHealth
Kerri Clancy, R.N., Brigham and Women's Hospital

Endoscopic Interventions
Coordinator
Lee Kaplan, M.D., Ph.D., Massachusetts General Hospital
Co-chair
Christopher C. Thompson, M.D., M.H.E.S., Brigham and Women's Hospital
William R. Brugge, M.D., Massachusetts General Hospital
Ram Chuttani, M.D., Beth Israel Deaconess Medical Center
David Desilets, M.D., Baystate Medical Center
James C. Ellsmere, M.D., Beth Israel Deaconess Medical Center
David W. Rattner, M.D., Massachusetts General Hospital
Michael Tarnoff, M.D., Tufts-New England Medical Center

SUPPLEMENTARY MATERIAL

To review task group appendices, go to www.mass.gov/dph and search "Weight Loss Surgery."

DISCLOSURE

The authors declared no conflict of interest.

REFERENCES

1. Bray GA. The missing link—lose weight, live longer. *N Engl J Med* 2007;357:818–820.
2. Dixon J. Survival advantage with bariatric surgery: report from the 10th International Congress on Obesity. *Surg Obes Relat Dis* 2006;2:585–586.
3. Maggard MA, McGory ML, Shekelle PG, Ko CY. Quality indicators in bariatric surgery: improving quality of care. *Surg Obes Relat Dis* 2006;2:423–429.
4. Sjöström L, Narbro K, Sjöström CD *et al*. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med* 2007;357:741–752.
5. Adams TD, Gress RE, Smith SC *et al*. Long-term mortality after gastric bypass surgery. *N Engl J Med* 2007;357:753–761.
6. Favretti F, Segato G, Ashton D *et al*. Laparoscopic adjustable gastric banding in 1,791 consecutive obese patients: 12-year results. *Obes Surg* 2007;17:168–175.
7. Perry CD, Hutter MM, Smith DB, Newhouse JP, McNeil BJ. Survival and changes in comorbidities after bariatric surgery. *Ann Surg* 2008;247:21–27.
8. Sowemimo OA, Yood SM, Courtney J *et al*. Natural history of morbid obesity without surgical intervention. *Surg Obes Relat Dis* 2007;3:73–77.
9. Christou NV, Sampalis JS, Liberman M *et al*. Surgery decreases long-term mortality, morbidity, and health care use in morbidly obese patients. *Ann Surg* 2004;240:416–423.
10. Flum DR, Dellinger EP. Impact of gastric bypass operation on survival: a population-based analysis. *J Am Coll Surg* 2004;199:543–551.
11. Kral JG. Studies show reduction in mortality after bariatric surgery. *Surg Obes Relat Dis* 2006;2:564.
12. McMahon MM, Sarr MG, Clark MM *et al*. Clinical management after bariatric surgery: value of a multidisciplinary approach. *Mayo Clin Proc* 2006;81(10 Suppl):S34–S45.
13. Clark MM, Balsiger BM, Sletten CD *et al*. Psychosocial factors and 2-year outcome following bariatric surgery for weight loss. *Obes Surg* 2003;13:739–745.
14. Sarwer DB, Wadden TA, Fabricatore AN. Psychosocial and behavioral aspects of bariatric surgery. *Obes Res* 2005;13:639–648.
15. Herpertz S, Kielmann R, Wolf AM *et al*. Does obesity surgery improve psychosocial functioning? A systematic review. *Int J Obes Relat Metab Disord* 2003;27:1300–1314.
16. Arcila D, Velázquez D, Gamino R *et al*. Quality of life in bariatric surgery. *Obes Surg* 2002;12:661–665.
17. Kolotkin RL, Crosby RD, Pendleton R *et al*. Health-related quality of life in patients seeking gastric bypass surgery vs non-treatment-seeking controls. *Obes Surg* 2003;13:371–377.
18. Elkins G, Whitfield P, Marcus J *et al*. Noncompliance with behavioral recommendations following bariatric surgery. *Obes Surg* 2005;15:546–551.
19. Collazo-Clavell ML, Clark MM, McAlpine DE, Jensen MD. Assessment and preparation of patients for bariatric surgery. *Mayo Clin Proc* 2006 81(10 Suppl):S11–S17.
20. Belle SH, Berk PD, Courcoulas AP *et al*. Safety and efficacy of bariatric surgery: Longitudinal Assessment of Bariatric Surgery. *Surg Obes Relat Dis* 2007;3:116–126.
21. Kendrick ML, Dakin GF. Surgical approaches to obesity. *Mayo Clin Proc* 2006;81(10 Suppl):S18–S24.

22. Lehman Center Weight Loss Surgery Expert Panel.Commonwealth of Massachusetts. Betsy Lehman Center for Patient Safety and Medical Error Reduction. Expert Panel on Weight Loss Surgery: Executive Report. *Obes Res* 2005;13:206–226.
23. Rand Corporation. Research Briefs.Obesity and Disability: The Shape of Things to Come. <http://www.rand.org/pubs/research_briefs/RB9043-1/> (2007). Accessed 9 June 2007.
24. Desjardins E, Schwartz AL. Collaborating to combat childhood obesity. *Health Aff (Millwood)* 2007;26:567–571.
25. Sturm R. Increases in morbid obesity in the USA: 2000–2005. *Public Health* 2007;121:492–496.
26. Haslam DW, James WP. Obesity. *Lancet* 2005;366:1197–1209.
27. Li Z, Bowerman S, Heber D. Health ramifications of the obesity epidemic. *Surg Clin North Am* 2005;85:681–701.
28. Finkelstein EA, Fiebelkorn IC, Wang G. National medical spending attributable to overweight and obesity: how much, and who's paying? *Health Aff (Millwood)* 2003;W3:219–226.
29. Economic Consequences. Centers for Disease Control and Prevention. <http://www.cdc.gov/nccdphp/dnpa/obesity/economic_consequences.htm> (2007). Accessed 9 June 2007.
30. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA* 2004;29:1238–1245. [Erratum in: *JAMA* 2005;293:293–294. *JAMA* 2005;293:298].
31. Olshansky SJ, Passaro DJ, Hershow RC *et al*. A potential decline in life expectancy in the United States in the 21st century. *N Engl J Med* 2005;352:1138–1145.
32. de Fatima, Haueisen Sander, Diniz M, de Azeredo, Passos VM, Diniz MT. Gut-brain communication: how does it stand after bariatric surgery? *Curr Opin Clin Nutr Metab Care* 2006;9:629–636.
33. Samuel I, Mason EE, Renquist KE *et al*. Bariatric surgery trends: an 18-year report from the International Bariatric Surgery Registry. *Am J Surg* 2006;192:657–662.
34. Elder KA, Wolfe BM. Bariatric surgery: a review of procedures and outcomes. *Gastroenterology* 2007;32:2253–2271.
35. DeMaria EJ. Bariatric surgery for morbid obesity. *N Engl J Med* 2007;356:2176–2183.
36. Zhao Y, Encinosa W. Bariatric Surgery Utilization and Outcomes in 1998 and 2004. Statistical Brief #23. Agency for Healthcare Research and Quality: Rockville, MD, 2007.
37. Kim TH, Daud A, Ude AO *et al*. Early U.S. outcomes of laparoscopic gastric bypass versus laparoscopic adjustable silicone gastric banding for morbid obesity. *Surg Endosc* 2006;20:202–209.
38. Smoot TM, Xu P, Hilsenrath P, Kuppersmith NC, Singh KP. Gastric bypass surgery in the United States, 1998–2002. *Am J Public Health* 2006;96:1187–1189.
39. Wolfe BM, Morton JM. Weighing in on bariatric surgery: procedure use, readmission rates, and mortality. *JAMA* 2005;294:1960–1963.
40. le Roux CW, Aylwin SJ, Batterham RL *et al*. Gut hormone profiles following bariatric surgery favor an anorectic state, facilitate weight loss, and improve metabolic parameters. *Ann Surg* 2006;243:108–114.
41. Peluso L, Vanek VW. Efficacy of gastric bypass in the treatment of obesity-related comorbidities. *Nutr Clin Pract* 2007;22:22–28.
42. Parikh MS, Laker S, Weiner M, Hajiseyedjavadi O, Ren CJ. Objective comparison of complications resulting from laparoscopic bariatric procedures. *J Am Coll Surg* 2006;202:252–261.
43. Chapman AE, Kiroff G, Game P *et al*. Laparoscopic adjustable gastric banding in the treatment of obesity: a systematic literature review. *Surgery* 2004;135:326–351.
44. Olbers T, Fagevik-Olsen M, Maleckas A, Lonroth H. Randomized clinical trial of laparoscopic Roux-en-Y gastric bypass versus laparoscopic vertical banded gastroplasty for obesity. *Br J Surg* 2005;92:557–562.
45. Scopinaro N, Gianetta E, Civalleri D, Bonalumi U, Bachi V. Bilio-pancreatic bypass for obesity: II. Initial experience in man. *Br J Surg* 1979;66:618–620.
46. Kotidis EV, Koliakos G, Papavramidis TS, Papavramidis ST. The effect of biliopancreatic diversion with pylorus-preserving sleeve gastrectomy and duodenal switch on fasting serum ghrelin, leptin and adiponectin levels: is there a hormonal contribution to the weight-reducing effect of this procedure? *Obes Surg* 2006;16:554–559.
47. Cottam D, Qureshi FG, Mattar SG *et al*. Laparoscopic sleeve gastrectomy as an initial weight-loss procedure for high-risk patients with morbid obesity. *Surg Endosc* 2006;20:859–863.
48. Roa PE, Kaidar-Person O, Pinto D *et al*. Laparoscopic sleeve gastrectomy as treatment for morbid obesity: technique and short-term outcome. *Obes Surg* 2006;16:1323–1326.
49. Silecchia G, Boru C, Pecchia A *et al*. Effectiveness of laparoscopic sleeve gastrectomy (first stage of biliopancreatic diversion with duodenal switch) on co-morbidities in super-obese high-risk patients. *Obes Surg* 2006;16:1138–1144.
50. Mognol P, Chosidow D, Marmuse JP. Laparoscopic sleeve gastrectomy (LSG): review of a new bariatric procedure and initial results. *Surg Technol Int* 2006;15:47–52.
51. Melissas J, Koukouraki S, Askoxylakis J *et al*. Sleeve gastrectomy: a restrictive procedure? *Obes Surg* 2007;17:57–62.
52. Langer FB, Reza Hoda MA, Bohdjalian A *et al*. Sleeve gastrectomy and gastric banding: effects on plasma ghrelin levels. *Obes Surg* 2005;15:1024–1029.
53. Kelly J, Shikora S, Jones DB *et al*. Best practice updates for surgical care in weight loss surgery. *Obes*, in press.
54. Apovian CM, Cummings S, Anderson W *et al*. Best practice updates for multidisciplinary care in weight loss surgery. *Obes*, in press.
55. Greenberg I, Sogg S, Perna F. Behavioral and psychological care in weight loss surgery—best practice update. *Obes*, in press.
56. Pratt J, Lenders CM, Dionne E *et al*. Best practice updates for pediatric/adolescent weight loss surgery. *Obes*, in press.
57. Schumann R, Jones SB, Cooper B *et al*. Update on best practice recommendations for anesthetic preioperative care and pain management in weight loss surgery, 2004–2007. *Obes*, in press.
58. Mulligan A, McNamara A, Boulton H *et al*. Best practice updates for nursing care in weight loss surgery. *Obes*, in press.
59. Wee CC, Pratt J, Fanelli R *et al*. Best practice updates for informed consent and patient education in weight loss surgery. *Obes*, in press.
60. Shikora SA, Kuger RS Jr, Blackburn GL *et al*. Best practices in policy and access for weight loss surgery. *Obes*, in press.
61. Lautz DB, Jiser ME, Kelly J *et al*. An update on best practice guidelines for specialized facilities and resources necessary for bariatric surgical programs. *Obes*, in press.
62. Hutter MM, Jones DB, Riley S *et al*. Best practice updates for weight loss surgery data collection. *Obes*, in press.
63. Ellsmere JC, Thompson CC, Brugge WR *et al*. Best practices in endoscopic intervention for weight loss surgery. *Obes*, in press.
64. Trus TL, Pope GD, Finlayson SR. National trends in utilization and outcomes of bariatric surgery. *Surg Endosc* 2005;19:616–620.
65. Galvani C, Gorodner M, Moser F *et al*. Laparoscopic adjustable gastric band versus laparoscopic Roux-en-Y gastric bypass: ends justify the means? *Surg Endosc* 2006;20:934–941.
66. DeMaria EJ, Jamal MK. Laparoscopic adjustable gastric banding: evolving clinical experience. *Surg Clin North Am* 2005;85:773–787.
67. Suter M, Giusti V, Worreth M, Heraief E, Calmes JM. Laparoscopic gastric banding: a prospective, randomized study comparing the Lapband and the SAGB: early results. *Ann Surg* 2005;241:55–62.
68. Suter M, Calmes JM, Paroz A, Giusti V. A 10-year experience with laparoscopic gastric banding for morbid obesity: high long-term complication and failure rates. *Obes Surg* 2006;16:829–835.
69. Kelly J, Tarnoff M, Shikora S *et al*. Best practice recommendations for surgical care in weight loss surgery. *Obes Res* 2005;13:227–233.
70. Schirmer BD, Schauer PR, Flum DR, Ellsmere J, Jones DB. Bariatric surgery training: getting your ticket punched. *J Gastrointest Surg* 2007;11:807–812.
71. Lautz DB, Jackson TD, Clancy KA *et al*. Bariatric operations in Veterans Affairs and selected university medical centers: results of the patient safety in surgery study. *J Am Coll Surg* 2007;204:1261–1272.
72. American Society for Metabolic and Bariatric Surgery. <http://www.asbs.org/> (2007). Accessed 23 August 2007.
73. Andrade SD. Planning and Design Guidelines for Bariatric Healthcare Facilities.American Institute of Architects. AIA Academy Journal. <http://www.aia.org/journal_aah.cfm?pagename=aah_jrnl_20061018_award_winner&dspl=1&article=article> (2006). Accessed 6 August 2007.
74. Rich DS. New JCAHO medication standards for 2004. *Am J Health-Syst Pharm* 2004;61:1349–1358.
75. ACS Division of Research and Optimal Patient Care. American College of Surgeons Bariatric Surgery Center Network (BSCN) Accreditation Program Manual.American College of Surgeons. <http://facs.org/index.html>. Accessed 23 August 2007.
76. Commonwealth of Massachusetts. Betsy Lehman Center for Patient Safety and Medical Error Reduction. Expert Panel on Weight Loss Surgery: Executive Report update. *Obes*, in press.

Barriers and Facilitators to Routine HIV Testing in VA Primary Care

Barbara G. Bokhour, PhD[1,2], Jeffrey L. Solomon, PhD[1], Herschel Knapp, PhD[3], Steven M. Asch, MD[3], and Allen L. Gifford, MD[1,2]

[1]Center for Health Quality, Outcomes & Economic Research, Edith Nourse Rogers Veterans Affairs Medical Center, Bedford, MA, USA; [2]Boston University School of Public Health, Boston, MA, USA; [3]Greater Los Angeles Veterans Affairs Healthcare System, Los Angeles, CA, USA.

BACKGROUND: Approximately 21% of the 1.1 million HIV-infected persons in the United States are unaware of their HIV status. The Centers for Disease Control (CDC) recommend routine opt-out HIV testing for all patients aged 13–64. Yet little is known about patient and provider perspectives on routine HIV testing.

OBJECTIVE: We sought to understand patient and provider perspectives on the adoption of routine HIV testing within the US Department of Veterans Affairs.

DESIGN: We conducted four focus groups with patients and two focus groups with primary care providers to explore perceptions of, communication about, and barriers and facilitators to routine HIV testing in primary care.

PARTICIPANTS: Convenience sample of patients and primary care providers at two geographically diverse Veterans' Affairs Medical Centers.

APPROACH: We conducted grounded thematic analyses of transcribed audio-recordings of focus groups to identify major themes, identifying similarities and differences between patient and provider perspectives.

MAIN RESULTS: Patients and providers concurred that implementation of routine HIV testing, treating HIV like other chronic diseases, and removing requirements for written informed consent and pre-test counseling were of benefit to patients and to public health. Patients, however, wished to have HIV testing routinely offered by providers so that they could decide whether or not to be tested. Veterans also stated that routinizing testing would help destigmatize HIV. Six steps to communicating about routine testing ("the 6 R's") were identified.

CONCLUSIONS: Patients and providers appear ready for implementation of routine HIV testing. However, providers should use patient-centered communication strategies to ease patients' concerns about confidentiality and stigma associated with HIV disease.

Received December 24, 2008
Revised June 25, 2009
Accepted July 16, 2009
Published online August 19, 2009

KEY WORDS: HIV/AIDS; screening; communication; qualitative research.
J Gen Intern Med 24(10):1109–14
DOI: 10.1007/s11606-009-1078-6

INTRODUCTION

With the availability of effective medications, early diagnosis of HIV reduces mortality and is cost-effective[1–3]. Patients diagnosed and treated for HIV avoid hospitalizations and can reduce risky behaviors, further reducing HIV transmission[4]. Despite longstanding guidelines recommending HIV testing, as many as 21% of the 1.1 million HIV-infected persons in the US are unaware of their status[5]. Consequently, in 2006 the Centers for Disease Control and Prevention (CDC) released new recommendations that *all* patients between the ages of 13 and 64 be offered opt-out HIV testing regardless of risk status [6]; these recommendations were recently endorsed by the American College of Physicians[7]. The guidelines suggest that providers routinely order HIV testing unless the patient declines. These guidelines differ from previous HIV testing practices in most jurisdictions, including the Department of Veterans Affairs (VA), in that previously HIV testing was only for patients at behavioral risk for HIV, or if patients requested a test. In addition, requirements for specified pre-test counseling and written informed consent have been, and remain, the policy in many places.

Routine 'opt-out' testing, in which patients are informed by the provider that they will be tested for HIV and assent is inferred unless the patients decline, is likely to reduce barriers to HIV testing [8]. Yet adoption of routine testing in general, and of opt-out testing practices in particular, may be challenging [9–11]. We know little about how patients will respond to the routine offering of HIV testing or how providers view this extension in scope of their primary care responsibilities. The persistent stigma associated with HIV and patients' fear of the health and social consequences of the diagnosis have been identified as significant barriers to being tested[10,12]. Many primary care providers find discussing the sensitive topics of HIV risk behavior and testing difficult and awkward[13,14]. Therefore, routine discussions of HIV testing may be more complex than screening for other diseases.

We sought to identify, from patient and provider perspectives, facilitators and barriers to implementing routine HIV testing in the VA according to current CDC guidelines. The issue is of particular importance in the VA, where a change in policy will affect the care of over 5 million veterans. Veterans are subject to social and behavioral factors that put them at higher risk for HIV, including lower income, more substance

abuse and other co-morbidities[15]. Thus, HIV prevalence rates are higher than in the general population[16], and early HIV detection is of particular concern. Understanding the concerns and perspectives of both VA patients and providers is central to succeeding with implementation of routine testing in this large health-care system.

METHODS

We conducted a qualitative study using focus groups of primary care providers and patients at two VA Medical Centers, one in California, one in New England. We explored provider and patient attitudes towards routine HIV testing, perceptions about what effective communication about HIV testing should entail, and perceptions of barriers and facilitators to implementing routine HIV testing. Institutional review boards at both sites approved the study.

Participants

Patients were recruited using informational posters, post cards with a drop-box placed in primary care clinics and at patient orientation sessions. Recruitment materials stated we were conducting focus groups about talking with providers about sensitive issues. A research assistant contacted each interested patient and indicated that HIV would be discussed in the focus groups. Patients were eligible to participate if they responded negatively to the question, "Have you ever been told you have HIV disease?" This requirement ensured that participants had not been diagnosed with HIV disease, which might alter their perceptions of HIV testing. Patients were eligible regardless of previous HIV testing, to include patients with HIV testing experiences.

Providers were recruited at regularly scheduled staff meetings at each medical center by a project investigator. All primary care providers were eligible to participate, including physicians, nurse practitioners and physician assistants. Written informed consent was obtained from all participants.

Data Collection

Patient focus groups were conducted at the medical centers by a study investigator, lasted approximately 90 min and were audio-recorded. We used a semi-structured focus group guide, containing broad lead questions and focused follow-up questions, designed to elicit patient experiences with, and perspectives on, HIV testing. The guide was used flexibly to follow the flow of conversation among participants while addressing all topics. Patients also commented on draft CDC patient HIV testing educational materials.

Provider focus groups were conducted during regularly scheduled 45-min staff meetings. The semi-structured focus group guide elicited provider experiences with discussing HIV testing with patients, barriers to conducting HIV testing and routinizing HIV testing as recommended by the CDC. Providers also commented on draft patient and provider CDC educational materials.

Analysis

Audio-recordings were transcribed verbatim. Transcripts were analyzed qualitatively using procedures informed by grounded theory methodology[17,18], a systematic approach to deriving qualitative themes from textual data. Accordingly, we first conducted open coding in which an investigator identifies key concepts emerging from the language used by participants and assigns codes (descriptive phrases) to segments of text. NVIVO qualitative analysis software was used to facilitate data coding and sorting[19]. Coded text segments were reviewed by two investigators to condense broad codes into distinct themes. Themes that emerged in both patient and provider groups were examined for similarities and differences in perspectives in a process known as constant comparison analysis. Subsequently, prominent themes and quotes exemplifying each were presented to the research team and refined through discussion.

RESULTS

Of the 67 patients contacted, 10 refused participation, 3 were ineligible, and 26 were unable to participate, predominantly due to scheduling problems. The 28 patient participants in four focus groups—2 at each site—were all men aged 35–88, were predominantly low income and had mixed levels of education (Table 1). Of the 47 providers contacted, 20 providers consented to participate; however, 7 dropped out because of scheduling difficulties. Thirteen providers participated in two provider focus groups, one at each site, and two additional providers who were unable to attend a focus group participated in individual interviews. Seven of the providers were women, eight were physicians, six were nurse practitioners and one a registered nurse.

In the focus groups, patients and providers discussed key issues that help or hinder conversations about, and achievement of, HIV testing in primary care encounters. Themes centered around perspectives on HIV testing becoming routine, what information should be communicated to patients and how, and procedural issues that hinder HIV testing. We present below the major themes that emerged from analyses, highlighting some strikingly similar as well as important differences in patient and provider perspectives.

Table 1. Patient Characteristics*

Age	
Range	35–88
Mean	60
Race (n)	
White	13
African American	3
*Income** (n)*	
$15,000 or less	4
15,001–20,000	4
20,001–40,000	5
40,001–60,000	2
Educational status (n)	
Some high school	1
HS graduate	4
Some college/ technical school	6
Completed college	3
Some graduate school	2

*Data missing for two patients
**One patient chose not to indicate income

HIV Testing Should Be Routine

Both patients and providers concurred that HIV testing should be routine because they thought it was (1) good for patients and (2) good for public health. The most salient reason for testing for both patients and providers was the value in knowing one's HIV status. Patients and providers advocated that HIV should be equated with other chronic diseases—such as diabetes and cholesterol—for which testing is routine.

> "I'd like to see [HIV] become like everything else, diabetes, tuberculosis, anything else that we test for. When they do the blood screening, do the whole thing. It's a deadly disease." *(Patient)*

> "I mean, why should [HIV] be any different, you know? I can order a CBC. I can order a PSA and somebody can come back with a PSA of 150 and I know he's got metastatic prostate disease. Why should this be any different?" *(Provider)*

Both groups stated the importance of learning results for chronic diseases/conditions—including HIV—in order to commence treatment as quickly as possible, or, as one patient put it, "nipping it [HIV] in the bud," before the disease progresses. In this respect, managing HIV was viewed as no different from managing other chronic health problems. Participants favored including HIV testing along with other routine blood work typically included in primary care visits.

Both patients and providers also evoked public health reasons for knowing one's HIV status. As a serious public health threat, HIV testing and subsequent treatment were viewed as effective means of stemming the spread of the disease and protecting sexual partners. Some patients were motivated by misunderstandings about HIV transmission, arguing that knowing one's status could protect others with whom they have even casual contact, such as kissing or sharing a glass.

Routine Testing May Decrease HIV Stigma

Patients and providers also asserted that testing for HIV routinely would likely reduce some of the stigma associated with the disease. Routinizing HIV testing would change an unusual screening into an ordinary event, rendering HIV testing "normal," thereby diminishing the stigma. This could be accomplished by associating HIV testing with screening for other chronic diseases that have much less discernible stigma, such as diabetes.

> "If it was something that happens all the time, you could take the stigmatism *[sic]* out of it, maybe." *(Patient)*

One provider reflected that current processes make HIV testing anything *but* routine, and that routinizing testing would further decrease HIV stigma.

> "My own feeling is that the stigma associated with it…it's certainly decreasing.… And it's almost like we're creating processes that make it different than routine care." (*Provider)*

Stigma as a Barrier to HIV Testing

At the same time, HIV stigma remained a barrier for patients as reflected in their concerns about confidentiality and the potential impact on the patient-provider relationship. Confidentiality was paramount to both patients and providers. Patients wanted guarantees that both the fact they were being tested, and the test results, would be confidential. One patient, stating that he didn't trust the VA or the government to maintain confidentiality, said that he might choose to be tested elsewhere.

Patients and providers indicated that a patient who trusts his/her provider would be more likely to agree to being tested. However, patients and providers alike were concerned about the impact raising this sensitive topic might have on the patient-provider relationship. Emphasizing the wish to be offered rather than told to have the test, one patient said that being told to be tested might lead someone to "walk out of [the doctor's] office and see another doctor." One provider stated that simply asking a patient to be tested could alter their rapport, affect the level of the patient's trust and negatively impact their future relationship.

Barriers to HIV Testing in the Clinical Encounter

Patients and providers described two important barriers to HIV testing that occur during the clinical encounter: obtaining written informed consent and pre-test counseling. Patients viewed signing consent forms as anxiety-provoking, intimidating and sometimes difficult to understand because of the dense legalistic language. They indicated that the documents rendered HIV different from other diseases, running counter to their desire to normalize HIV by equating it with other chronic diseases. Once HIV is construed as exceptional—somehow different from other diseases—patients noted they were likely to be more apprehensive about being tested because of fears of a positive result.

> "They do all [other tests] over here, but [HIV testing] is different. This is 'Ooh, keep away from that.' You know, and that's what that form does, in my opinion." *(Patient)*

In addition, patients voiced concern about legal language in the consent forms indicating the possibility that test results might need to be released to "third parties." Because of fears of breached confidentiality, some patients expressed reservations about having to sign a consent form.

Providers agreed with patients that requiring written informed consent made HIV different from testing for other chronic diseases. In addition, the informed consent forms represented a formidable and time-consuming logistical obstacle. At one site, social workers conducted the consent process, but were not located in the primary care clinic, making, the providers said, the "whole situation so difficult, with the limited amount of time we have." Providers from the other site were encouraged to use computer-based consent forms and signature pads intended to streamline the process, but these were cumbersome in practice.

The second barrier to HIV testing was pre-test counseling, in which providers discuss the process of HIV testing, the potential implications of a positive test result and behaviors that may put individuals at risk for HIV. Patients viewed pre-test counseling as another process that rendered HIV excep-

tional, and therefore preferred not to participate in it. They noted that counseling would be likely to instill fear about HIV, and may even prevent them from being tested. Equating HIV once again with other chronic illnesses, patients said they wished to talk in-depth about HIV with a provider only if a test result were positive, at which time they would want information about the implications of a positive result and how best to manage the disease. For example, one patient, talking about routine TB testing, stated:

> "It was automatic that we'd get tested. I didn't need an explanation at the time of the testing. However, if it came back positive, you'd best like some explanation".

Providers viewed pre-test counseling as an obstacle to HIV testing because it took up many "precious primary care minutes." One provider feared that talking with patients about HIV testing in depth would lead to a flood of patient questions and concerns, and yet others acknowledged the moral necessity of discussing potential ramifications of a positive test. Doing justice to such questions and concerns was viewed as taking away time from accomplishing other important tasks during clinical appointments.

Patients identified an additional barrier to getting tested—the wait time between being tested and receiving the results. Several patients stated that waiting for results made them anxious and depressed, and when they considered being tested again would delay the test. One patient even stated, "I'd be swallowing tranquilizers while I'm waiting for the phone to ring." Providers did not discuss this aspect of testing.

Communicating About Routine Testing

Although patients suggested HIV testing be handled as routine rather than unusual, they repeatedly stated that they wished to be *asked* if they would like to have an HIV test, rather than be *told* they were going to be tested or have it be done automatically without discussion. One man, after saying he would agree to be tested stated, "But I don't think it should be automatic. A lot of people might take umbrage at that." Others reported similar views, with statements such as:

> "…as long as they use the word 'should,' and not 'have to." *(Patient)*

> "I'm not against it, absolutely not. But again I just don't want something else shoved down my throat." *(Patient)*

Above all, patients desired the autonomy to be able to choose and consent to be tested for HIV. Some patients equated their wish for the autonomy to decide about testing with their fundamental "civil liberties," one stating that he wanted a choice in order to "preserve as much of my freedom as possible."

Patients stated they would be more likely to agree to be tested if providers communicated several key information points when raising the topic. First, patients wanted to be told, "We're going to test everybody, everybody," indicating the goal of making the test not only routine, but also universal.

Patients indicated that they needed to be reassured they were not being singled out to be tested based on clinical signs of the disease. One patient stated:

> "If he suggested he wanted to test me for AIDS, he must have saw something in my crit [*sic*] counts or something like that, that are fallen enough that he wanted to do the research on me." *(Patient)*

Thus, patients who were asymptomatic and had no apparent risk factors wished to be told that HIV testing was being suggested as a general policy, not because of suspected HIV disease.

Patients found that learning that 25% of all HIV infected patients are unaware of their status, in combination with the test being 'offered,' would be influential in achieving testing (note that this figure has changed to 21% since the time of the focus groups). One man said it most succinctly:

> I think any time you go see a doctor, it should be offered to you. "Would you like it?" And this information should be given to each patient, that a quarter of the people don't know they have it, and offer them, "Would you like to have an HIV test because of this information?" *(Patient)*

Providers similarly found that information from the CDC materials about HIV testing would be helpful in talking with patients, particularly the information noting that 25% of all HIV infected patients are unaware of their status. Several providers stated that they felt ill-prepared to raise the topic of HIV testing with patients. One even said that some sort of written script would be helpful. Others suggested that one way to routinize testing would be to inform patients that they would be tested for HIV just as they inform patients about blood tests for other health conditions, such as thyroid disease or elevated cholesterol. They argued that by making it completely matter-of-fact, patients would be more likely to agree to being tested. One provider stated that he would say the following to patients:

> "I'm going to order a bunch of blood tests including cholesterol, kidney function, liver function. One of the tests that's now recommended is that we also check for HIV. I'm going to go ahead and order that." *(Provider)*

Veterans may be Different

The fact that the patients were veterans of the US Armed Forces is an important point that was raised by the patients themselves. They contended that veterans as a group are more likely than the general population to engage in unprotected sex and use drugs and alcohol, putting them at higher risk of HIV infection. Therefore, they said, veterans may be more open to being tested than civilians. One man stated:

> "I was tested when I was in California the first time in '95 when I was in the Marine Corps. And, you know, I was nervous then, just because, you know, we had had unprotected sex with, you know, a female, and it was a

bunch of us and it was just crazy.... It was just really reckless." *(Patient)*

Another man stated:

"I think, you know, the veterans and stuff [who's] been around the lower side of life a little bit, could accept it more than somebody that goes to the Calvary Baptist Church every week." (Others laugh.) *(Patient)*

Although this man recognized that "being on the lower side of life," leads to a higher risk for HIV, the idea that anyone could become infected remained salient to him. Thus, he advocated for routine testing. Providers did not comment on the uniqueness of the veteran population regarding HIV risk.

DISCUSSION

Under rapidly changing accepted HIV testing recommendations, providers may take different approaches to discussing testing with patients. The CDC-recommended 'opt-out' approach to routine HIV testing is a dramatic change from typical past practices[20]. At present, there is no clear consensus on exactly how providers should present 'opt-out' testing to patients, and no other studies have examined patient perspectives on routine testing. Patients and providers in our study expressed clear perspectives about how to best achieve routine testing. Our findings have implications for implementing the new CDC recommendations, supporting routine testing of all adults, eliminating written informed consent and reducing the time burden of HIV testing by eliminating extensive pre-test counseling.

We found that patients were open to being tested and supportive of testing as a routine aspect of regular care, similar to others' findings that hospitalized patients responded positively to an unsolicited HIV test[10]. However, our findings highlight veterans' wishes to be tested only after being given information to make an independent decision and provide verbal consent for testing. This is important for providers to note, especially as some of our providers, in line with the CDC 'opt-out' guideline, recommended informing the patient they would order an HIV test unless it was actively refused.

The concomitant perspective that both routine testing and this more 'opt-in' scenario are desirable for patients is understandable given the stigma associated with HIV disease[21–23]. Although HIV is less stigmatized than in the past, patients remain concerned about the confidentiality of testing and results. Previous research shows that providers can effectively communicate and request routine testing with at-risk individuals, contributing to destigmatization of the disease[24]. To conduct HIV testing in a patient-centered manner, we must consider patients' concerns alongside those of the medical and public health communities.

Barriers to testing noted by our providers, including concern about the patient-provider relationship, time constraints and the consent process, have all been noted by other studies in a variety of settings[25]. Like most primary care providers, the providers in our study felt constrained by time limits imposed on them for clinical appointments, and adding routine HIV testing to their responsibilities seemed onerous. Easing requirements for pre-test counseling and written consent should result in reduced provider burden. Providers can be patient-centered in ordering this sensitive test by *offering* the test to patients, reassuring patients that HIV testing is now routine for all adults, stating that patients are not exhibiting clinical signs of HIV, providing some basic information about HIV and asking patients if they would like to be tested.

Based on these findings, we developed the *6 R*'s for routine testing—six steps for providers to use in patient-centered discussions of HIV testing: (1) ***R***aise the topic of HIV testing; (2) ***R***eassure the patient that he/she is not showing clinical signs of the disease; (3) provide ***R***ationale that many patients infected with HIV are not aware of their status; (4) ***R***espond to any questions that the patient may have about HIV disease; (5) ***R***equest permission to order the test; (6) tell the patient when he/she can expect to get the ***R***esults. A sample script for using the *6 R's* is provided in Table 2.

There are several limitations to our study. We conducted our study at only two VAs; patients and providers from other regions in the US may have different views not expressed here. All veteran patients enrolled in the focus groups were men, and women may have different perspectives about routine HIV testing[10]. As noted by the veterans themselves, veterans may be different from the civilian population. In addition, other studies have identified the costs of HIV testing and funding for HIV treatment as barriers to testing[8]. In the VA, cost is not a barrier because HIV testing and treatments are covered fully for most HIV-positive veterans or require minimal out-of-pocket expense. Approximately half of patients who were eligible were unable to participate in our focus groups due to scheduling constraints. Nonparticipating individuals may have different views. We also had a limited number of providers who participated in the focus groups, meaning that we may not have fully appreciated the range of perspectives of the providers at these sites. Finally, providers who choose to work within the VA system may differ in their perspectives about care for vulnerable patients from primary care providers elsewhere.

Using the *6 R's*, we propose, may help providers discuss HIV testing with their patients, and make patients comfortable with being testing for HIV disease. In contrast to 'opt-out' testing, this communication strategy may alleviate the risk of patients' being tested without their knowledge, which could put them at additional psychological and social risk if they are positive.

Future research is needed to fully explore implementation of routine testing guidelines, examining both an opt-out ap-

Table 2. Sample Script for Using the 6 R's for Routine Testing

Raise	"We are now offering to test everybody for HIV just as we test for other things like diabetes or heart disease, through a simple blood test"
Reassure	"There's nothing I know about you or your health that makes me think you might have HIV"
Rationale	"It is possible to have HIV for a long time without knowing it. Of the people who have HIV as many as 25% of them don't know they have it"
Respond	"Lots of people have questions about HIV disease or getting tested. What questions do you have for me?"
Request	"Given what we've discussed, would you like me to test you for HIV along with your other blood tests today?"
Results	"It takes about 2 weeks for the results to come back. Waiting for the results can make people nervous. We will contact you as soon as we receive the results"

proach as proposed by CDC and the more patient-centered approach proposed here. Such research should examine the impact of these communication approaches on HIV testing rates, the impact on the patient-provider relationship and the cost-effectiveness of each approach.

Increasing the number of patients tested for HIV in the US is a high priority, to achieve earlier detection, earlier treatment and earlier secondary prevention through education about HIV transmission [26]. However, the ultimate goal of early detection is to transition newly positive patients into care smoothly and rapidly. If routine HIV testing is achieved at the cost of weakening veterans' trust in the health-care system, this could lead patients to delay returning for test results, or it could lead to delayed transition into care after testing.

Putting an end to the 'exceptionalism' of HIV requires an ongoing effort to routinize and destigmatize HIV testing and the disease itself [20]. In the hands of some doctors, a brief script in which HIV is embedded in a list of other needed blood tests may stifle opportunities for question-asking and fall short of standards for patient-centered communication. A patient-centered approach to streamlined HIV test discussion may improve the likelihood that patients accept testing and eventually lead to fully routine HIV testing in primary care.

Acknowledgements: *A version of this paper was presented at the VA Health Services Research and Development Annual Meetings, February 2008, and at the Society for General Internal Medicine Meetings, April 2008. Funding for this study was provided by the Department of Veterans Affairs, Health Services Research and Development Service grant no. RRP 07–281. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. We would like to thank Dr. Wendell Ching and Dr. James Schlosser for their assistance in conducting this study, as well as research assistants, Linda McCoy, Anne Taylor and Lee Ann Lowe.*

Conflict of interest: *The authors have no conflict of interest with the research presented in this paper.*

Corresponding Author: *Barbara G. Bokhour, PhD; Center for Health Quality Outcomes Economic Research, ENRM Veterans Affairs Medical Center, 200 Springs Road (152), Bedford, MA 01730, USA (e-mail: Barbara.Bokhour@va.gov).*

REFERENCES

1. **Walensky RP, Freedberg KA, Weinstein MC, Paltiel AD.** Cost-effectiveness of HIV testing and treatment in the United States. Clin Infect Dis. 2007;45 (Suppl 4):S248–54.
2. **Owens DK, Sundaram V, Douglass LR, et al.** Seroprevalence of HIV infection at VA health care systems. Med Decis Mak. 2003;23:569.
3. **Sanders GD, Bayoumi AM, Sundaram V, et al.** Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. N Engl J Med. 2005;352(6):570–85.
4. **Paltiel AD, Weinstein MC, Kimmel AD, et al.** Expanded screening for HIV in the United States–an analysis of cost-effectiveness. N Engl J Med. 2005;352(6):586–95.
5. Centers for Disease Control and Prevention. HIV prevalence estimates - United States, 2006. 2008;57:1073–1076.
6. Centers for Disease Control and Prevention. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. MMWR. 2006;55(RR-14):1–17.
7. **Qaseem A, Snow V, Shekelle P, Hopkins R, Jr., Owens DK.** Screening for HIV in Health Care Settings: A Guidance Statement From the American College of Physicians and HIV Medicine Association. Ann Intern Med. 2008.
8. **Holtgrave D.** Costs and consequences of the US Centers for Disease Control and Prevention's recommendations for opt-out HIV testing. PLoS Med. 2007;4(6):e.
9. **Bartlett JG, Branson BM, Fenton K, Hauschild BC, Miller V, Mayer KH.** Opt-out testing for human immunodeficiency virus in the United States: progress and challenges. Jama. 2008;300(8):945–51.
10. **Obermeyer CM, Osborn M.** The utilization of testing and counseling for HIV: a review of the social and behavioral evidence. Am J Public Health. 2007;97(10):1762–74.
11. **Millen JC, Arbelaez C, Walensky RP.** Implications and impact of the new US centers for disease control and prevention HIV testing guidelines. Curr Infect Dis Rep. 2008;10(2):157–163.
12. **Chesney MA, Smith AW.** Critical delays in HIV testing and care - The potential role of stigma. Am Behav Sci. 1999;42(7):1162–1174.
13. **Epstein R, Morse D, Frankel R, Frarey L, Anderson K, Beckman H.** Awkward moments in patient-physician communication about HIV risk. Ann Intern Med. 1998;128(6):435–42.
14. **Epstein RM, Levenkron JC, Frarey L, et al.** Improving physicians' HIV risk-assessment skills using announced and unannounced standardized patients. J Gen Intern Med. 2001;16:176–180.
15. **Rogers WH, Kazis LE, Miller DR, et al.** Comparing the health status of VA and non-VA ambulatory patients: the veterans' health and medical outcomes studies. J Ambul Care Manage. 2004;27(3):249–62.
16. **Owens DK, Sundaram V, Lazzeroni LC, et al.** Prevalence of HIV infection among inpatients and outpatients in Department of Veterans Affairs health care systems: implications for screening programs for HIV. Am J Public Health. 2007;97(12):2173–8.
17. **Strauss AL. Qualitative Analysis for Social Scientists.** Cambridge [Cambridgeshire]; New York: Cambridge University Press; 1987.
18. **Charmaz K.** Constructing Grounded Theory. Thousand Oaks, CA: Sage; 2006.
19. NVIVO 7.0 qualitative data software program]. Qualitative Solutions and Research, International; 2007.
20. **Bayer R, Fairchild AL.** Changing the paradigm for HIV testing-the end of exceptionalism. N Engl J Med. 2006;355:647–69.
21. **Anderson M, Elam G, GErvere S, Solarin I, Fenton K, Estabrook P.** HIV/AIDS stigma and discrimination: Accounts of HIV-positive Carribean people in the United Kingdom. Soc Sci Med; 2008.
22. **Talley A, Bettencourt B.** A relationship-oriented model of HIV-related stigma derived from a review of the HIV-affected couples literature. AIDS Behavior; 2008.
23. **Meiberg A, Bos A, Onya H, Schaalma H.** Fear of stigmatization as barrier to voluntary HIV counselling and testing in South Africa. East Afr Public Health. 2008;5(2):49–54.
24. **Sobo E, Bowman C, Asch SA, Goetz MB, Gifford AL.** "A routine thing": Implementation of a computerized HIV test clinical reminder for veterans. Anthroplogy and Medicine. 2008;15(3):213–25.
25. **Burke RC, Sepkowitz KA, Bernstein KT, et al.** Why don't physicians test for HIV? A review of the US literature. Aids. 2007;21(12):1617–24.
26. **Valdisseri R.** Late HIV diagnosis: Bad medicine and worse public health. PLoS Med. 2007;4:e200.

Predictors of Medication Adherence for African American Patients Diagnosed with Hypertension

Julia Braverman, PhD; Julien Dedier, MD, MPH

Background: The prevalence, morbidity and mortality of hypertension are strikingly higher for African Americans than for Whites. Poor adherence to the antihypertensive medication regimen is a major cause of inadequate blood pressure control. In this study, we assess the relationship of antihypertensive medication adherence to sociodemographic, clinical and cognitive characteristics of urban African American adults.

Method: Data were drawn from a larger randomized controlled trial assessing the effect of a behavioral intervention to improve medication adherence and blood pressure control among hypertensive African American patients followed in an urban primary care network. Medication adherence was assessed at baseline using the Medication Event Monitoring System (MEMS) – a method regarded as the gold standard for assessing medication adherence in clinical research. Information on potential correlates of medication adherence (sociodemographic, clinical and cognitive) was obtained at baseline by computer-assisted interview. We assessed the cross sectional association of these factors to medication adherence in baseline data.

Results: Medication adherence was significantly associated with systolic blood pressure ($r=.253$, $P<.04$) and self-reported medication adherence ($r=.285$, $P<.03$). The relationship of education to medication adherence varied significantly by sex ($P<.05$ for interaction). Specifically, lower educational attainment was related to higher adherence among men, but lower adherence among women.

Conclusion: Identifying correlates of low antihypertensive medication adherence and their interactions, as in this study, will help health providers to better recognize patients at higher risk for worse hypertension-related outcomes. This knowledge can also inform interventions which target a higher-risk subset of hypertensive patients. (*Ethn Dis*. 2009;19:396–400)

Key Words: Hypertension, Medication Adherence, African Americans, Urban Population

From Division of Clinical Informatics, Beth Israel Deaconess Medical Center (JB), Medical Information Systems Unit, Boston Medical Center (JD).

Address correspondence and reprints request to Julia Braverman; Beth Israel Deaconess Medical Center; Division of Clinical Informatics; 1330 Beacon St., Suite 400; Boston, MA 02215; 617-278-8166; 617-278-8188 (fax); 617-278-8166; braverju@gmail.com

Introduction

In the United States, hypertension is 30%–50% more prevalent in African Americans than Whites[1] and accounts for half of the excess cardiovascular mortality observed in African Americans vs Whites.[2] Of those diagnosed with hypertension, African Americans have lower rates of blood pressure (BP) control than Whites, and this disparity has increased over time.[3] Poor adherence to prescribed antihypertensive medications contributes significantly to lower rates of BP control and, among modifiable risk factors, increasing such adherence is considered to have the greatest potential to improve BP control.[4] Further, studies suggest that as a group African Americans may have lower adherence to the antihypertensive medication regimen than Whites[5–6] and may benefit substantially from interventions to promote adherence to the BP medication regimen.

Unfortunately, physicians are generally not good at identifying poor medication adherence among their African American patients. In a study of young, urban, hypertensive, African Americans, primary care providers were unable to identify poor medication adherence 60% of the time.[7] Knowledge of the correlates of poor medication adherence would help physicians and health educators to identify patients at risk for poorer BP-related health outcomes. This knowledge is also a first step toward developing interventions to improve antihypertensive medication adherence among African Americans.

In this study, we assess the relationship of antihypertensive medication adherence to sociodemographic, clinical and cognitive characteristics of urban African American adults.

In this study, we assess the relationship of antihypertensive medication adherence to sociodemographic, clinical and cognitive characteristics of urban African American adults.

Method

Participants

This study was part of a larger randomized control trial of a telephone-based behavioral intervention to improve medication adherence, physical activity and dietary behaviors among hypertensive, urban, African American adults of low socioeconomic status. We present results of a cross sectional analysis using baseline data.

Participants were drawn from the primary care practices of a large urban teaching hospital, and four of its affiliated neighborhood health centers. Eligibility criteria were as follows: 1) a physician diagnosis of hypertension; 2) being at least 35 years old; 3) being non-adherent to dietary recommendations for hypertension; 4) having an active prescription for at least one antihypertensive medication; 5) having two elevated clinic blood pressure

readings (≥140/90 or ≥135/85 if diabetic) within the proceeding 6 months; 6) non-adherence to physical activity recommendations; 7) non-adherence to dietary recommendations for hypertension; 8) the ability to understand spoken English; 9) regular access to a telephone; and 10) self-reported African American race/ethnicity.

Procedure

The electronic health record (EHR) was used to identify individuals who satisfied the first four eligibility criteria. Apparent race/ethnicity, also captured in the EHR, was used as an initial proxy for African American race/ethnicity. Next, each participating primary care provider was shown a list of potentially eligible patients and asked to remove anyone who did not meet eligibility criteria. Study personnel then contacted the remaining patients by telephone to establish eligibility. Five weeks later, a research assistant went to the subject's home to confirm eligibility, obtain written informed consent, and to place Medication Event Monitoring System (MEMS) caps on up to 3 antihypertensive medication pill bottles. MEMS caps contain a microprocessor that records the time and date that a patient opens the pill bottle to obtain a dose of medication. Such electronic monitoring devices are considered the gold standard for assessing rates of medication adherence in clinical research.[8] During the next home visit 6 weeks to one year later, a research assistant downloaded MEMS caps adherence data and collected baseline data on all other variables used in this analysis. The Boston Medical Center institutional review board approved the study.

Measures

Medication adherence

Medication adherence was viewed as the dependent variable in these analyses, and was measured using MEMS caps applied to up to 3 BP medication bottles as described above. Because the duration between placing the caps and uploading data varied, by patient, from 6 weeks to 1 year, we used data from a middle 30-day period in the analyses. Powerview communication software (Aardex Corporation, Union City, CA) was used to read and download adherence data from each MEMS cap. For each medication, we calculated the percentage of prescribed doses taken each day during the 30-day period. We then averaged this number over all 30 days to determine the average percentage of prescribed doses taken per day. For participants who had more than one monitored medication, we averaged this value across all medications. As a second measure of medication adherence, we calculated the average percentage of days the monitored medication was taken as prescribed. Therefore, we had two summary measures of medication adherence for each subject.

Candidate correlating variables

A list of potential correlates of poor medication adherence was derived from the literature[9] and supplemented by clinical opinion. As noted above, information on each of these factors was obtained by trained research assistants at the second home visit via computer-assisted personal interviews.

The participants' subjective financial status was assessed by asking them to classify their current financial situation as: 1) comfortable, with enough money for extras; 2) enough to pay the necessary bills without cutting back, but not extras; 3) enough to pay the bills, but have had to cut back; or 4) cannot pay some bills no matter how hard I try.

Household income was assessed by asking participants to approximate their household income in the previous year as: <$10,000; $10,000 to $20,000; $20,001 to $30,000; $30,001 to $40,000; $40,001 to $50,000; or >$50,000.

Employment type was assessed by asking the participants to indicate if they were employed and, if so, whether full- or part-time. If unemployed, participants could indicate if they were disabled, retired, a student or a homemaker.

Insurance type was assessed using the question, "How do you cover your health care costs?" Possible responses were: Medicare/Medicaid, self-insured, employer-paid, veteran's benefits, free care, or other.

Self-efficacy for taking medication was assessed using a 51-item measure consisting of all 43 candidate items tested by Ogedegbe et al,[10] and 8 additional experimental items. On a scale from 1 to 5, participants rated how confident they were that they could take their blood pressure medication as prescribed under certain adverse conditions, such as: when you are busy at home, when you are tired, or in the presence of people other than relatives or friends. Individual item scores were summed to create an overall score for each subject ranging from 51 (low self-efficacy) to 255 (maximal self-efficacy).

We assessed self-reported medication adherence using the 7-item version of the Morisky survey, a validated measure of adherence to medication regimens.[11] Each response tail contained 2 to 5 possible options and was scored from 1 to 5. Individual item scores were summed to create an overall score for each subject ranging from 7 (poor adherence) to 17 (maximum adherence). This total score was used in the analyses.

Physician support was assessed using the single item, "How much encouragement for taking your hypertension medication as prescribed do you get from your doctor?" scored on a 5-point Likert scale.

Family support was assessed using a similar item, "How much encouragement for taking your hypertension medication as prescribed do you get from your friends and family?"

In addition, every subject was asked to select 3 values of personal importance from a list of 12 including responsibil-

Table 1. Baseline characteristics of study participants

	Current sample (*N*=70)	Sample for the larger RCT (*N*=337)
Age in years (mean±SE)	58±11	56.5
Males	21 (30%)	100 (30%)
Full-time or part-time employment	27 (39%)	132 (39%)
Education (years)	12.2	12.1
History of stroke	6 (8%)	25 (7.5%)
Diabetic	31 (44%)	129 (38%)
Systolic BP	131±17	131.2
Diastolic BP	80±10	80.6
Mean annual household income		
<10 K	9 (27%)	122 (36%)
10–20 K	21 (30%)	85 (25%)
21–30 K	10 (14%)	48 (14%)
31–40 K	4 (6%)	23 (7%)
41–50 K	3 (4%)	9 (3%)
>50 K	3 (4%)	17 (5%)
Non reporters	10 (14%)	33 (10%)
# with medication insurance benefit	68 (97%)	333 (99%)
Number of BP medications		
1	39%	36%
2	30%	37%
3	31%	27%

ity, independence, God's will, physical strength, etc.

Statistical Analysis

The following were coded as categorical variables in the analyses: sex, financial status (4 groups, as above), living alone or not, insurance and employment types, diabetes status (Y/N), history of stroke (Y/N), medication class (beta blocker, diuretic or calcium channel blocker), and personal values. The following were coded as continuous variables in the analyses: education, self-reported income, blood pressure, number of prescribed medications, weight, self-efficacy for taking medication, and level of physician and family support. Frequencies were calculated for categorical variables, and means and standard errors for continuous variables. We conducted a Pearson correlation and a curve analysis for all continuous variables with MEMS-derived medication adherence to identify possible linear and non-linear correlates of medication adherence. For binary categorical variables, we conducted *t* tests to assess their relation to medication adherence. For those with 3 or more categories we performed a one-way ANOVA. We used a two-way ANOVA to test for interaction effects of 2 or more categorical variables on medication adherence.

RESULTS

MEMS data were collected for all 337 randomized participants in the larger clinical trial as described above. However, we report findings on the 70 participants' baseline MEMS data available for analysis. Table 1 shows the baseline characteristics of these 70 participants and of all participants enrolled in the larger clinical trial. The 2 groups are comparable across the variables shown and, therefore, are also likely to have had similar baseline levels of medication adherence. Twenty-seven (39%) participants took one antihypertensive medicine, 21 (30%) took two, and 22 (31%) participants took three. The two MEMS-derived measures of overall adherence were highly correlated with each other (ie, the mean percentage of prescribed doses taken per day and the mean percentage of days taken as prescribed, $r=.94$, $P<.001$). For this reason, we used only the latter measure as the dependent variable in subsequent analyses. Mean medication adherence was 71.6%, median=82.3%, sd=26%.

Table 2 shows the correlation between MEMS-derived medication adherence and selected study variables. Medication adherence was significantly correlated with systolic blood pressure ($r=.253$, $P<.04$) and self-reported medication adherence ($r=.285$, $P<.03$). It was marginally correlated with self-efficacy for medication adherence ($r=.198$, $P<.09$). No curve fit analysis resulted in significant results.

There was no significant difference in adherence between males and females, however, there were strong interactions between: 1) sex and education (Figure 1), and 2) sex and whether or not the individual lived alone (Figure 2). With respect to education, when education was coded as high school and above or below high school, females with a less than high school education were less adherent (61%) than females with more formal education (74%). The opposite pattern was observed for males as those with lower education were more adherent than

Table 2. Significant and marginally significant correlations between MEMS-derived medication adherence and other variables

Variable	*r*	Significance level
Systolic blood pressure	.253	$P<.04$
Self-efficacy	.285	$P<.03$
Self-reported medication adherence	.198	$P<.09$

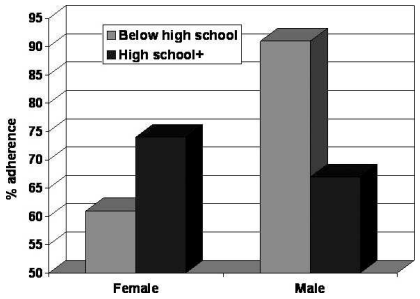


Figure 1. Effects of sex and education level on medication adherence.

those with higher education (91% vs. 72%, respectively; F [1, 47]=4.244; $P<.05$). (Figure 1).

The interaction between sex and whether or not the individual lived alone approached marginal statistical significance (F [1,66]=3.38, $P<.08$). Of the 70 subjects, 31 lived alone and 39 lived with someone else. Females who lived alone were more adherent (74%) than females who lived with someone else (65%); whereas, males who lived alone were less adherent (69%) than males who lived with someone else (86%) (Figure 2).

We assessed for confounding of the interaction effects described above. We were particularly interested in the effects of per capita income and age as these two variables are intuitively related to education, sex and to whether or not one lives alone.

Correlation analysis did not demonstrate any relation between age and medication adherence. However, it was found that people who lived alone were

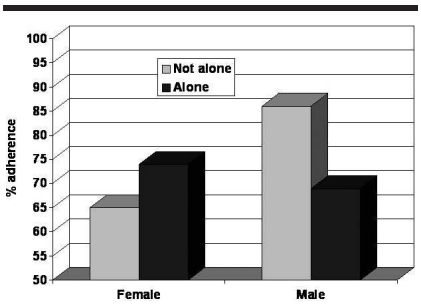


Figure 2. Effects of sex and living alone or not on medication adherence.

older (65 vs 54, $P<.01$). No significant interactions between age, sex and living alone were found.

The per capita income was obtained by dividing the upper limit of the houshold income category by the number of people in the houshold. Not surprisingly, females' self-reported income was lower than males' ($10.6 K vs $18.4 K, $P<.001$). No effect of income on medication adherence was found, either by itself or in an interaction with other variables.

A separate one-way ANOVA test and a single multiway ANOVA were conducted to determine if adherence was related to insurance type, actual income or participants' perceived financial status. No test reached statistical significance or showed a promising trend. This may be explained by the fact that all but 2 subjects reported that their insurance covered the cost of medications.

DISCUSSION

The purpose of this study was to identify correlates of medication adherence among urban, hypertensive, African American adults of low socioeconomic status. Unlike most other studies of medication adherence in this population, we assessed adherence using MEMS technology which is considered a more objective and rigorous assessment approach than self-report alone. We observed a mean adherence rate of 71.6% in our sample, which is similar to that of other studies in similar socio-demographic populations. We also observed a trend between self-efficacy for medication taking and actual medication adherence which is consistent with prior research.[9,10]

A novel finding is the interaction observed between education and medication adherence by sex. While the least educated females tended to demonstrate the lowest adherence, the least educated males demonstrated the highest adherence, on average. Similarly, high school

While the least educated females tended to demonstrate the lowest adherence, the least educated males demonstrated the highest adherence, on average.

and above education was associated with relatively high adherence among women, but only moderate to low adherence among men. A potential explanation for this finding is that women with lower education are more likely than more educated women to be preoccupied with children and families. Indeed, the literature does suggest that caring for dependents is associated with lower medication adherence. [9] This being the case, one might expect to find a direct correlation between education and employment status, and between employment status and medication adherence. No such relation was found.

The moderating effect of education on medication adherence observed in our study may also parallel the effect of general IQ on behavior change described in McGuire's Informational Processing Theory.[12] This model states two factors must be present in order for someone to comply with a request or agree with a message: 1) reception (ie, understanding of the message), and 2) yielding (ie, accepting of the message). People with a higher IQ or level of education typically understand the message better than those with a lower IQ or educational level. However, they are also more likely to mount counterarguments and to display resistance to the request or message. This tendency toward resistance is also higher among people with higher self-esteem.[13] In our study sample, one can argue that the males with higher educational attainment possessed higher self-esteem[14] and were, therefore, less adherent to their

medication regimen on that basis. On the other hand, educational attainment has less influence on women's level of self-esteem. This is because women's self-esteem is shown to be more influenced by emotional factors (eg, relational harmony) rather than by achievement.[14,15] Therefore, for the women in our sample higher educational attainment may have increased understanding of the message about improving medication adherence without increasing resistance to that message.

We also noted a marginal interaction of sex, and whether one lived alone, on medication adherence. Specifically, women who lived alone were more adherent than women who lived with someone else (presumably, a family member), whereas, men who lived alone were less adherent than men who lived with someone else. As argued above, females who live with others may invest in meeting the needs of their dependents such that their own medication adherence suffers. Conversely, medication adherence among men who have a live-in partner would be expected to improve, as was observed in this study.

Although electronic monitoring systems such as the one used in this study are considered to be the most reliable measure of drug adherence, their use may potentially increase adherence by the so-called Hawthorne effect.[16] This effect is strongest during the initial period of monitoring, but wanes over time. [17] In this study we used a middle 30-day period of MEMS data, rather than an earlier period, to calculate medication adherence. We believe this approach mitigates the influence of MEMS monitoring on actual medication adherence. Moreover, the adherence rate reported in our study is similar to rates reported in studies of medication adherence that did not use electronic monitoring systems.[9] In addition, the 70 participants in our study were similar to those in the larger randomized trial across a range of characteristics. These facts argue against the presence of significant monitoring or selection bias in the medication adherence rates we observed.

Multiple factors that influence patient adherence to prescribed therapies have been described. These include quality of life, complexity and side effects of medications, health care system issues, demographic, behavioral, treatment and clinical variables, lack of knowledge regarding hypertension, to name only a few.[9] Although these factors are variably important when considered in isolation, our findings highlight the need to consider how individual risk factors for poor medication adherence interact to increase or reduce the likelihood of medication non-adherence. Identifying correlates of low antihypertensive medication adherence will help healthcare providers to better recognize patients at higher risk for poor adherence and worse hypertension-related outcomes. A knowledge of these risk factors will also inform the development of interventions which target a higher-risk subset of hypertensive patients.

REFERENCES

1. Shea S, Misra D, Ehrlich MH, Field L, Francis CK. Correlates of nonadherence to hypertension treatment in an inner-city minority population. *Am J Public Health*. 1992;82: 1607–1612.
2. Wong MD, Shapiro MF, Boscardin WJ, Ettner SL. Contribution of major diseases to disparities in mortality. *N Engl J Med*. 2002;347(20):1585–1592.
3. Hertz RP, Unger AN, Cornell JA, Saunders E. Racial disparities in hypertension prevalence, awareness, and management. *Arch Intern Med*. 2005;165(18):2098–2104.
4. Halpern MT, Khan ZM, Schmier JK, et al. Recommendations for evaluating compliance and persistence with hypertension therapy using retrospective data. *Hypertension*. 2006;47:1039–1048.
5. Siegel D, Lopez J, Meier J. Antihypertensive medication adherence in the Department of Veterans Affairs. *Am J Med*. 2007;120:26–32.
6. Bosworth HB, Powers B, Grubber JM, et al. Racial differences in blood pressure control: potential explanatory factors. *J Gen Intern Med*. 2008;23(5):692–698.
7. Hill MN, Bone LR, Kim MT, Miller DJ, Dennison CR, Levine DM. Barriers to hypertension care and control in young urban black men. *Am J Hypertens*. 1999;2: 951–958.
8. Farmer KC. Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice. *Clin Ther*. 1999;21:1073–90.
9. Hyre AD, Krousel-Wood MA, Muntner P, Kawasaki L, DeSalvo KB. Prevalence and predictors of poor antihypertensive medication adherence in an urban health clinic setting. *J Clinical Hypertens*. 2007;9:179–186.
10. Ogedegbe G, Mancuso C, Allegrante J, Charlson M. Development and evaluation of a medication adherence self-efficacy scale in hypertensive African-American patients. *J Clin Epidemiol*. 2003;56:520–529.
11. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care*. 1986;24(1):67–74.
12. McGuire WJ. Personality and susceptibility to social influence. In: Borgatta EF, Lambert W, eds. *Handbook of Personality Theory and Research*. Chicago: Rand McNally; 1968.
13. Gudjonsson GH, Sigurdsson JF. The relationship of compliance with coping strategies and self-esteem. *Eur J of Psychol Assess*. 2003;19: 117–123.
14. Hollender J. Sex differences in sources of social self-esteem. *J Consult Clin Psychol*. 1972;38: 343–347.
15. Reid A. Gender and sources of subjective well-being. *Sex Roles*. 2004;51:618–632.
16. Bailey DM. *Research for the Health Professional: A Practical Guide*. Philadelphia: FA Davis; 1997.
17. Levine A, Hinkin C, Marion S, et al. Adherence to antiretroviral medications in HIV: differences in data collected via self-report and electronic monitoring. *Health Psychol*. 2006;25(3):329–335.

AUTHOR CONTRIBUTIONS

Design concept of study: Braverman, Dedier
Acquisition of data: Braverman, Dedier
Data analysis and interpretation: Braverman, Dedier
Manuscript draft: Braverman, Dedier
Statistical expertise: Braverman, Dedier
Acquisition of funding: Braverman, Dedier
Administrative, technical, or material assistance: Braverman, Dedier
Supervision: Braverman, Dedier

The Effectiveness of Limiting Alcohol Outlet Density As a Means of Reducing Excessive Alcohol Consumption and Alcohol-Related Harms

Carla Alexia Campbell, MHSc, Robert A. Hahn, PhD, MPH, Randy Elder, PhD, Robert Brewer, MD, MSPH, Sajal Chattopadhyay, PhD, Jonathan Fielding, MD, MPH, MBA, Timothy S. Naimi, MD, MPH, Traci Toomey, PhD, Briana Lawrence, MPH, Jennifer Cook Middleton, PhD, the Task Force on Community Preventive Services

Abstract: The density of alcohol outlets in communities may be regulated to reduce excessive alcohol consumption and related harms. Studies directly assessing the control of outlet density as a means of controlling excessive alcohol consumption and related harms do not exist, but assessments of related phenomena are indicative. To assess the effects of outlet density on alcohol-related harms, primary evidence was used from interrupted time–series studies of outlet density; studies of the privatization of alcohol sales, alcohol bans, and changes in license arrangements—all of which affected outlet density. Most of the studies included in this review found that greater outlet density is associated with increased alcohol consumption and related harms, including medical harms, injury, crime, and violence. Primary evidence was supported by secondary evidence from correlational studies. The regulation of alcohol outlet density may be a useful public health tool for the reduction of excessive alcohol consumption and related harms.

(Am J Prev Med 2009;37(6):556–569) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Excessive alcohol consumption, including both binge drinking and heavy average daily alcohol consumption, is responsible for approximately 79,000 deaths per year in the U.S., making it the third-leading cause of preventable death in the nation.[1] Approximately 29% of adult drinkers (≥18 years) in the U.S. report binge drinking (five or more drinks on one or more occasions for men and four or more drinks for women) in the past 30 days, as do 67% of high school students who drink.[2,3] The direct and indirect costs of excessive alcohol consumption in 1998 were $184.6 billion.[4] The reduction of excessive alcohol consumption is thus a matter of major public health and economic interest.

From the Community Guide Branch of the National Center for Health Marketing (Campbell, Hahn, Elder, Chattopadhyay, Lawrence, Middleton); National Center for Chronic Disease Prevention and Health Promotion (Brewer, Naimi), CDC, Atlanta, Georgia; Los Angeles County Department of Health Services (Fielding), Los Angeles, California; and University of Minnesota School of Public Health (Toomey), Minneapolis, Minnesota

Address correspondence and reprint requests to: Robert A. Hahn, PhD, MPH, Community Guide Branch, Division of Health Communication and Marketing, Centers for Disease Control and Prevention, 4770 Buford Highway, Mailstop E-69, Atlanta GA 30333. E-mail: rhahn@cdc.gov.

The density of retail alcohol outlets is often regulated to reduce excessive alcohol consumption and related harms. Alcoholic beverage outlet density refers to the number of physical locations in which alcoholic beverages are available for purchase either per area or per population. An outlet is a setting in which alcohol may be sold legally for either on-premises or off-premises consumption. On-premises settings may include restaurants, bars, and ballparks; off-premises settings may include grocery and convenience stores as well as liquor stores. In 2005, the most recent year for which data are available, there were more than 600,000 licensed retail alcohol outlets in the U.S., or 2.7 outlets per 1000 population aged ≥18 years.[5] The number of outlets per capita in states with state-owned retail outlets varied from a low of 0.48 per 1000 residents in Mississippi to a high of 7.25 per 1000 in Iowa.[5]

Alcohol outlet density is typically controlled by states. Under state jurisdiction, outlet density may be regulated at the local level through licensing and zoning regulations, including restrictions on the use and development of land.[6] This regulation may be proactive as part of a community development plan, or in response to specific issues or concerns raised by community leaders. However, local control can be limited by state pre-emption laws, in which state governments explicitly or implicitly curtail the ability of local authorities to

0749-3797/09/$–see front matter
doi:10.1016/j.amepre.2009.09.028

regulate outlet expansion.[7] Thus, both state and local policies need to be considered when assessing factors that affect outlet density.

The WHO has published a review that identifies outlet density control as an effective method for reducing alcohol-related harms.[8] Similarly, in 1999, the Substance Abuse and Mental Health Services Administration's Center for Substance Abuse Prevention review concluded that there was a "medium" level of evidence supporting the use of outlet density control as a means of controlling alcohol-related harms.[9] In addition, several organizations have advocated the use of outlet density regulation for the reduction of alcohol consumption and alcohol-related harms. These include the European Union (in their 2000–2005 Alcohol Action Plan)[10] and the WHO Western Pacific Region.[11] The criteria used in the WHO report are not specified and may be expert opinion rather than systematic assessment of the characteristics of available studies. The SAMHSA review uses specified characteristics of included studies in drawing conclusions; however, the studies included are not up to date. In the present synthesis, 14 of the studies reviewed were published after 2000. Finally, a recent review by Livingston et al.[12] presents useful conceptual hypotheses and notes the importance of outlet "bunching"—which the team referred to as "clustering"—density at a more micro level.

Further, the present review assesses whether interventions limiting alcohol outlet density satisfy explicit criteria for intervention effectiveness of the *Guide to Community Preventive Services* (*Community Guide*), and assesses studies available as of November 2006. In addition, unlike any of the prior documents, the present review considers evidence from assessments of policies that are not explicitly considered density-related but that have direct effects on outlet density (i.e., privatization, liquor by the drink, and bans). If effective, policies limiting alcohol outlet density might address several national health objectives related to substance abuse prevention that are specified in *Healthy People 2010*.[13]

Guide to Community Preventive Services

The systematic review described in this report represents the work of CDC staff and collaborators on behalf of the independent, nonfederal Task Force on Community Preventive Services (Task Force). The Task Force is developing the *Community Guide* with the support of the USDHHS in collaboration with public and private partners. The book *The Guide to Community Preventive Services. What Works to Promote Health?* presents the background and the methods used in developing the *Community Guide*.[14]

Methods

The methods of the *Community Guide* review process[15,16] were used to assess whether the control of alcohol outlet density is an effective means of reducing excessive alcohol consumption and related harms. In brief, this process involves forming a systematic review development team (the team); developing a conceptual approach to 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; drawing conclusions about the body of evidence of effectiveness; and translating the evidence on intervention effectiveness into recommendations. Evidence is collected on 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 the applicability of evidence (i.e., the extent to which available effectiveness data might generalize to diverse population segments and settings), the economic impact of the intervention, and barriers to implementation. The results of this review process are then presented to the Task Force on Community Preventive Services (Task Force), an independent scientific review board from outside the federal government, which considers the evidence on intervention effectiveness and determines whether the evidence is sufficient to warrant a recommendation.[15]

Conceptual Approach and Analytic Framework

Outlet density is hypothesized to affect excessive alcohol consumption and related harms by changing physical access to alcohol (i.e., either increasing or decreasing proximity to alcohol retailers), thus changing the distance that drinkers need to travel to obtain alcohol or to return home after drinking. Increases in the density of on-premises outlets can also alter social aggregation, which may adversely affect those who are or who have been drinking excessively, leading to aggressive or violent behavior (Figure 1). With alcoholic beverages acquired in off-premises settings, the consumption more often occurs at the purchaser's home, and excessive consumption may be associated with domestic violence and suicidal behavior.

Decreases in off-premises or on-premises alcohol outlets, or both, are expected to decrease access to alcoholic beverages by increasing the distance to alcohol outlets, increasing alcohol prices, reducing exposure to on-premises alcohol marketing, and potentially by changing social norms around drinking, thereby decreasing excessive alcohol consumption and related harms. Decreases in outlet density are expected to decrease social aggregation in and around on- and off-premises alcohol outlets which, in turn, may decrease aggressive behavior potentially exacerbated by alcohol consumption.[17] Finally, decreased density increases distances traveled to and from alcohol outlets, thus increasing the potential for alcohol-related crashes. However, this potential harm could be mitigated by decreased alcohol consumption and hence decreased alcohol-impaired driving.[18,19] Thus, the expected effect of outlet density on motor-vehicle crashes may be mixed.[20]

The effect that density has on consumption and harms may be further influenced by at least seven characteristics

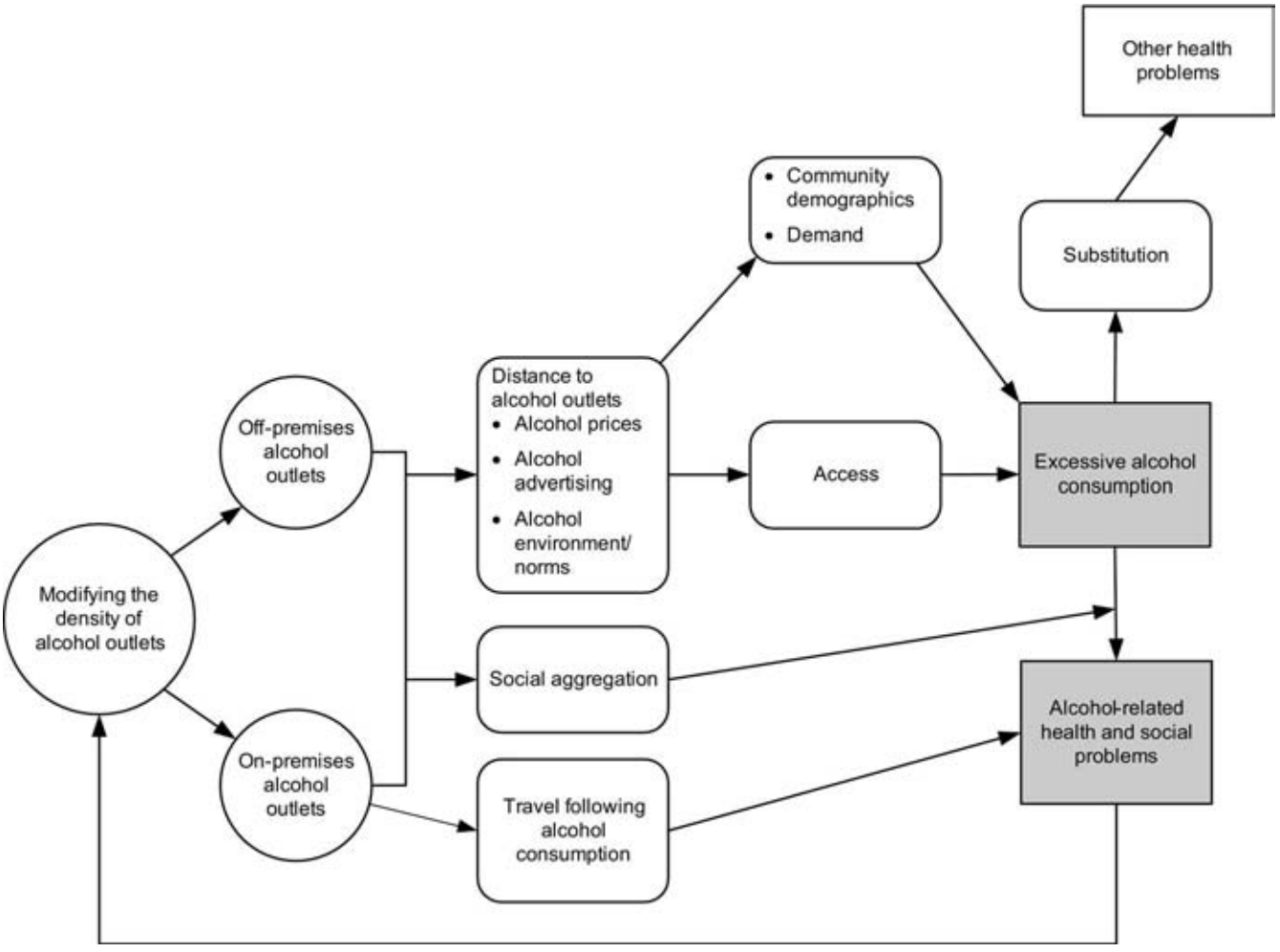


Figure 1. Analytic framework showing the hypothesized effects of changes in outlet density on excessive alcohol consumption and related harms

of retail alcohol outlets and the communities in which they are located: (1) outlet size (i.e., the physical size of the retail premises or the volume of its sales); (2) clustering (i.e., the level of aggregation of outlets within a given area); (3) location (i.e., the proximity of alcohol retail sites to places of concern, such as schools or places of worship); (4) neighboring environmental factors (e.g., demographics of the community and the degree of isolation of a community); (5) the size of the community (which may affect access to other retail sites); (6) the type and number of alcohol outlets (e.g., bar, restaurant, liquor store, grocery store) in a community may also influence whether and how outlet density affects drinking behavior[21]; and (7) alcohol outlets may be associated with illegal activities, such as drug abuse, which may also contribute to public health harms. As with other policies and regulations, the effects of regulations affecting outlet density may depend on the degree to which the policies are implemented and enforced.

There are several challenges to directly evaluating the effectiveness of local policies in changing outlet density on alcohol consumption and related harms. Direct studies of the effects of policies changing density on alcohol-related public health outcomes have not been conducted. Policy changes may occur in small communities in which documentation and data may be unavailable and where the number of retail alcohol outlets, alcohol-related outcomes, or both may be small; thereby it may be difficult to assess the relationship between outlet density and excessive alcohol consumption and related harms. Further, the effects of policy decisions on outlet density may be gradual. Other changes in alcohol control policies (e.g., enhanced enforcement of the minimum legal drinking age) may occur simultaneously, making it difficult to isolate the effect of changes in outlet density on drinking behavior.

The team used both primary and secondary scientific evidence to help address these challenges and to comprehensively assess the impact of changes in alcohol outlet density on excessive alcohol consumption. Primary evidence included studies comparing alcohol-related outcomes before and after a density-related change. In this category were (1) studies assessing the impact of privatizing alcohol sales—commonly associated with increases in density; (2) studies assessing the impact of bans on alcohol sales—associated with decreases in density; and (3) studies of other alcohol licensing policies that directly affect outlet density (e.g., the sale of liquor by the drink). Time–series studies (i.e., studies in which the association between changes in outlet density and alcohol-related outcomes is assessed over time) were also used to provide primary evidence

of intervention effectiveness, even when the cause of the observed change in outlet density was unknown. The team did not include studies of strikes in the production or distribution of alcoholic beverages or studies of interventions among college populations. Secondary evidence included cross-sectional studies, which do not allow the inference of causality.

Inclusion and Exclusion Criteria

To be included in this review, studies had to meet the following criteria: First, they had to evaluate changes in outlet density or policy changes that clearly resulted in changes in outlet density. Studies of policy changes (e.g., privatization or the legalization of liquor by the drink) had to provide evidence that there was a corresponding change in alcohol outlet density. Second, studies had to be conducted in high-income nations,[a,22] be primary research (rather than a review of other research), and be published in English. Third, studies had to report outcome measures indicative of excessive alcohol consumption or related harms. Direct measures that had the strongest association with excessive alcohol consumption included binge drinking, heavy drinking, liver cirrhosis mortality, alcohol-related medical admissions, and alcohol-related motor-vehicle crashes, particularly single-vehicle nighttime crashes, which are widely used to indicate motor-vehicle crashes due to drinking and driving.[23] Less direct measures included per capita ethanol consumption, which is a well-recognized proxy for the prevalence of heavy drinkers in a population[8,24]; unintentional injuries; suicide; and crime, such as homicide and aggravated assault. In most studies included in this review, consumption is measured by sales data; the team referred to this measure as "consumption" and note the exceptional study in which self-reported consumption is directly assessed. Fourth, studies had to be published in a peer-reviewed journal or in a government report. Reports not published or published by private organizations were not included.

Search for Evidence

The following databases were searched from inception up to November 2006 to identify studies assessing the impact of changes in alcohol outlet density and other review topics: EconLit, PsycINFO, Sociological Abstracts, MEDLINE, EMBASE, and EtOH (no longer available after 2003). The search yielded 6442 articles, books, and conference abstracts, of which 5645 were unique. After screening titles and abstracts, 251 papers and articles and 17 books were retrieved specifically related to outlet density; five articles could not be retrieved. After assessing quality of execution and design suitability (see below), 88 articles or books were included in the review. The actual number of studies that qualified for the review was less than this, however, because some studies were described in more than one report or publication.

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 review criteria (available at www.thecommunityguide.org/library/ajpm355_d.pdf) to assess the suitability of the study design and threats to validity. Uncertainties and disagreements between the reviewers were reconciled by the team. The classification of study design was based on *Community Guide* standards, and thus may differ from the classification reported in the original studies. Studies with greatest design suitability were those in which data on exposed and control populations were collected prospectively. Studies with moderate design suitability were those in which data were collected retrospectively or in which there were multiple pre- or post measurements but no concurrent comparison population. Studies with least-suitable designs were cross-sectional studies or those in which there was no comparison population and only a single pre- and post-intervention measurement. On the basis of the number of threats to validity (maximum: nine; e.g., poor measurement of exposure or outcome, lack of control of potential confounders, or high attrition) studies were characterized as having good (one or fewer threats to validity); fair (two to four threats); or limited (five or more threats) quality of execution. Studies with good or fair quality of execution, and any level of design suitability (greatest, moderate, or least), qualified for the body of evidence synthesized in the review.

The team summarized the results of cross-sectional studies based on whether drinking occurred on- or off-premises. However, some studies did not stratify their findings by outlet type and so were presented in a combined category. For each outcome and setting, the team summarized study findings by comparing the relative number of positive and negative findings. Finally, elasticities—summary effect measures showing the percentage change in an outcome per 1% change in an exposure (e.g., outlet density)—were calculated if the study provided sufficient information.

Other Harms and Benefits, Applicability, Barriers, and Economics

Harmful and beneficial outcomes not directly related to public health (e.g., vandalism or public nuisance) were noted if they were described in the studies reviewed or if the team regarded them as plausible. In addition, if an intervention was found to be effective, the team assessed barriers to implementation; the applicability of the intervention to other settings, populations, or circumstances; and the economic costs and benefits of the intervention.

Results

Intervention Effectiveness—Primary Evidence

Time–series studies of alcohol outlet density change. The team found ten studies[20,25–33] that directly evaluated the effect of changes in outlet density over time without identifying the causes for density changes. Of these, eight were "cross-sectional time–series" (i.e., panel)

[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.)

studies of greatest design suitability[20,25–29,31,33] and two were single-group time–series studies of moderate design suitability.[30,32] Eight of the studies were of good execution[25–31,33] and two were of fair execution.[20,32] Few took spatial lag (i.e., the likelihood that neighboring geographic units are not statistically independent) into account. Five studies assessed associations between changes in outlet density and population-level alcohol consumption,[25,26,28,31,33] and the remainder assessed specific alcohol-related harms.[20,27,29,30,32]

Consumption. All five studies that assessed the association between outlet density and population-level alcohol consumption found that they were positively associated; increased density was associated with increased consumption, and vice versa. Three studies examined the relationship between outlet density and the consumption of spirits in the U.S. The first study estimated that, from 1955 to 1980, for each additional outlet license per 1000 population, there was an increase of 0.027 gallons in per capita consumption of spirits ethanol ($p<0.01$).[28] The second study reported an elasticity of 0.14 ($p<0.01$) for outlet density and spirits for the period 1970–1975.[31] The third study examined the association of outlet density and the sale of spirits and wine in 38 states over a period of 18 years; the effects of consumption on density were separated out by use of two-stage least squares regression. The elasticity for spirits and wine was found to be 0.033 (NS) and 0.015 (NS), respectively.[26]

A study assessing trends from 1952 to 1992 in the United Kingdom[25] reported an elasticity of 2.43 ($p<0.05$) for off-premises density and beer consumption but no significant association for other beverages (except hard cider). Finally, a study[33] examining data from 1968 to 1986 in Canada reported a significant association between reductions in off-premises density and reductions in alcohol consumption. This study also found an association between changes in outlet density and cirrhosis mortality, which was mediated by changes in alcohol consumption. When the alcohol consumption variable was added to the analytic model, the coefficient for cirrhosis mortality was no longer significant.

Motor-vehicle crashes and other injury outcomes. Two studies by one author,[20,30] using the same methods and database in California, found mixed results when evaluating the association between on- and off-premises outlet density and fatal and nonfatal motor-vehicle crashes in small California cities (i.e., with total populations <50,000) during two different time periods and among different populations. The first study assessed the association between outlet density and crashes from 1981 through 1989 across all age groups. The author found a negative association between off-premises outlet density and both fatal and nonfatal crashes, and a positive association between on-premises outlets and both fatal and nonfatal crashes.[20] The second study assessed the association between outlet density and fatal and nonfatal crashes from 1981 through 1998 among people aged ≥60 years. This study reported a negative association for nonfatal crashes (elasticity: −0.69, $p<0.05$) and a positive association for fatal crashes (elasticity: 1.18, $p<0.05$).

Three studies[27,29,32] assessed the relationship between outlet density and suicide or interpersonal violence. A study of young people aged 10–24 years in the U.S. from 1976 through 1999 found positive associations between outlet density (on- and off-premises outlets combined) and suicides for most gender and age strata assessed, but only the findings for boys/men aged 15–19 years were significant (elasticities ranged from −0.03 to 0.10 for girls/women and from 0.05 to 0.12 for boys/men).[29]

The effect of changes in the density of on-premises outlets and violent crime was investigated in Norway from 1960 through 1995.[32] The researcher used autoregressive integrated moving average (ARIMA) modeling and found that each alcohol outlet was associated with 0.9 violent crimes investigated (by the police) per year. A supplementary analysis found that this association persisted even after controlling for amount of alcohol consumption, suggesting that the effect of increased density was independent of the effect of increased alcohol consumption ($p<0.03$). This suggests that the social aggregation of drinkers in and around alcohol outlets directly affects assaults, as indicated in Figure 1 (under "social problems").

Finally, a study of 581 California neighborhoods identified by ZIP code from 1996 through 2002[27] indicated that an increase in on- and off-premises outlet density was associated with an increase in hospitalizations for assault, but that this association varied for on-premises and off-premises locations, and among various types of on-premises locations (e.g., bar or restaurant) as well. The researchers used random-effects regression models, taking spatial lag into account, thus allowing for the lack of independence of neighborhoods in the association of outlets and alcohol-related harms. Within a given ZIP code, the elasticity for off-premises outlets and alcohol-related assaults on residents was 0.167 ($p<0.001$); for restaurants, it was −0.074 ($p<0.01$); and for bars, 0.064 ($p<0.001$). The elasticity for bars and assaults involving residents of neighboring ZIP codes was also significant (0.142, $p<0.001$); however, the elasticities for off-premises alcohol outlets and for restaurants relative to assaults involving residents of neighboring ZIP codes were not significant. Based on these results, the authors estimated that, on average, eliminating one bar per ZIP code in California would reduce the number of assaults requiring overnight hospitalization by 290 per year in the state.

Summary

Seven of nine time–series studies found positive associations between changes in outlet density and alcohol consumption and related harms, particularly interpersonal violence. However, two studies assessing the relationship between alcohol outlet density and motor-vehicle crashes in small California cities during two different time periods[20,30] had inconsistent findings for which no clear explanation was apparent. The studies reviewed also suggested that the association between outlet density and interpersonal violence may at least partially be due to social aggregation in and around alcohol outlets, and that the density of outlets in a given locale can also influence the probability of assaults involving residents of neighboring communities.

Privatization Studies

Alcohol privatization involves the elimination of government monopolies for off-premises alcohol sales to allow sales by privately owned enterprises. In the U.S. and Canada, privatization occurs at the state or provincial level; in many European nations, privatization may occur at a national level, currently guided by policies of the European Union. In the U.S., one alcoholic beverage may be privatized at a time; for example, wine might be privatized (i.e., subsequently for sale in commercial settings) while spirits may not be privatized, or may be privatized at a different time. Typically, privatization results not only in a substantial increase in the number of outlets where alcohol can be purchased but also in changes in alcohol price, days and hours of sale, and marketing.[21,34] This combination of events limits the ability to attribute subsequent changes in alcohol consumption and related harms to changes in outlet density alone. Nonetheless, because of the impact privatization generally has on outlet density, the team concluded that privatization studies were relevant for assessing the impact of changes in outlet density on excessive alcohol consumption and related harms.

The effects of privatization on the privatized beverages are assessed first, followed by an assessment of the effects of privatization on beverages other than those for which sales were privatized. If privatization affects consumption and related harms by means of increased outlet density, the consumption (and related harms) of the privatized beverage should increase, while consumption of other beverages might decline if usual drinkers of these other beverages now switch to the newly available privatized beverage. Comparing the association between alcohol consumption and alcohol-related harms associated with privatized and nonprivatized alcoholic beverages, respectively, provides a basis for assessing the impact of privatization on alcohol consumption and related harms while controlling for other factors that might be occurring simultaneously.

Following an analysis of the effects of privatization, this section then reviews the effects of remonopolization, that is, reversing privatization by reinstatement of government monopoly control over the retail sales of alcohol beverages. This policy change would be expected to have the opposite effects of privatization and result in lower alcohol outlet density.

Eleven events of privatization and one of remonopolization, analyzed in 17 studies and reported in 12 papers,[35–45] met the review inclusion criteria. The units of analysis were eight U.S. states (AL, ID, IA, ME, MT, NH, WA, WV); two Canadian provinces (Quebec and Alberta); and (in the sole study of remonopolization) Sweden. Several studies assessed overlapping privatization events. For example, two research teams assessed the privatization of wine and then spirits in Iowa,[34,38,39,45] and two researchers assessed early phases of the privatization of wine in Quebec, while one of these researchers also assessed the later phases, with each phase counted as a separate privatization event.[36,46] In addition, several papers assessed the effects of privatization in more than one state and provided separate effect estimates for the privatization in each state; for purposes of this review, each state-level assessment was treated as a separate study. Finally, a single state or province could privatize different beverages at different times, resulting in separate privatization events. Altogether, the events assessed in these studies occurred between 1978 and 1993. In all areas assessed, the number of outlets increased dramatically following privatization. The studies used ARIMA time–series study design; all except two studies[36,46] reported results for comparison populations.

All studies used alcohol sales data as a measure of population-level alcohol consumption. One study also assessed fatal motor-vehicle crashes (MVCs),[42] another study[34] also evaluated single-vehicle nighttime crashes and liver cirrhosis. The single study of remonopolization[40] assessed hospitalizations for alcoholism, alcohol intoxication, and alcohol psychosis combined, alcohol intoxication alone, assaults, suicides, falls, and MVCs.[40] Fourteen studies (in seven papers)[35,38,39,42–44,46] were of greatest design suitability; three studies (in two papers)[37,40] were of moderate design suitability. All studies were of fair execution.

Effects of Privatization on Privatized Beverages

Seventeen studies[35–44] assessed the effects of privatization on the sale of at least one of four beverage types (wine, spirits, full-strength beer, and medium-strength beer) in ten settings. The median relative increase in alcohol sales subsequent to privatization was 42.0%, with an interquartile interval of 0.7% to 136.7%. That is, among the studies reviewed, compared with consumption prior to privatization, the median effect was

an increase of 42.0% in consumption of the privatized alcoholic beverage. Studies of three events of privatization, two in Iowa and one in Alberta, yielded inconsistent findings, which merit further description.

In Iowa, wine was privatized in 1985, and spirits in 1987. Wagenaar and Holder[35,43] reported that wine consumption increased 93.0% (95% CI=69.3, 120.2) from baseline to 44 months after privatization of retail wine sales. Following the subsequent privatization of retail spirits sales in Iowa 2 years later, these researchers[35,43] reported a 9.5% (95% CI=3.5, 15.9) increase in spirits consumption; they also found no evidence that privatization affected cross-border alcohol purchasing.[35,43] In contrast, Mulford and Fitzgerald[39] found that wine privatization in Iowa was associated with a nonsignificant increase of only 0.5% (95% CI= −13.2, 16.4) in wine sales, and that spirits privatization was associated with a nonsignificant increase of 0.7% (95% CI= −4.3, 6.0) in spirits sales. Differences between the findings of these research groups may be due to differences in time periods assessed, modeling variables and procedures, beverage types included in the assessment (e.g., Mulford and Fitzgerald exclude wine coolers that were not affected by the policy change and Wagenaar and Holder do not), use of a control population, and outcome measurement. Fitzgerald and Mulford[34] also report small unadjusted rate decreases in single-vehicle nighttime crashes (−1.6%) and alcoholic cirrhosis mortality (−5.5%) associated with the privatization of wine and spirits in Iowa.

A study in Alberta, Canada, estimated that gradual privatization over a period of 20 years resulted in an increase in spirits consumption of 12.7% (95% CI=2.2, 24.4) and no change in either wine or beer consumption.[42] Although the process of privatization occurred over an extended period, the major events of privatization occurred essentially at the same time (in 1992); thus, considered in aggregate, privatizing spirits in Alberta increased total alcohol sales by 5.1% (95% CI= −2.8, 13.7) over this 20-year period. Despite the increased alcohol sales, the authors reported that there was an estimated 11.3% (95% CI= −33.8, 19.0) decrease in traffic fatalities. However, neither the increase in total alcohol sales nor the decrease in traffic fatalities was significant.

Effects of Privatization on Beverages Not Subject to Privatization

Five publications[37,38,43,44,47] assessed the effects of privatization in eight settings on the concomitant sales of alcoholic beverages that were not privatized during the same period. Overall, these studies reported that there was a minimal decline: a median of 2.1% (interquartile interveral [IQI]: −4.8% to 2.7%) in the sales on nonprivatized beverages.

Effects of Remonopolization on Alcohol-Related Outcomes

A single before-and-after study[40] evaluated the effects of remonopolization of sales of medium-strength beer in Sweden. This study compared the association between the number of retail alcohol outlets and the occurrence of six different alcohol-related outcomes during a 51-month period following the remonopolization of medium-strength beer, with that for a similar period prior to remonopolization. Among young people aged 10–19 years, alcoholism, alcohol intoxication, and alcohol psychosis (which were considered in combination) decreased by 20% ($p<0.05$) following remonopolization. These outcomes also decreased by >5% among people aged ≥40 years, although the change was not significant ($p>0.05$). Hospitalizations for acute alcohol intoxication also decreased between 3.5% and 14.7% ($p>0.05$); suicides decreased by 1.7% to 11.8% ($p>0.05$); and falls decreased by 3.6% to 4.9% ($p>0.05$) following remonopolization, although none of these changes were significant either. Motor-vehicle crashes (MVCs) significantly decreased by 14% ($p<0.05$) in all age categories except one (those aged 20–39 years). Other nonsignificant changes include assaults, which decreased by 1.4% among those aged 20–39 years, but increased by 6.9% to 14.8% ($p>0.05$) in the other age groups: 10–19, 40–59, ≥60 years. The authors did not provide any explanation for this seemingly inconsistent finding.

Summary

These studies indicate that privatization increases the sales of privatized beverages but has little effect on the sales of nonprivatized alcoholic beverages. The one study that evaluated the reintroduction of government monopoly control of sale of an alcoholic beverage (medium-strength beer) found that remonopolization led to a significant decrease in motor-vehicle crashes for most age groups and a significant decrease among youth for several, but not all, alcohol-related harms.

Studies of Alcohol Bans

The team found seven studies[18,41,48–52] that examined the effects of bans on local on- or off-premises alcohol sales or consumption (i.e., "dry" towns, counties, or reservations). Five studies examined the effects of bans in American Indian and Native settings in Alaska,[49,50,53] northern Canada,[52] and the southwestern U.S.[51] Two studies assessed the effects of bans in nontribal areas of the U.S. and Canada.[18,41] Two studies were of greatest design suitability[18,41]; two of moderate design suitability[50,51]; and three of least suitable design.[49,52,53] All were of fair execution. The studies examined events that occurred from 1970

through 1996. Two additional studies modeled the association of multiple policies, including local policies of dry counties, with spirits consumption[28] and with juvenile suicide.[29] Both of these studies were of greatest design suitability and good execution, and the team considered them comparable to studies of bans and as primary evidence.

An additional cross-sectional study of bans[54] was not used as primary evidence of effectiveness, but provided insights into the effect that alcohol availability in areas surrounding dry communities (e.g., outside Indian reservations) has on the occurrence of alcohol-related harms among residents of the dry communities.

Effects of Alcohol Bans in Isolated Communities

All of the studies that evaluated the effect of bans in isolated northern communities found substantial reductions in alcohol-related harms with the exception of suicide.[18,41,49,51–59] In the communities that instituted bans, rates of harm indicated by alcohol-related medical visits were reduced by 9.0% for injury deaths to 82% for alcohol-related medical visits (CIs not calculable). One of these studies[50] found that the effects were reversed when the ban was lifted, and found similar benefits when the ban was then reimposed (Figure 2).[50] Two of these studies suggest that bans on alcohol sales in isolated communities led residents to decrease their use of other intoxicants. In Barrow, Alaska, medical visits for use of isopropyl alcohol declined during ban periods.[50]

An additional study qualitatively evaluated a Canadian Inuit community[52] that overwhelmingly voted to ban alcohol in 1978. Although comparative data are not available from this study (and the study thus does not meet review inclusion criteria), it is notable that during the 3 years following the implementation of this prohibition there were only five arrests for the illegal possession of alcohol and, of these, four were associated with a single incident. The reported reduction in alcohol consumption in general and among youth in particular was linked with several societal benefits, including improved mental and physical health among community members, and a reduction in conflicts within the community. The ban on alcohol sales was associated with a reduction in the use of other substances of abuse (e.g., inhalants) by youth.

Effects of Alcohol Bans in Less-Isolated Communities

Studies assessing the impact of bans (particularly bans on on-premises sales) in less-isolated communities have produced mixed results. Some studies have found that bans are associated with increases in alcohol-related harms, including motor-vehicle crashes[18,46] and alcohol-related arrests.[51] However, two studies[28,29] found that states that had a larger proportion of their population living in dry counties had less alcohol consumption and related harms than states that had a smaller proportion of their population living in dry counties. One study[28] found that living in dry counties was associated with lower rates of spirits consumption ($p<0.01$). The other study found small, nonsignificant associations with male suicide (elasticities of −0.002 to −0.066) and female suicide (elasticities of −0.021 to −0.038).[29]

A cross-sectional study of injury deaths in New Mexico[54] highlights the potential harms associated with alcohol sales bans in areas (in this case reservations, 80% of which are dry) that are adjacent to other areas where alcohol is readily available. This study found that in these settings, although the relative risk (RR) of total injury deaths was greater for American Indians than for whites (RR=3.1; 95% CI=2.6, 3.6), the relative risk was greatest for deaths involving pedestrians struck by vehicles (RR=7.5; 95% CI=5.3, 10.6) and for hypothermia (i.e., freezing to death; RR=30.5; 95% CI=17.7, 48.7). Furthermore, American Indians in New Mexico who died of these causes were likely to

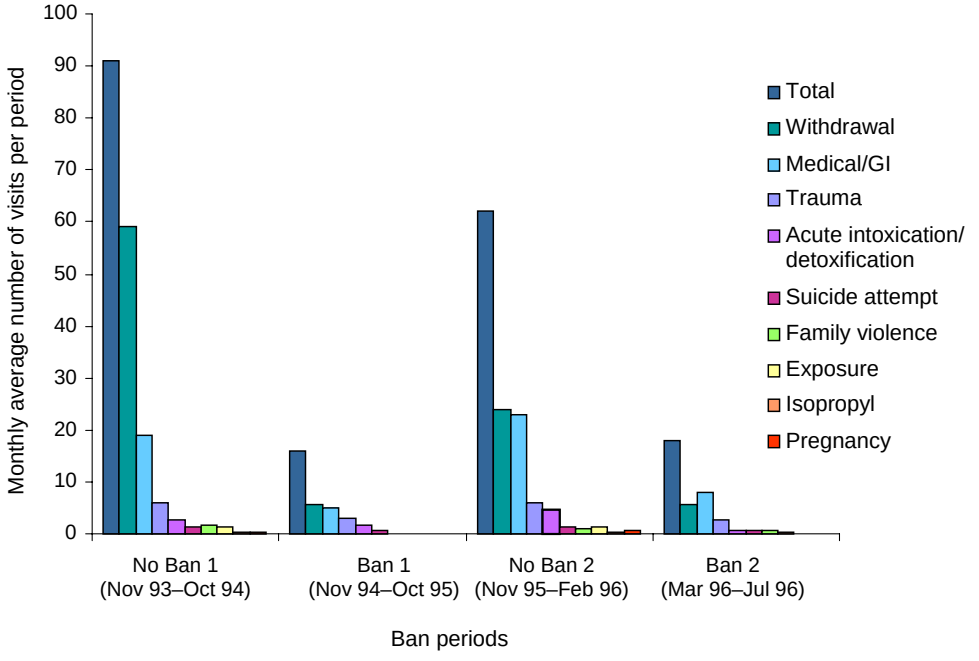


Figure 2. Alcohol-related outpatient visits associated with changes in alcohol ban policy, Barrow, Alaska, 1993–1996[50]

have elevated blood alcohol levels (an average of 0.24 g/dL and 0.18 g/dL for pedestrian deaths and hypothermia, respectively). A disproportionate number (67%) of these deaths occurred in counties bordering reservations, despite the fact that most American Indians live on reservations. Although the design of this study does not allow causal inference regarding the effect of bans, these findings suggest that travel between dry reservations and adjacent areas where alcohol is readily available may increase the risk of death from these external causes among those traveling off-reservation to purchase alcohol.

Summary

The effectiveness of bans in reducing alcohol-related harms appears to be highly dependent on the availability of alcohol in the surrounding area. In isolated communities, bans can substantially reduce alcohol-related harms. However, where alcohol is available in areas nearby those with bans, travel between these areas may lead to serious harms.

Studies of Licensing-Policy Changes Affecting Outlet Density

The team identified four studies of national or local licensing-policy changes that resulted in increased outlet density. The studies were conducted in Iceland,[60] Finland,[47] New Zealand,[61] and North Carolina.[62] The policy changes assessed occurred between 1969 and 1990. The North Carolina study was of greatest design suitability and good execution. The other three studies were of moderate design suitability and good execution.[47,60,61] These studies examined various indices of alcohol consumption; the North Carolina study also assessed effects on alcohol-related motor-vehicle crashes. Another study assessed the effect of a change in national policy controlling the sale of table wine in New Zealand.

Effects on Excessive Alcohol Consumption and Related Harms

The only U.S. study that met criteria for this category of interventions evaluated the decision by several North Carolina counties to allow on-premises sale of spirits (i.e., "liquor by the drink" [LBD]), replacing the previous option of "brown-bagging,"[62] in which patrons of an establishment bring their own alcoholic beverage (in a bag) and the establishment supplies other items (e.g., a drink glass, ice, water). Of the 100 counties in North Carolina, three approved liquor by the drink in November 1978 and eight approved it in January 1979. The policy change was followed by the opening of many bars and lounges adjacent to restaurants. Interrupted time–series models indicated that, relative to counties that did not change their policies, sales of spirits increased in LBD counties by 8.2% ($p<0.05$) among the first group of counties to adopt the new policy, and by 4.3% ($p<0.05$) among the second group. Nighttime single-vehicle crashes among men of legal drinking age also increased in both early- and late-adopting counties by 18.5% ($p<0.01$) and 15.7% ($p<0.01$), respectively. However, there were no significant changes in rates of nighttime single-vehicle crashes among boys/men aged <21 years, who were not permitted to drink spirits and were thus not (legally) affected by the policy change.

In Finland, the enactment in 1969 of a policy allowing the sale of medium-strength beer resulted in a 22% increase in the number of monopoly alcohol outlets and a 46% increase in restaurant liquor licenses, and permitted 17,400 grocery stores to sell medium-strength beer. During the year following these changes, overall alcohol sales in Finland increased by 46%. Of the increase, 86% was attributed by the researchers to the increased availability of beer. Overall alcohol consumption increased by 56%, with the greatest volume increases among those drinking more than a half liter of pure alcohol per year (1/2 liter of pure alcohol is equivalent to 1/3 gallon of 80-proof liquor). However, alcohol consumption increased significantly among all adults at all levels of alcohol consumption in Finland subsequent to this policy change, regardless of their baseline pattern of consumption, including those who had previously reported that they had not consumed alcohol during the past year.

In Iceland,[60] a policy change in 1989 resulted in an expansion in off-premises monopoly outlets and commercial on-premises outlets in Reykjavik and in rural areas. Over the subsequent 4-year period, consumption increased by 43% among men who drank more than 350 centiliters of alcohol per year at baseline, but changed minimally among women and men who drank at lower levels.

In New Zealand,[61] a policy change in 1989 allowed the sale of table wine in grocery stores, resulting in an increase of approximately 25% in the number of wine outlets in the country over a 2-year period. This resulted in a 17% (95% CI=9.8%, 24.9%) increase in wine sales during this time, but in no change in the sales of other alcoholic beverages. This indicates that there was an overall increase in alcohol consumption in New Zealand subsequent to this policy change, and that wine, the privatized beverage, was not being substituted for other nonprivatized alcoholic beverages.

Summary

These studies consistently indicated that more permissive licensing procedures increased the number of on- and off-premises alcohol outlets, which in turn led to increases in alcohol consumption. Two of these studies specifically reported increases in alcohol consumption among heavy drinkers, and one study reported an increase in drinking among survey subjects who reported not drinking during a specified period at the

baseline assessment. The single study that evaluated alcohol-related harms (alcohol-related motor-vehicle crashes) found that they increased substantially after allowing the sale of liquor by the drink.

Intervention Effectiveness—Secondary Evidence

Although the primary evidence just reviewed is heterogeneous in topic and design and does not allow summary tabular presentation, the secondary evidence presented below is based on consistent statistical procedures and readily allows a summary table.

Cross-Sectional Studies

Findings from studies of on- and off-premises outlets combined. The 28 cross-sectional studies[19,55–57,63–86] that assessed the association of outlet density (on-premise and off-premise, not distinguished) assessed 47 alcohol-related outcomes. Of these outcomes, 41 (87.2%) found a positive association, that is, as density increased, so did consumption and alcohol-related harms, and vice versa (Table 1, A). Positive associations were found for consumption-related outcomes (e.g., per capita alcohol consumption); violence and injury outcomes; and several medical conditions (e.g., liver disease). The mean elasticities ranged from 0.045 for crime to 0.421 for motor-vehicle crashes.

Findings from studies of on-premises outlets. The 23 studies[23,58,78,79,87–105] that assessed the association of outlet density and alcohol-related outcomes in on-premises outlets reported on 25 outcomes. Of these, 21 (84.0%) indicated a positive association (Table 1, B). Positive associations were also found for consumption-related outcomes, several forms of violence and injury outcomes related to alcohol consumption, and one medical condition. Mean study elasticities could be estimated for most outcome types, and values ranged from 0.021 for child abuse to 0.250 for population consumption.

Findings from studies of off-premises outlets. The 23 studies[58,79,89–92,94–99,101–111] that assessed the association of outlet density and alcohol-related outcomes in off-premises outlets reported on 24 outcomes. Of these, 18 (75.0%) also indicated a positive association (Table 1, C). Positive associations were found for consumption-related outcomes, several forms of violence and injury outcomes related to alcohol consumption, and one medical condition. Mean study elasticities could be estimated for most outcome types and values ranged from −0.15 for injury to 2.46 for population consumption. Mean elasticity was also high (0.483) for violent crime.

Summary

Cross-sectional studies generally show consistent positive associations between alcohol outlet density and excessive alcohol consumption and related harms, with the possible exception of injuries, for which the findings were less consistent. The largest effect sizes were for studies relating outlet density to population consumption and violent crime.

Table 1. Cross-sectional studies, outcomes by setting type

Outcomes	# of studies	% positive	M elasticity
A. ON- AND OFF-PREMISES AGGREGATED			
Consumption			
Population consumption	7	85.7	0.27
Binge drinking	5	80.0	
Underage drinking	2	100.0	
Violence and injury			
Violent crime	15	93.3	0.32
Injury	3	100.0	0.23
Motor-vehicle crashes	6	50.0	0.42
Drunk driving	1	100.0	
Crime	2	100.0	0.04
Medical conditions			
Alcohol medical visits	1	100.0	
Alcoholism	1	100.0	
Liver disease	4	100.0	
Total all premises	47	87.2	
B. ON-PREMISES			
Consumption			
Population consumption	3	33.3	0.25
Binge drinking	1	100.0	
Violence and injury			
Violent crime	4	100.0	0.12
Injury	3	100.0	0.14
Motor-vehicle crashes	6	66.7	0.05
Drunk driving	2	100.0	
Crime	1	100.0	
Child abuse	2	100.0	0.02
Medical conditions			
Liver disease	3	100.0	0.06
Total on-premises	25	84.0	
C. OFF-PREMISES			
Consumption			
Population consumption	2	100.0	2.46
Binge drinking	1	100.0	
Violence and injury			
Violent crime	6	100.0	0.48
Injury	3	66.7	−0.15
Motor-vehicle crashes	5	80.0	0.10
Drunk driving	2	50.0	
Crime	1	100.0	
Child abuse	2	100.0	0.01
Medical conditions			
Liver disease	2	50.0	−0.05
Total off-premises	24	76.9	

Summary of the Body of Scientific Evidence on Alcohol Outlet Density and Excessive Drinking and Related Harms

Using a variety of different study methods, study populations, and alcohol measures, most of the studies included in this review reported that greater outlet

density is associated with increased alcohol consumption and related harms, including medical harms, injuries, crime, and violence. This convergent evidence comes both from studies that directly evaluated outlet density (or changes in outlet density) and those that evaluated the effects of policy changes that had a substantial impact on outlet density, including studies of privatization, remonopolization, bans on alcohol sales and the removal of bans, and changes in density from known policy interventions and from unknown causes. Studies assessing the relationship between alcohol outlet density and motor-vehicle crashes produced mixed results.[18,20,62,112]

Other Benefits and Harms

Communities commonly seek limits on alcohol outlet density, either through licensing or zoning, for purposes that may not be directly related to public health (e.g., the reduction of public nuisance, loitering, vandalism, and prostitution).[7,113] Although the team did not specifically search for studies that assessed these outcomes, some of the studies the team reviewed suggested that there may be an association between outlet density and these outcomes as well. For example, a study from New South Wales, Australia, reported an association between outlet density and "neighborhood problems with drunkenness" but did not find a significant association with property damage.[114] There was evidence of one potential harm of decreased outlet density (i.e., an increase in fatal single-vehicle nighttime vehicle crashes) presumably associated with an increase in driving in response to greater distances between alcohol outlets.[19]

Applicability

Evidence of the association of outlet density and alcohol consumption and related harms derives from studies conducted primarily in North American and in Scandinavian countries. One study[27] indicated that the impact of changes in outlet density may be affected by demographic characteristics (e.g., gender distribution) of the population; in this case, the association of outlet density with assaults requiring hospitalization was stronger where there was a greater proportion of boys/men in the population. Most of the studies reviewed assessed the effects of increased outlet density, which is a consequence of the general trend toward liberalization of alcohol policies associated with outlet density. Few data were found from which to draw inferences about regulations that control or reduce outlet density.

Studies of bans on alcohol sales, conducted primarily among American Indian and Alaska Native populations, consistently report a reduction in excessive consumption and related harms following the implementation of a ban on alcohol sales, possession, or both, provided the area affected by the ban was not surrounded by other sources of alcoholic beverages.

Barriers

Reductions in outlet density, with resultant reductions in consumption, are likely to have substantial commercial and fiscal consequences, and thus may be opposed by commercial interests in the manufacture, distribution, and sale of alcoholic beverages. In keeping with its commercial interests, the alcoholic beverage industry has tended to support policies that facilitate outlet expansion.[115]

State pre-emption laws (i.e., laws that prevent implementation and enforcement of local restrictions) can also undermine efforts by local governments to regulate alcohol outlet density.[7] 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*.[13] However, there is no similar objective in *Healthy People 2010* related to the sale of alcoholic beverages.

Economic Evaluation

The team's systematic economic review did not identify any study that examined the costs and benefits of limiting alcohol outlet density. Although there has been speculation that reducing the number of alcohol outlets may result in a loss of revenue to state and local governments owing to a loss of licensing fees and alcohol tax revenues, the team found no studies that have documented this speculation. In addition, there may be economic gains resulting from revenue generation from merchants and consumers who would otherwise avoid areas known to have a high alcohol outlet density; however, the team found no studies about this topic. Moreover, in 2006, alcoholic beverage licenses accounted for only $406 million (0.9%) of the $45 billion that state governments received from all licensing fees, and alcohol taxes accounted for only 0.7% of all taxes ($4.9 billion of $706 billion) collected by state governments (www.census.gov//govs/statetax/0600usstax.html).

Even in the absence of published data on program implementation costs and other costs related to this intervention, it should be expected that the cost of restricting access to alcohol by limiting the number of alcohol outlets is likely to be small relative to the societal cost of excessive alcohol consumption in the U.S. For example, in 1998, the most recent year for which data are available, the societal cost of excessive alcohol consumption in the U.S. was $185 billion, including, among other costs, approximately $87 billion in lost productivity due to morbidity, $36 billion in lost future earnings due to premature deaths, $19 billion in medical care costs, $10 billion in lost earnings due to crime, $6 billion in costs to the criminal justice

system, and $16 billion in property damage related to motor-vehicle crashes.[4] Moreover, each state alcohol enforcement agent is responsible for monitoring an average of 268 licensed establishments[116]; thus, reducing the number of retail alcohol outlets might reduce their enforcement responsibilities. In summary, no existing study examines the economic costs and benefits of limiting alcohol outlet density.

Research Gaps

Although the scientific evidence reviewed indicates that the regulation of alcohol outlet density can be an effective means of controlling excessive alcohol consumption and related harms, it would be useful to conduct additional research to further assess this relationship:

- There are few if any studies evaluating how local decisions are made regarding policies affecting alcoholic beverage outlet density or the consequences of such policy changes. Such case studies may be difficult to conduct, but they could provide important insights to guide policy decisions regarding alcohol outlet density in other communities.
- The majority of outlet density research explores the impact of increasing alcohol outlet density on alcohol-related outcomes; there is a lack of research on the impact of reducing outlet density. This might be done by observing the impact of temporal changes in outlet density on excessive alcohol consumption and related harms.
- The association of on- and off-premises alcoholic beverage outlets with illegal activities such as prostitution and drug abuse should be examined. In themselves, these may have adverse public health and other outcomes; in addition, they may confound the apparent association of alcohol outlets with these outcomes.
- Relatively little is known about the impact of density changes relative to baseline density levels. Some authors (e.g., Mann[117]) have proposed that the association between outlet density and alcohol consumption follows a demand curve, such that when density is relatively low, increases in density may be expected to have large effects on consumption, and when density is relatively high, increases in density should be expected to have smaller effects.[21,117] Thus, it would be useful to assess this hypothesis empirically using econometric methods, with different kinds of alcohol-related outcomes. Such information would allow communities at different alcohol outlet density "levels" to project the possible benefits of reducing density by specific amounts or the potential harms of increasing density.
- For public health practitioners, legislators, and others attempting to control alcohol outlet density to reduce alcohol-related harms, it would be useful to catalog approaches to regulation beyond licensing and zoning that may have an effect on outlet density (e.g., traffic or parking regulations that, in effect, control the number of driving patrons who may patronize an alcohol outlet).
- A primary rationale for limiting alcohol outlet density is to improve public health and safety. Furthermore, the economic efficiency of limiting outlet density is difficult to assess without data on the economic impact of this intervention. To remedy this, future studies on the impact of changes in alcohol outlet density should assess both health and economic outcomes, so that the economic impact of this intervention can be assessed empirically.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

The authors are grateful for the contributions of Ralph Hingson, ScD, MPH (National Institute of Alcohol Abuse and Alcoholism), and Steve Wing (Substance Abuse and Mental Health Services).

No financial disclosures were reported by the authors of this paper.

References

1. CDC. Alcohol-attributable deaths and years of potential life lost—United States, 2001. MMWR Morb Mortal Wkly Rep 2004;53(37):866–70.
2. 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.
3. National Center for Health Statistics. Health, United States, 2005 with chartbook on trends in the health of America. Hyattsville MD: National Center for Health Statistics, 2005. USDHHS No. 2005–1232.
4. Harwood H. Updating estimates of the economic costs of alcohol abuse in the United States: estimates, update methods, and data. Rockville MD: National Institute on alcohol Abuse and Alcoholism, 2000. NIH Publication No. 98-4327.
5. National Alcoholic Beverage Control Association. NABCA survey book. Alexandria VA: NABCA, 2005.
6. Ashe M, Jernigan D, Kline R, Galaz R. Land use planning and the control of alcohol, tobacco, firearms, and fast food restaurants. Am J Public Health 2003;93(9):1404–8.
7. Mosher J. Alcohol issues policy briefing paper: the perils of preemption. Chicago IL: American Medical Association, 2001.
8. Babor TF, Caetano R, Casswell S, et al. Alcohol: no ordinary commodity—research and public policy. Oxford: Oxford University Press, 2003.
9. Grover PL, Bozzo R. Preventing problems related to alcohol availability: environmental approaches. Rockville MD: SAMHSA/CSAP, 1999. DHHS Publication No. (SMA) 99-3298.
10. WHO. European Alcohol Action Plan, 2000–2005. Copenhagen: WHO, 2000.
11. WHO. Draft regional strategy to reduce alcohol-related harm. Auckland, New Zealand: WHO, 2006.
12. Livingston M, Chikritzhs T, Room R. Changing the density of alcohol outlets to reduce alcohol-related problems. Drug Alcohol Rev 2007;26:557–66.
13. USDHHS. Healthy People 2010. Washington DC: USDHHS, 2001.
14. Task Force on Community Preventive Services. The Guide to Community Preventive Services. What works to promote health? In: Zaza S, Briss PA, Harris KW, eds. New York: Oxford University Press, 2005. www.thecommunityguide.org.
15. 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.

16. Zaza S, Wright-De Aguero LK, Briss PA, et al. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Am J Prev Med 2000;18(1):44–74.
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. Vol. XIII. Alcohol and violence. New York: Plenum Press, 1997:245–82.
18. Baughman R, Conlin M, Dickert-Conlin S, Pepper J. Slippery when wet: the effects of local alcohol access laws on highway safety. J Health Econ 2001;20(6):1089–96.
19. Gruenewald PJ, Treno AJ, Nephew TM, Ponicki WR. Routine activities and alcohol use: constraints on outlet utilization. Alcohol Clin Exp Res 1995;19(1):44–53.
20. McCarthy P. Alcohol-related crashes and alcohol availability in grass-roots communities. Appl Econ 2003;35(11):1331–8.
21. Her M, Giesbrecht N, Room R, Rehm J. Privatizing alcohol sales and alcohol consumption: evidence and implications. Addiction 1999;94(8): 1125–39.
22. World Bank. World Development Indicators 2006. devdata.worldbank.org/wdi2006/contents/cover.htm.
23. 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.
24. Cook PJ, Skog OJ. Alcool, alcoolisme, alcoolisation: comment. Alcohol Health Res World 1995;19(1):30–1.
25. Blake D, Nied A. The demand for alcohol in the United Kingdom. Appl Econ 1997;29:1655–72.
26. Gruenewald PJ, Ponicki WR, Holder HD. The relationship of outlet densities to alcohol consumption: a time series cross-sectional analysis. Alcohol Clin Exp Res 1993;17(1):38–47.
27. Gruenewald PJ, Remer L. Changes in outlet densities affect violence rates. Alcohol Clin Exp Res 2006;30(7):1184–93.
28. Hoadley JF, Fuchs BC, Holder HD. The effect of alcohol beverage restrictions on consumption: a 25-year longitudinal analysis. Am J Drug Alcohol Abuse 1984;10(3):375–401.
29. Markowitz S, Chatterji P, Kaestner R. Estimating the impact of alcohol policies on youth suicides. J Ment Health Policy Econ 2003;6(1):37–46.
30. McCarthy P. Alcohol, public policy, and highway crashes: a time-series analysis of older-driver safety. Transp Econ Policy 2005;39(1):109–25.
31. McCornac DC, Filante RW. The demand for distilled spirits: an empirical investigation. J Stud Alcohol 1984;45(2):176–8.
32. Norstrom T. Outlet density and criminal violence in Norway, 1960–1995. J Stud Alcohol 2000;61(6):907–11.
33. Xie X, Mann RE, Smart RG. The direct and indirect relationships between alcohol prevention measures and alcoholic liver cirrhosis mortality. J Stud Alcohol 2000;61(4):499–506.
34. Fitzgerald JL, Mulford HA. Consequences of increasing alcohol availability: the Iowa experience revisited. Br J Addict 1992;87(2):267–74.
35. Holder HD, Wagenaar AC. Effects of the elimination of a state monopoly on distilled spirits' retail sales: a time-series analysis of Iowa. Br J Addict 1990;85:1615–25.
36. Trolldal B. The privatization of wine sales in Quebec in 1978 and 1983 to 1984. Alcohol Clin Exp Res 2005;29(3):410–6.
37. MacDonald S. The impact of increased availability of wine in grocery stores on consumption: four case histories. Br J Addict 1986;81:381–7.
38. Mulford HA, Fitzgerald JL. Consequences of increasing off-premise wine outlets in Iowa. Br J Addict 1988;83(11):1271–9.
39. Mulford HA, Ledolter J, Fitzgerald JL. Alcohol availability and consumption: Iowa sales data revisited. J Stud Alcohol 1992;53(5):487–94.
40. Ramstedt M. The repeal of medium-strength beer in grocery stores in Sweden—the impact on alcohol-related hospitalizations in different age groups. Finland: Nordic Council for Alcohol and Drug Research (NAD), 2002. No. 42.
41. Smart RG, Docherty D. Effects of the introduction of on premise drinking on alcohol-related accidents and impaired driving. J Stud Alcohol 1976;37:683–6.
42. Trolldal B. An investigation of the effect of privatization of retail sales of alcohol on consumption and traffic accidents in Alberta, Canada. Addiction 2005;100(5):662–71.
43. Wagenaar AC, Holder HD. A change from public to private sale of wine: results from natural experiments in Iowa and West Virginia. J Stud Alcohol 1991;52(2):162–73.
44. Wagenaar AC, Holder HD. Changes in alcohol consumption resulting from the elimination of retail wine monopolies: results from five U.S. states. J Stud Alcohol 1995;56(5):566–72.
45. Fitzgerald JL, Mulford HA. Privatization, price and cross-border liquor purchases. J Stud Alcohol 1993;54(4):462–4.
46. Smart RG. The impact on consumption of selling wine in grocery stores. Alcohol Alcohol 1986;21(3):233–6.
47. Makela P. Whose drinking does the liberalization of alcohol policy increase? Change in alcohol consumption by the initial level in the Finnish panel survey in 1968 and 1969. Addiction 2002;97(6):701–6.
48. Berman MHT. Alcohol control by referendum in Northern native communities: the Alaska local option law. Arctic 2001;54(1):77–83.
49. Bowerman RJ. The effect of a community supported alcohol ban on prenatal alcohol and other substance abuse. Am J Public Health 1997;87(8):1378–9.
50. Chiu AY, Perez PE, Parker RN. Impact of banning alcohol on outpatient visits in Barrow, Alaska. JAMA 1997;278(21):1775–7.
51. May P. Arrests, alcohol and alcohol legalization among an American Indian tribe. Plains Anthropol 1975;20(68):129–34.
52. O'Neil JD. Community control over health problems: alcohol prohibition in a Canadian Inuit village. Int J Circumpolar Health 1984;84:340–3.
53. Berman M, Hull T, May P. Alcohol control and injury death in Alaska native communities: wet, damp and dry under Alaska's local option law. J Stud Alcohol 2000;61(2):311–9.
54. Gallaher MM, Fleming DW, Berger LR, Sewell CM. Pedestrian and hypothermia deaths among Native Americans in New Mexico: between bar and home. JAMA 1992;267(10):1345–8.
55. Britt H, Carlin BP, Toomey TL, Wagenaar AC. Neighborhood-level spatial analysis of the relationship between alcohol outlet density and criminal violence. Environ Ecol Stat 2005;12:411–26.
56. Escobedo LG, Ortiz M. The relationship between liquor outlet density and injury and violence in New Mexico. Accid Anal Prev 2002;34(5):689–94.
57. Parker DA, Wolz MW, Harford TC. The prevention of alcoholism: an empirical report on the effects of outlet availability. Alcohol Clin Exp Res 1978;2(4):339–43.
58. Treno AJ, Gruenewald PJ, Johnson FW. Alcohol availability and injury: the role of local outlet densities. Alcohol Clin Exp Res 2001;25(10):1467–71.
59. Wood DS, Gruenewald PJ. Local alcohol prohibition, police presence and serious injury in isolated Alaska Native villages. Addiction 2006;101(3):393–403.
60. Olafsdottir H. The dynamics of shifts in alcoholic beverage preference: effects of the legalization of beer in Iceland. J Stud Alcohol 1998; 59(1):107–14.
61. Wagenaar AC, Langley JD. Alcohol licensing system changes and alcohol consumption: introduction of wine into New Zealand grocery stores. Addiction 1995;90(6):773–83.
62. Blose JO, Holder HD. Public availability of distilled spirits: structural and reported consumption changes associated with liquor-by-the-drink. J Stud Alcohol 1987;48(4):371–9.
63. Dull RT. An assessment of the effects of alcohol ordinances on selected behaviors and conditions. J Drug Issues 1986;16(4):511–21.
64. Dull RT. Dry, damp, and wet: correlates and presumed consequences of local alcohol ordinances. Am Drug Alcohol Abuse 1988;14(4):499–514.
65. Gorman DM, Speer PW, Labouvie EW, Subaiya AP. Risk of assaultive violence and alcohol availability in New Jersey. Am J Public Health 1998;88(1):97–100.
66. Gorman DM, Labouvie EW, Speer PW, Subaiya AP. Alcohol availability and domestic violence. Am J Drug Alcohol Abuse 1998;24(4):661–73.
67. Gorman D, Speer P, Gruenewald P, Labouvie E. Spatial dynamics of alcohol availability, neighborhood structure and violent crime. J Stud Alcohol 2001;62(5):628–36.
68. Gyimah-Brempong K. Alcohol availability and crime: evidence from census tract data. South Econ J 2001;68(1):2–21.
69. Jewell R, Brown R. Alcohol availability and alcohol-related motor vehicle accidents. Appl Econ 1995;27:759–65.
70. Markowitz S, Grossman M. Alcohol regulation and domestic violence towards children. Contemp Econ Policy 1998;16(3):309–20.
71. Neuman C, Rabow J. Drinkers' use of physical availability of alcohol: buying habits and consumption level. Inter Addict 1985;20(11–12):1663–73.
72. Nielsen AL, Martinez R, Lee MT. Alcohol, ethnicity, and violence: the role of alcohol availability for Latino and Black aggravated assaults and robberies. Sociol Q 2005;46:479–502.
73. Ornstein S, Hanssens D. Alcohol control laws and the consumption of distilled spirits and beer. J Consum Res 1985;12(2):200–13.
74. Parker DA. Alcohol problems and the availability of alcohol. Alcohol Clin Exp Res 1979;3(4):309–12.

75. Pollack CE, Cubbin C, Ahn D, Winkleby M. Neighbourhood deprivation and alcohol consumption: does the availability of alcohol play a role? Int J Epidemiol 2005;34(4):772–80.
76. Reid RJ, Hughey J, Peterson NA. Generalizing the alcohol outlet-assaultive violence link: evidence from a U.S. Midwestern city. Subst Use Misuse 2003;38(14):1971–82.
77. Rush BR, Gliksman L, Brook R. Alcohol availability, alcohol consumption and alcohol-related damage. I. The distribution of consumption model. J Stud Alcohol 1986;47(1):1–10.
78. Scribner RA. The risk of assaultive violence and alcohol availability. Am J Public Health 1995;85(3):335–40.
79. Scribner RA, MacKinnon DP, Dwyer JH. Alcohol outlet density and motor vehicle crashes in Los Angeles County cities. J Stud Alcohol 1994;55(4):447–53.
80. Smart RG. Effects of two liquor store strikes on drunkenness, impaired driving and traffic accidents. J Stud Alcohol 1977;38(9):1785–9.
81. 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.
82. Stevenson RJ, Lind B, Weatherburn D. The relationship between alcohol sales and assault in New South Wales, Australia. Addiction 1999;94(3):397–410.
83. Tatlow JR, Clapp JD, Hohman MM. The relationship between the geographic density of alcohol outlets and alcohol-related hospital admissions in San Diego County. J Community Health 2000;25(1):79–88.
84. Treno A, Grube J, Martin S. Alcohol availability as a predictor of youth drinking and driving: a hierarchical analysis of survey and archival data. Alcohol Clin Exp Res 2003;27(5):835–40.
85. Weitzman ER, Folkman A, Folkman MP, Wechsler H. The relationship of alcohol outlet density to heavy and frequent drinking and drinking-related problems among college students at eight universities. Health Place 2003;9(1):1–6.
86. Zhu L, Gorman DM, Horel S. Alcohol outlet density and violence: a geospatial analysis. Alcohol Alcohol 2004;39(4):369–75.
87. Colon I. Alcohol availability and cirrhosis mortality rates by gender and race. Am J Public Health 1981;71(12):1325–8.
88. Colon I, Cutter HS. The relationship of beer consumption and state alcohol and motor vehicle policies to fatal accidents. J Safety Res 1983;14(2):83–9.
89. Freisthler B, Gruenewald PJ, Treno AJ, Lee J. Evaluating alcohol access and the alcohol environment in neighborhood areas. Alcohol Clin Exp Res 2003;27(3):477–84.
90. Freisthler B. A spatial analysis of social disorganization, alcohol access, and rates of child maltreatment in neighborhoods. Child Youth Serv Rev 2004;26(9):803–19.
91. Freisthler B, Needell B, Gruenewald PJ. Is the physical availability of alcohol and illicit drugs related to neighborhood rates of child maltreatment? Child Abuse Negl 2005;29(9):1049–60.
92. Godfrey C. Licensing and the demand for alcohol. Appl Econ 1988;20:1541–58.
93. Harford T, Parker D, Pautler C, Wolz M. Relationship between the number of on-premise outlets and alcoholism. Stud Alcohol 1979;40(11):1053–7.
94. Gruenewald PJ, Johnson FW, Treno AJ. Outlets, drinking and driving: a multilevel analysis of availability. J Stud Alcohol 2002;63(4):460–8.
95. Kelleher KJ, Pope SK, Kirby RS, Rickert VI. Alcohol availability and motor vehicle fatalities. J Adolesc Health 1996;19(5):325–30.
96. Lascala EA, Gerber D, Gruenewald PJ. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. Accid Anal Prev 2000;32:651–8.
97. LaScala EA, Johnson FW, Gruenewald PJ. Neighborhood characteristics of alcohol-related pedestrian injury collisions: a geostatistical analysis. Prev Sci 2001;2(2):123–34.
98. Lipton R, Gruenewald P. The spatial dynamics of violence and alcohol outlets. J Stud Alcohol 2002;63(2):187–95.
99. Rabow J, Watts RK. Alcohol availability, alcoholic beverage sales and alcohol-related problems. J Stud Alcohol 1982;43(7):767–801.
100. Roncek DW, Maier PA. Bars, blocks, and crimes revisited: linking the theory of routine activities to the empiricism of "hot spots." Criminology 1991;29(4):725–53.
101. Scribner R, Cohen D, Kaplan S, Allen SH. Alcohol availability and homicide in New Orleans: conceptual considerations for small area analysis of the effect of alcohol outlet density. J Stud Alcohol 1999;60(3):310–6.
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. Watts RK. Alcohol availability and alcohol-related problems in 213 California cities. Alcohol Clin Exp Res 1983;7(1):47–58.
104. Wieczorek WF, Coyle JJ. Targeting DWI prevention. J Prev Interv Community 1998;17(1):15–30.
105. van Oers JA, Garretsen HF. The geographic relationship between alcohol use, bars, liquor shops and traffic injuries in Rotterdam. J Stud Alcohol 1993;54(6):739–44.
106. Alaniz ML. Immigrants and violence: the importance of neighborhood context. Hisp J Behav Sci 1998;20(2):155–74.
107. Cohen DA, Mason K, Scribner R. The population consumption model, alcohol control practices, and alcohol-related traffic fatalities. Prev Med 2002;34(2):187–97.
108. Colon I. The influence of state monopoly of alcohol distribution and the frequency of package stores on single motor vehicle fatalities. Am J Drug Alcohol Abuse 1982;9(3):325–31.
109. Freisthler B, Midanik LT, Gruenewald PJ. Alcohol outlets and child physical abuse and neglect: applying routine activities theory to the study of child maltreatment. J Stud Alcohol 2004;65(5):586–92.
110. Gorman DM, Zhu L, Horel S. Drug 'hot-spots,' alcohol availability and violence 220. Drug Alcohol Rev 2005;24(6):507–13.
111. Scribner RA, Cohen DA, Fisher W. Evidence of a structural effect for alcohol outlet density: a multilevel analysis. Alcohol Clin Exp Res 2000;24(2):188–95.
112. McCarthy PS. Public policy and highway safety: a city-wide perspective. Reg Sci Urban Econ 1999;29(2):231–44.
113. Wittman FD, Hilton ME. Uses of planning and zoning ordinances to regulate alcohol outlets in California cities. In: Holder HD, ed. Control issues in alcohol abuse prevention: strategies for states and communities. Greenwich CT: JAI Press, 1987:337–66.
114. Donnelly N, Poynton S, Weatherburn D, Bamford E, Nottage J. Liquor outlet concentrations and alcohol-related neighborhood problems. Alcohol Stud Bull 2006;8:1–16.
115. Giesbrecht N. Roles of commercial interests in alcohol policies: recent developments in North America. Addiction 2000;95(4):S581–95.
116. National Highway Traffic Safety Administration. The role of alcohol beverage control agencies in the enforcement and adjudication of alcohol laws. No. DOT HS 809 877.
117. Mann R, Rehm J, Giesbrecht N, et al. The effect of different forms of alcohol distribution and retailing on alcohol consumption and problems—an analysis of available research. Toronto: Center for Addiction and Mental Health, 2005.

Collaboration in Academic Medicine: Reflections on Gender and Advancement

Phyllis L. Carr, MD, Linda Pololi, MBBS, Sharon Knight, PhD, and Peter Conrad, PhD

Abstract

Purpose
Collaboration in academic medicine is encouraged, yet no one has studied the environment in which faculty collaborate. The authors investigated how faculty experienced collaboration and the institutional atmosphere for collaboration.

Method
In 2007, as part of a qualitative study of faculty in five disparate U.S. medical schools, the authors interviewed 96 medical faculty at different career stages and in diverse specialties, with an oversampling of women, minorities, and generalists, regarding their perceptions and experiences of collaboration in academic medicine. Data analysis was inductive and driven by the grounded theory tradition.

Results
Female faculty expressed enthusiasm about the potential and process of collaboration; male faculty were more likely to focus on outcomes. Senior faculty experienced a more collaborative environment than early career faculty, who faced numerous barriers to collaboration: the hierarchy of medical academe, advancement criteria, and the lack of infrastructure supportive of collaboration. Research faculty appreciated shared ideas, knowledge, resources, and the increased productivity that could result from collaboration, but they were acutely aware that advancement requires an independent body of work, which was a major deterrent to collaboration among early career faculty.

Conclusions
Academic medicine faculty have differing views on the impact and benefits of collaboration. Early career faculty face concerning obstacles to collaboration. Female faculty seemed more appreciative of the process of collaboration, which may be of importance for transitioning to a more collaborative academic environment. A reevaluation of effective benchmarks for promotion of faculty is warranted to address the often exclusive reliance on individualistic achievement.

Acad Med. 2009; 84:1447–1453.

Over the past several years, there has been a greater emphasis on faculty collaboration in academic medicine, particularly in research. The National Institutes of Health (NIH) Roadmap initiative has provided incentives for such collaboration,[1] and Darrell Kirch,[2] in his President's Address at the Association of American Medical Colleges (AAMC) annual meeting in November 2007, noted the need for the culture of academic medicine to change from one of individualistic achievement to one that values and rewards collaboration and teamwork. Collaboration, as defined by Chrislip and Larson,[3] is "a mutually beneficial relationship between two or more parties who work together toward common goals by sharing knowledge, learning, responsibility, authority and accountability for achieving results." In the fields of business and education, collaboration is widely and effectively utilized. Collaborative teamwork garners greater resources, recognition, and rewards when facing competition for finite resources.[4]

Few have studied collaboration in academic medicine, where faculty collaboration is usually an intellectual endeavor that is creative in nature. Successful academic collaboration in medicine, as opposed to more task-oriented instrumental collaboration, entails participants learning from and with one another. Although the collaborative process seems appropriate in academic medical research, an inherent contradiction exists: the promotion and advancement of faculty in academe traditionally requires an individual body of work.[2] We sought to understand how faculty in medical academe define and view collaboration and how they describe the atmosphere for collaboration in academic medicine. We were especially interested in the influences of gender and career stage on perceptions about collaboration.

Dr. Carr is professor of medicine and associate dean for student affairs, Boston University School of Medicine, Boston, Massachusetts.

Dr. Pololi is senior scientist and resident scholar, Brandeis University, Waltham, Massachusetts.

Dr. Knight is associate dean and director of graduate studies and professor, Department of Health Education and Promotion, East Carolina University, Greenville, North Carolina.

Dr. Conrad is Harry Coplan Professor of Social Sciences at Brandeis University, Waltham, Massachusetts.

Correspondence should be addressed to Dr. Carr, Office of Student Affairs, Boston University School of Medicine, A213, 72 East Concord Street, Boston, MA 02118; telephone: (617) 638-4166; fax: (617) 638-4491; e-mail: (PLCARR@BU.EDU).

Method

In 2007, as part of a larger qualitative study on faculty at diverse U.S. medical schools, we selected five schools to represent the diverse organizational characteristics (i.e., public/private, NIH research intensive, primary care orientation/community orientation) of the then 126 medical schools in the United States. The sample included representation from each of the designated AAMC regions, with two schools from the Northeastern region where the bulk of medical schools are concentrated. Faculty from two public and three private schools, two research-oriented schools as determined by NIH funding, one school with a primary care orientation, and three with a community orientation participated in the study. The demographics regarding female and underrepresented minority faculty in these five schools were almost identical to

statistics regarding women and minorities at medical schools nationally.

Participant criteria

We used both stratified purposeful and chain referral sampling strategies (i.e., interviewed faculty suggested other faculty who fit particular descriptions) to identify medical faculty from the five C–Change (The National Initiative on Gender, Culture and Leadership in Medicine) medical schools[5] on the basis of their medical school affiliation, gender, race/ethnicity, department/discipline, and career stage. Participants were research scientists, medical and surgical subspecialists, and generalist medical faculty holding doctoral degrees (84% MD/DO, 16% PhD). We interviewed 96 faculty members who were in one of four possible career stages that we felt would provide a thoughtful spectrum of medical faculty experience: (1) early career, those who were in their initial faculty appointment for two to five years, (2) "plateaued," those faculty who had not advanced as expected in rank and responsibility and who had been faculty members for 10 or more years, (3) faculty in leadership roles such as deans, department chairs, and center directors (identified below as "senior" faculty), and (4) former faculty who had left academic medicine.

Data collection and analysis

We four authors personally conducted in-depth, open-ended interviews—15% in person and the remainder by telephone. Interviews, typically one hour in length, were audio-recorded and transcribed verbatim. We used an interview guide that we developed through a pilot series of interviews. The guide included questions such as

- What is it about your work that energizes you?
- What do you see as valued in your institution?
- How does your institution encourage interdisciplinary collaboration?
- When have you felt most successful in your work?

The data were coded and all names and identifying information were removed. We analyzed the data in aggregate by repeatedly reading the interview transcripts to develop understanding and interpret meaning. The transcribed narrative data were stored, coded, and sorted using Atlas.ti software. Analysis involved reducing or condensing data to identify patterns and themes as they emerged from the coded data. The analysis process was inductive and driven by the Grounded Theory tradition.[6,7] We drew and verified our conclusions by continually reviewing the transcripts and challenging one another's findings until we developed consensus. The institutional review board at each of the investigators' institutions (Brandeis University, Boston University School of Medicine, and East Carolina Brody School of Medicine) approved this study.

Results

Participant demographics

We invited 170 faculty to participate. Eight individuals refused, primarily because of time constraints, and 54 did not respond; we were unable to schedule interviews with 12 potential participants. We identified plateaued and early career faculty in advance through key informant interviews. Identifying male plateaued faculty was more difficult than locating female faculty at a similar stage. Overall, those in leadership were more likely to agree to participate, and we found securing interviews with early career faculty the most difficult (Table 1). However, the smaller number of early career faculty in the study was due to the fact that we reached saturation and were no longer obtaining new data from this group. Female and underrepresented minority faculty were purposefully oversampled, as were generalists, a group comprising general internal medicine, family medicine, and general pediatrics. Fifty-five percent of the participants were women; 17% were African American/black, 4% were Hispanic/Latino, 79% were Caucasian/white, and 20% were generalists. Interviewees were similarly represented in the plateaued group (n = 23) and among those who had left academic medicine (n = 24). The largest number of participants fell into the senior leadership group (n = 28). Fewer were in the early career group (n = 21), because, as discussed above, we stopped interviewing such faculty when we no longer obtained new data. We interviewed similar numbers of faculty from each school.

Table 1

Faculty Participants (N = 96) by Career Stage and Gender in a 2007 Qualitative Study of Collaboration in Academia

Career stage*	Female faculty	Male faculty
Early	14	7
Plateau	13	10
Senior	16	12
Left	13	11
Total	*53*	*43*

* Early faculty are those who have been faculty for two to five years; plateaued faculty are those who have not advanced as expected in rank and responsibility and who have been faculty members for 10 or more years; senior faculty are those who are in leadership roles such as deans, department chairs, and division and center directors; faculty who have left are no longer working in academia.

Over 4,000 pages of data were recorded. In this manuscript, participants' quotations are identified by gender, degree, and the faculty group (defined above) to which the participant belonged, as well as the area in which the collaboration occurred (research, administrative, clinical, or educational).

The results below are grouped by the seven emergent themes extracted from the data and presented in the following order: (1) the gendered meaning, value, and experience of collaboration, (2) career stage perspectives on collaboration, (3) barriers to collaboration, (4) collaboration in research—barriers and disadvantages, (5) collaboration in medical education, and (6) outcomes of collaboration.

Gendered meaning, value, and experience of collaboration

Faculty reflected on the meanings and experiences of collaboration in several ways. For some faculty, collaboration was a reason for choosing a career in academe. Several female faculty members described it as a way of working together and learning and growing in their work. Female faculty tended to value the concept and process of collaboration at all career stages:

> It's fascinating and exciting to me to see how it's done elsewhere and get ideas . . . it's really a shared knowledge . . . collaboration to learn, impart knowledge and look at things in a new way . . . broader than just what we each do as individuals and you can actually reach more people.

—Plateaued female physician (educational collaboration)

Academic medicine offered a wide variety of opportunities . . . it offered the opportunity to teach and to work with colleagues . . .. I also liked talking about ideas and thinking about ideas and collaborating on projects with colleagues.

—Female physician who left academic medicine (educational collaboration)

It was the greatest thing, I mean having that collaboration . . . I think, more and more, we think less about the departmental structure and more about collaboration than ever before.

—Senior female physician (administrative collaboration)

Compared with male faculty, more female faculty commented on the interpersonal relationships associated with collaboration. The relationships that could result from collaboration were important to faculty. Whether the collaboration resulted in the positive bonds or whether these bonds helped to bring about collaboration was unclear, but relationships were clearly an important aspect of the collaboration process, especially for female faculty in our study:

I actually get energized by the relationships that I develop with both medical students and residents . . . as well as relationships with other faculty. I do a lot of community work and outreach, and so I think that's what draws me to it [collaboration] more than anything. It's the contacts that I develop with people and feel a bond with.

—Plateaued female physician (clinical and educational collaboration)

Female faculty viewed collaboration as a sharing of ideas and knowledge and as a way of looking at things from a new perspective. Male faculty often recounted more pragmatic reasons for collaboration such as the provision of research opportunities:

I actually have been ask[ed] to join that group several times and have had other collaborative things done because of it . . . and again . . . it came about because of my work ethic and the way I do things clinically, and for being a team player because it's very political.

—Early career male physician (clinical and research collaboration)

Male faculty recognized that chances for collaboration could be political and thus intentionally positioned themselves to acquire these opportunities. Male faculty focused on outcomes, such as research opportunities, from their collaborative work, whereas female faculty typically focused on the process of collaboration, the generation of ideas, and the enjoyment of actual participation.

Faculty expressed the notion that a collegial, collaborative atmosphere would be the ideal environment in which to work in academic medicine. Collaboration was also seen as a means of having faculty interaction as part of the process of decision making.

It's a much better environment if there's collaboration, cooperation and support . . .. I really try to make a collaborative environment where everybody feels that they're part of the decision making . . . it's a much better environment if there's collaboration and support.

—Senior female physician (administrative collaboration)

Career stage perspectives

The career stage of faculty influenced their perspectives on collaboration. Senior faculty in leadership positions, regardless of gender, viewed collaboration as an opportunity for brainstorming, collective decision making, problem solving, and accessing other points of view. They expressed a consistently positive view of these opportunities:

The chairs get together as a group to address common problems . . . and to try to evolve collegial, collaborative ways of overcoming and addressing these problems.

—Senior male PhD faculty (administrative collaboration)

Certainly at a high administrative level, we have a small core of administrative staff . . . we have two or three hours of . . . brainstorming meetings every week where we really . . . wrestle with issues and try to make decisions in a collaborative way.

—Senior male physician (administrative collaboration)

Senior male faculty sometimes viewed the relationships stemming from collaboration as networking, a more formal relationship to other colleagues, both at their institution and in the larger community of academic medicine.

Part of it [collaboration] is a network of colleagues—I mentioned locally, but also nationally . . . we get together sporadically . . . they tend to be in leadership roles in other institutions and we support each other in trying to affect these big complex institutions.

—Senior male physician (administrative collaboration)

These faculty clearly enjoyed the process of collaboration and gained from the input of other senior leaders in solving administrative problems in the workings of academic institutions. Administrative collaboration tended to be more pragmatic than collaboration in research, which more formally evaluates and studies a scientific question by obtaining data.

I really loved faculty development and building programs and working with my colleagues, fellow chairs, on just trying to build the institution . . .. I enjoyed it so much because I do like the interaction with the faculty and other leaders in the institution.

—Senior female physician (administrative collaboration)

For many senior faculty, a form of instrumental collaboration provided opportunities to interact with other faculty and to build interpersonal bonds or networks through collaboration.

Unlike senior faculty leaders who endorsed and embraced collaboration in their own work, early career faculty faced important impediments to collaboration. One of the major impediments was criteria for career advancement.

Barriers to collaboration

Faculty tended to work in academic environments that did not pervasively value collaboration regardless of career stage. Early career faculty, for example, tended to describe an environment that was one of competitive self-interest rather than collaboration and clearly recognized the lack of support for collaboration:

It's like playing high school football again, you know? Really it's the same sort of mentality. You watch out for the other person, I'll help you, what can I do to help you get out [of work] sooner, rather than everybody for themselves . . . in academic medicine, unfortunately, I see more of that attitude than the former, meaning that it's everybody for

themselves . . . I think more selfish than competitive.

—Early career male physician (research collaboration)

I mean that would be a positive in my life if I felt we were all sort of on this team working towards this common goal, but I don't have that.

—Early career male physician (clinical collaboration)

The hierarchical nature of academic medicine served to inhibit opportunities and collegial environments for effective collaboration. In fact, for some faculty the traditional hierarchy was the antithesis of collaboration:

What kind of structure would that (function- or purpose-driven organization) foster . . . it would look a whole lot different I would think . . . a lot less hierarchical, a whole lot more collaborative.

—Female physician who left academic medicine (research and educational collaboration)

Hierarchy creates a structure that makes communication up the line more difficult for junior faculty who would benefit from more interaction with senior and leadership faculty. Junior faculty saw senior faculty as contributing to the hierarchical structure. A senior male physician stated, "I'm the boss, don't hesitate to talk to somebody who's keeping me from ever having to talk to you." These junior faculty were concerned that departments mirrored some of their more senior faculty who could be averse to change, making it hard for new ideas to take hold. An early career stage female physician put it this way: "I think that the upper administration does not appear to be aware of the problems we have, which I think is very strange, because at one point they had to be where we are now." And a senior male physician said, "It's very hierarchical [so] those at lower levels have minimum input. . . . Certainly not into major strategic decisions at a departmental level. It's all held at a very high level."

The lack of cooperation between departments in academic medicine also works against effective collaboration as stated by this former faculty member:

I would like to see people . . . be more curious and more respectful of each other, so not be as guarded and defensive as specialties sometimes become.

—Male physician who left academic medicine (research collaboration)

Collaboration in research

Barriers. There were barriers to research collaboration as well, including the difficulty faculty encountered learning about the work of other researchers in their institution with similar interests.

It's [collaboration] great when it happens. I mean, but when it happens, it's because you are sitting with somebody at lunch . . . it's because you are sitting there and talking about your work and you think, "Oh, I work on that. Let's do something together" . . . that's how it happens.

—Senior female PhD (research collaboration)

These informal pathways to research collaboration are common, and many faculty saw the potential of having more formal ways to encourage collaboration, such as through institutional research databases, rather than to have it occur through chance. Early career faculty involved in clinical research also expressed low expectations of other faculty contributing and collaborating with them on their research:

It's not that they don't want to be involved, it's just that they are really busy and to expect other people to care about what I'm doing, for the most part is—I don't. So I don't expect it, so I'm not very disappointed by it . . . I think isolated . . . the only time isolated bothers me is when I am trying to enroll patients in studies that I need my partners' patients. . . . I never get resistance, but I never get the level of support that would make this all better.

—Early career male physician (research collaboration)

In addition to the concerns by some male faculty about the lack of cooperation from other faculty in furthering their research studies, clinical faculty had other concerns. They identified the heavy clinical load expected by hospitals from their physician staff as a barrier to collaboration and wanted to ensure that somehow research collaboration would not be stifled for clinical faculty:

Cultivating that environment for collaborative work between that handful of researchers who are struggling to get time off from their clinical expectations.

—Early career woman PhD (research collaboration)

A major barrier for research collaboration is faculty advancement. The system of promotion usually requires an individual body of work that is clearly that of the faculty member who is under consideration for promotion. Although faculty might wish to collaborate and realize that they could gain greater productivity through collaboration, they also realize that they would be promoted or promoted earlier if their work were completely their own and separate from their mentor's line of research. Collaboration in academic medical research is the most difficult to achieve because of this constraint. As one plateaued female physician (research collaboration) stated,

My model for research is more a group model and I still work very closely with my scientific mentor, which is, you know, great, in terms of being able to use his resources, where I haven't had them, but it also hurts you in that it keeps you from looking independent, which is part of advancement.

Faculty were also aware of the contradictions between, on one side, this policy of independence and the reality of academic life and, on the other, collaboration and the need for academic medicine to change.

The goal, this whole goal of becoming an independent researcher which truthfully I think is kind of bunk at the end of the day. I think, to me, and I think the new Institute Roadmap is acknowledging that we do get a lot more done when we work together, in group mission, greater productivity. . . . You always felt like you were working together, helping one another. I have not seen that. . . . I think they [large labs that collaborate] are few and far between.

—Early career male physician (research collaboration)

Collaboration was valued and noted to provide a means of greater research productivity, but at the end of the day, faculty were very aware of the need to have accomplished goals independently.

Disadvantages. One faculty member observed that the pace of the work that was accomplished could be negatively impacted by collaboration. Investigators needed to rely on other people to complete their work in a timely way. Faculty sensed a loss of control:

> [When] you need to rely on other people . . . other factors come into play, like how hard others work.
>
> —Early career female PhD (research collaboration)

Working with others, therefore, had its limitations as well as advantages.

Collaboration in medical education

Many faculty enjoyed medical education because of the many possible avenues for collaboration. They valued the excellence of work that resulted from effective collaboration, as well as having a pragmatic outcome such as a grant or curricula.

> I wrote a bunch of curricula while I was there and changed some courses fairly dramatically, and I also liked talking about ideas and thinking about ideas and collaborating on projects with other colleagues.
>
> —Female physician who left academic medicine (educational collaboration)

Faculty also noted that collaboration as part of the leadership structure of a residency program was effective, valuable, and common to many programs.

> So when I was offered the position of training director, there was no associate training director and I advocated aggressively to name an associate training director . . .. I had a lot of resistance about that from my chair. Eventually she agreed and the program has been so much better.
>
> —Early career male physician (educational collaboration)

> It's a lot of work, but when you see it [new curriculum] come together, or even if it falls apart, it's such a learning process for me, and I think for them, and hopefully that we are able to—as a group—to impart some knowledge or at least a new way of looking at something to groups of students, residents, or our colleagues.
>
> —Plateaued female physician (educational collaboration)

Outcomes

Many faculty, researchers, clinicians, and educators, particularly male faculty, expressed the need for tangible outcomes from collaboration. They felt that if they engaged in collaboration, there needed to be something in it for them, whether it was a publication, a grant, or curricula. Female faculty again imparted a sense of enjoying the process as much as the outcome (see above quotation), and of having the ability to choose faculty for their committee as a way of increasing the probability of having more thoughtful and effective members:

> What I liked about it was working with people from different backgrounds and coming together to work on a common problem and the other people there made me think of things in a way that I hadn't before. And they were people I was able to handpick . . . for being very thoughtful people . . .. That was a project that involved a lot of intellectual energy, that had an outcome, you know that had a grant, a product that was really a challenge to produce.
>
> —Plateaued female physician (educational collaboration)

This female faculty member has not advanced as expected, yet she is supportive of the idea of collaboration and finds the process in and of itself valuable, regardless of the outcome.

Some faculty, often female, were generous with the outcomes that they produced collaboratively, as illustrated in the following statement:

> So we set major goals, we worked together and one of the things that happened . . .. I was very generous with authorship. If you did something for the [research] project, you could be an author. And so that way everybody felt a part of it.
>
> —Senior female physician (research collaboration)

> It's several different individuals collaborating to work together to achieve a goal and publish. Just recently we recognized that there was a problem with stress, depression, and burnout . . . we had an intervention that we studied as a research project.
>
> —Senior female physician (research collaboration)

Female faculty design opportunities that allow others to share in the outcome of their work; they value the process of collaboration, which they use as team building with colleagues. Male faculty, particularly those still in the early stages of their careers, did not express the idea of providing such a shared outcome but, rather, a more competitive approach that would result in greater research opportunities for taking on additional work (see "Barriers to Research," the first quote: "It's everybody for themselves . . . I think more selfish than competitive").

There were no male faculty at any stage of their careers in this study who expressed value in the process of collaboration regardless of an outcome.

Discussion

Leaders have promulgated effective collaboration as facilitating excellence in research in academic medicine, but often neither the atmosphere nor the advancement criteria in medical institutions are conducive to such collaboration. The need to transform the individualistic and competitive environment to one that is collaborative and cooperative has been voiced by senior leadership in medicine.[1,2] There is almost no literature about collaboration in academic medicine, despite the importance of this concept. The lone, autonomous investigator who competes for research grants and advancement has long been the model for success in academic medicine. Many of the early career male faculty in our study expressed a competitiveness in the academic environment that female faculty did not experience, recognize, or create. Rather, female faculty revealed an enthusiasm about the process of collaboration, which was at odds with early career male faculty, who seemed outcome-oriented in their vision of collaboration. The concepts that female faculty associated with collaboration, such as learning and growing, sharing ideas, and gaining new perspectives—similar to those expressed by Chrislip and Larson[3]—are at the heart of an effective process of collaboration.

Our respondents described ways of collaborating that fit two of three descriptions of collaboration detailed by Schneider[8]: (1) brainstorming or collaboration by interest, in which collaborators spontaneously discuss a common problem at hand—this is described by senior faculty who approach institutional problems through this approach—and (2) collaboration by the leader, in which the collaborators are chosen by the committee chair and thus often have compatible values and come from similar working environments—this type of collaboration is referred to in our study by a plateaued female physician who found it essential to have thoughtful faculty whom she had selected in her committee (first quote in the outcome

section). Schneider's third way of collaboration (not specifically mentioned by our participants) is collaboration by acuity, in which the collaborators are selected to create a balanced skill set among the members. In addition to these types of collaboration, we also noted a form of instrumental collaboration through which faculty worked on large projects with multiple-author papers, essentially completing work together but not necessarily learning and growing together. Collaboration by acuity is particularly important in academic medicine because of the highly specialized nature and increasing complexity of medical knowledge. The lone, autonomous investigator is at a disadvantage as collaboration is even more vital now than in times past.

Both the lack of formal pathways in many institutions to facilitate collaboration among faculty and the barriers between departments hinder effective collaboration. One of our key findings is that faculty readily concurred that the best environment in academe is one in which cooperation, support, and collaboration exist, yet few of our faculty participants, and particularly early career faculty, described such an environment at their current institution.

Paul T.P. Wong[9] describes five types of "toxic" cultures that inhibit collaboration: authoritarian–hierarchical, competing–conflicting, laissez faire, dishonest–corrupt, and rigid–traditional. Several aspects of academic medicine contain features of these "toxic" cultures: the hierarchical and competitive aspects of the culture, which faculty participants repeatedly noted as barriers to collaboration; chairs who are not accessible to talk with faculty; and the competitive environment for individualistic advancement sensed by early career male faculty. In business, leaders have understood the importance of a healthy corporate culture that includes trust building, transparency, accountability, and empowerment of the workforce. A hierarchical culture does not permit the equal sharing of ideas so that all participants can contribute, learn, and grow. Instead, in medical academe there is often the unilateral control of the chair or chief, resulting in lost opportunity for mutual learning, equal input, and high-quality decisions.[9]

Whereas the hierarchy and competitiveness of academe are prominent to junior faculty, senior faculty have reached another stage in their careers at which they tend to view collaboration as brainstorming; consequently, developing ideas and solving problems are forms of collaboration. Initial brainstorming often sets the tone for further work, which involves agreeing on similar group values. Senior leaders in academic medicine are likely to share a common set of academic and institutional priorities, which aid in making such sessions collaborative.[4] Senior leaders are at the top of the hierarchy and can follow a mutual learning model in which there is transparency, sharing of information, decision making by building a consensus, and a greater commitment to implementing such decisions.[10] For effective collaboration to occur, the ideas of every person need to be respected and the value of each person's contribution appreciated. From this mutual respect and appreciation, a shared field of meaning can emerge,[11] and differences can be seen as opportunities for learning.[10] These qualities of effective collaboration seem to thrive in the brainstorming of senior faculty. Leaders seem to accept collaboration as an effective means to an end for themselves in the roles that they play, but perhaps not necessarily for junior faculty. The role of the leader suggests that advancement has occurred to some degree before a role change. Views on collaboration may change once someone meets at least initial advancement criteria and moves into an academic leadership role, such as chair or dean.

These senior faculty almost uniformly described a connection to colleagues, a readiness to collaborate with other senior faculty for problem solving, and an enhanced satisfaction from this process. Junior faculty expressed concerns both about the atmosphere in academic medicine being adverse to sharing work and about their lack of connection to other faculty. This lack of connection, despite the greater ease of communication through e-mail, faxing, and teleconferencing, suggests that relationship formation cannot be adequately achieved without direct interpersonal meetings and exchanges. Likewise, in a culture in which rank or job title is important, it can be difficult for a lower-ranking faculty member to access more senior faculty for the purpose of collaboration. The disconnect between junior and senior faculty in our study was vivid and problematic. Particularly concerning is that early career faculty were very aware of the barriers to collaboration and pessimistic that effective, productive collaboration was possible. They were highly aware of the hierarchy, as they were close to the bottom of it. Greater efforts are necessary to support junior faculty and to create a more collaborative culture at all levels of academe.

Female faculty remained enthusiastic and positive about collaboration, whereas male faculty, particularly early career male faculty, seemed much more pragmatic about collaboration, valuing it for its outcomes. Nonetheless, these early career male faculty still perceived the individualistic and competitive atmosphere in academe. Female faculty in this study expressed greater interest in collaboration than did their male colleagues, with less concern for the outcome than the processes and the learning involved with the work. Notably, though, many of these female faculty had either plateaued or left medical academe. It is also disconcerting that it was easier to find plateaued female faculty compared with male faculty. If effective collaboration is more valued, women may find it easier to advance in their academic careers and thus be less likely to plateau.

Our study has a number of limitations. The in-depth and time-consuming nature of the interviews necessitated a small sample size, so whether our findings are generalizable is not clear. Another investigator examining our data may not extract the same themes, but we found the themes consistent and congruent. Our study also has significant strengths. This work is the first on collaboration in academic medicine and provides national information from five medical schools which are quite representative of medical schools nationally by geography, in public–private status, and demographically. The qualitative methods reveal the experience of faculty in their own rich language and according to their own understanding.

Conclusions

The descriptions of the academic environment provided by the faculty we interviewed, especially junior faculty, are concerning. Creating a supportive, collegial, and collaborative atmosphere in medical academe needs to be a high priority. Female faculty could be a particularly valuable asset in this transition because they seem to value collaboration, both the process and the outcomes. Reciprocally, building a more collaborative atmosphere could improve the position of female faculty in academic medicine. Better benchmarks to judge the accomplishments of faculty are needed so that the desire for and actual steps toward advancement are not at odds with the increasingly complex and interdisciplinary process of academic work. The criteria for promotion need to change to reflect the value of effective collaboration.

Acknowledgments

The authors gratefully acknowledge the critical funding support of the Josiah Macy Jr., Foundation and supplemental funding provided by Office of Public Health and Science, Office of Women's Health, and Office of Minority Health; National Institutes of Health, Office of Research on Women's Health; Agency for Healthcare Research and Quality; Centers for Disease Control and Prevention; and Health Resources and Service Administration. The authors also thank Marianne McPherson and Meg Lovejoy for assistance with data coding. The authors are indebted to the medical faculty who willingly shared their thoughts with them.

References

1 Zerhouni E. Medicine: The NIH roadmap. Science. 2003;302:63–72.
2 Kirch DG. President's address: Culture and the courage to change. Available at: (http://www.aamc.org/meetings/annual/2007/highlights/president.htm). Accessed June 17, 2009.
3 Chrislip DD, Larson CE. Collaborative leadership: How citizens and civic leadership can make a difference. In: Schuman S, ed. Creating a Culture of Collaboration. San Francisco, Calif: Jossey-Bass; 1994.
4 Wagner CS, Loet L. Globalisation in the network of science in 2005: The diffusion of international collaboration and the formation of a core group. Available at: (http://203.199.213.48/1271/1/globalisation.pdf). Accessed June 17, 2009.
5 Biernacki P, Waldorf D. Snowball sampling: Problems and techniques of chain referral sampling. Sociol Methods Res. 1981;10:141–163.
6 Glaser B, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago, Ill: Aldine; 1967.
7 Charmaz K. Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis. Thousand Oaks, Calif: Sage Publications; 2006.
8 Schneider F. Collaboration: Some thoughts of new ways of working and learning together. Available at: (http://Summit.Kein.org/node/190). Accessed August 18, 2009.
9 Wong PTP. Is your organization an obstacle course or a relay team? A meaning centered approach to a collaborative culture. In: Schuman S. Creating a Culture of Collaboration: The International Association of Facilitators Handbook. San Francisco, Calif: Jossey-Bass; 2006.
10 Schwartz R. Using the facilitative leader approach to create an organizational culture of collaboration. In: Schuman S. Creating a Culture of Collaboration: The International Association of Facilitators Handbook. San Francisco, Calif: Jossey-Bass; 2006.
11 Zubizarreta R. Practical dialogue: Emergent approaches for effective collaboration In: Schuman S. Creating a Culture of Collaboration: The International Association of Facilitators Handbook. San Francisco, Calif: Jossey-Bass; 2006.

UPDATE

Update in HIV Medicine for the Generalist

Amina A. Chaudhry, MD MPH[1], Allen L. Gifford, MD[2], Joseph Cofrancesco Jr., MD MPH[1], James Sosman, MD[3], Gail Berkenblit, MD, PhD[1], and Lynn E. Sullivan, MD[4]

[1]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [2]Department of Medicine, VA New England Healthcare and Boston University School of Medicine, Boston, MA, USA; [3]Department of Medicine, University of Wisconsin School of Medicine, Madison, WI, USA; [4]Department of Medicine, Yale University School of Medicine, New Haven, CT, USA.

J Gen Intern Med 24(2):276–82
DOI: 10.1007/s11606-008-0851-2

INTRODUCTION

This update reviews the literature in the prevention and management of HIV infection as pertinent to the general internist. Our objectives were (1) to review the most recent data regarding HIV prevention, counseling, testing, and treatment, with a special focus on cost-effectiveness; (2) to discuss new findings regarding survival and general management of HIV, including a brief review of updated treatment guidelines and new antiretroviral medications; and (3) to discuss the intersection of HIV with other chronic diseases commonly encountered by generalist physicians. We included papers with both domestic and international relevance.

We performed a literature search of peer-reviewed studies published since July 2006. Initially, we performed a PUBMED search using the Medical Subject Heading (MeSH) term "HIV," limiting our search to English articles dealing with human subjects published in core clinical journals on or after July 1, 2006. We narrowed our results to studies focusing on the following categories: prevention, counseling and testing, survival, management, and HIV and other chronic diseases. Additionally, we reviewed studies published since July 2006 in the major internal medicine journals and HIV specialty journals, and included articles based on recommendations by experts in the field. Final selection of articles was by group consensus of HIV experts and practicing HIV clinicians.

Received July 29, 2008
Revised October 7, 2008
Accepted October 14, 2008
Published online November 11, 2008

PREVENTION: THE ROLE OF MALE CIRCUMCISION

Gray RH, Kigozi G, Serwadda D, et al. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *Lancet*. 2007;369:657–666.

More than 40 epidemiologic studies have suggested a beneficial effect of male circumcision in preventing HIV acquisition among heterosexual men.[1] The biological explanation for this may be limiting the vulnerable mucosal surface. Several recent randomized controlled trials were performed to assess safety and efficacy of male circumcision in reducing female-to-male HIV transmission.[2,3] Gray et al. reported on the largest of these trials, performed in Rakai, Uganda.

This study enrolled 4,996 uncircumcised, HIV-negative men aged 15–49 years who were randomly assigned to receive immediate circumcision or circumcision delayed for 24 months. The primary outcome was HIV incidence. This trial was stopped early after interim analysis showed significant efficacy for circumcision. In a modified intention to treat analysis, 24-month incidence was 0.66/100 person-years in the immediate circumcision group versus 1.33/100 person-years in the control group, corresponding to a 51% efficacy (95% CI, 16%–72%; P=0.006). Accounting for crossovers in an as-treated analysis, efficacy increased to 55% (95% CI, 22%–75%; P=0.003). HIV incidence was lower in the immediate circumcision group in all sociodemographic, behavioral, and sexually transmitted disease subgroups. This effect appeared to be stronger among men with two or more partners, extramarital partners, and during later follow-up periods. Moderate or severe adverse events were relatively rare. All adverse events resolved with treatment. Behaviors were similar in both groups. The Rakai study findings are consistent with the other two circumcision trials performed in Africa.[2,3]

One caveat is that circumcision in these studies was performed by trained personnel using sterile technique in well-equipped facilities. Complications might increase under less controlled conditions. Additionally, preliminary data from another study raise concern about potential increased HIV transmission to female partners of HIV-positive men circumcised as adults, although this difference was not statistically significant.[4] Still, the World Health Organization

and UNAIDS recommend that circumcision be considered in high-prevalence settings.[5] The Rakai study supports this approach. However, these conclusions should not be extended to imply reduced transmission to circumcised men who have sex with men, or to women with circumcised HIV-positive male partners.

COST-EFFECTIVENESS

Paltiel AD, Walensky RP, Schackman BR, et al. Expanded HIV screening in the United States: Effect on clinical outcomes, HIV transmission, and costs. *Ann Intern Med*. 2006 Dec 5;145(11):797–806

Sanders GD, Bayoumi AM, Holodniy M, Owens DK. Cost-effectiveness of HIV screening in patients older than 55 years of age. *Ann of Int Med*. 2008;148(12):889–903.

Goldie SJ, Yazdanpanah Y, Losina E, et al. Cost-effectiveness of HIV treatment in resource-poor settings–the case of Cote d'Ivoire. *N Engl J Med*. 2006 Sep 14;355(11):1141–53.

During the last 3 years, there has been renewed emphasis on expanding HIV testing in primary care.[6] Studies have previously demonstrated that under most conditions in the US, cost-effectiveness ratios for routine HIV testing are substantially less than the commonly used benchmark of $50,000 per quality-adjusted life year (QALY).[7,8] More recent research taking into account new testing and treatment strategies and effects on future HIV transmission has strengthened these findings.

Paltiel and colleagues used cost-effectiveness modeling to incorporate the effects of testing and treatment on future transmission. The number of infections that one person may transmit depends on the number of infective contacts, the behavioral risk of each contact, the efficiency of transmission, and the lifetime duration of HIV infectivity. In theory, HIV testing may *decrease* subsequent infections by decreasing risk behaviors and through viral suppression with antiretroviral therapy. However, testing and treatment might conversely *increase* subsequent infections by disinhibiting behavior and prolonging life, leading to more contacts.

They set the mean base lifetime transmission rate from a single infected individual (R_0) at 1.44. In their simulation model, they found that for a typical population (1% HIV prevalence) one-time screening had a cost-effectiveness ratio of $30,800 per QALY, and screening every 5 years a ratio of $32,300 per QALY. Assuming an adverse effect of HIV testing on future transmission (if testing were to *increase* transmission), one-time screening still conferred a net health benefit, with the cost-effectiveness ratio increased, but still a good value at $44,200/QALY. Screening every 3 to 5 years became quite costly, with cost-effectiveness ratios exceeding $100,000/QALY. However, using more favorable assumptions (that testing would *reduce* transmission) improved estimates of cost-effectiveness so that one-time routine HIV screening ratios remained below $50,000/QALY in settings where HIV prevalence was as low as 0.2%. One limitation is that the model incorporates effects on transmission directly from the screened individual to others, but not on more distant downstream transmission.

Sanders and colleagues focused on adults aged 55–75 years. In their Markov model incorporating prevalence, treatment, years and quality of life saved, transmission, and costs, they found that for older adults with a spouse or other sexual partner, one-time screening of populations with prevalence as low as 0.5% yielded favorable incremental cost-effectiveness in those 65 ($30,020/QALY) and those 75 years old ($41,520/QALY). Screening older adults without sexual partners was less cost-effective ($55,440/QALY for a 65 year old), but still approached the conventional threshold of $50,000/QALY.

Goldie and colleagues studied cost-effectiveness of HIV treatment in resource-limited settings. In Côte d'Ivoire, they compared treatment strategies not considered in developed settings, such as use of trimethoprim–sulfamethoxazole prophylaxis without antiretrovirals or prophylaxis plus antiretroviral therapy monitored with clinical criteria alone and not laboratory tests. Using conservative assumptions, they demonstrated that for all antiretroviral strategies, those combined with trimethoprim–sulfamethoxazole prophylaxis were in each case more effective (greater life expectancy) and more efficient (less incremental cost per year of life gained).

Prophylaxis alone was more efficient (incremental cost-effectiveness ratio of $240/year gained) than antiretroviral therapy and prophylaxis without CD4 testing ($620/year gained) or antiretrovirals and prophylaxis with CD4 testing ($1,180/year gained). However, prophylaxis without antiretrovirals was also less effective in terms of increased undiscounted life expectancy (1.6 months) than strategies that included antiretroviral therapy, either without CD4 testing (10.7–45.9 months depending on criteria for initiation of therapy) or with CD4 testing (14 months). The authors used three times the local per capita GDP ($2,124) as a cost-effectiveness threshold. Using this threshold, each strategy was economically attractive. The conclusions held up well to sensitivity analyses, with the exception of variation in the cost of antiretrovirals. Increases in antiretroviral costs had the greatest risk of driving costs per year of life saved significantly higher than the $2,124 GDP line. This study suggests that delivering HIV treatment can be economically reasonable in resource-limited settings.

SURVIVAL

Lohse, N, A Hansen, G Petersen, et al. Survival of persons with and without HIV infection in Denmark, 1995–2005. *Ann Intern Med* 2007; 146:87–95.

For most studies of antiretroviral therapy, the endpoint is HIV RNA suppression, yet the most important clinical indicator is impact on mortality rate. Lohse et al. use data from a population-based cohort study of all Danish residents to investigate survival in persons with HIV in the era of highly active antiretroviral therapy (HAART). Each of 3,990 HIV-infected persons living in Denmark was matched

with 95 controls from the general population based on age, sex, and geographic region. Patients were observed for an average of 5.8 and 8.4 person-years, respectively. Analyses were perfomed for three clinically relevant periods: 1995–1996 (pre-HAART), 1997–1999 (early HAART), and 2000–2005 (late HAART) and for both those with and without hepatitis C coinfection.

The study found that persons with HIV infection had a median survival beyond age 25 of 19.9 years, compared with 51.1 years in the general population. However, subgroup analysis found that in the late HAART era (2000–2005), survival beyond age 25 had increased to 32.5 years. Those patients who were hepatitis C negative and diagnosed in the late HAART era had the best survival at 38.9 years beyond age 25.

The strength of this study is the ability to capture accurate and comprehensive data for an entire population. There were no exclusion criteria, and all persons with HIV were included regardless of CD4 count, viral load, stage of disease, comorbidities, or treatment adherence. In addition, less than 3% of the cohort was lost to follow up. One potential limitation is that the measurement of survival from age 25 likely leads to lead time bias as most patients were diagnosed after age 40 and so presumably did not have disease at age 25. A second limitation is that the limited observation period may be too short to extrapolate long-term survival. Finally, because of Denmark's national health system, patients had excellent access to health care (>75% of patients were on HAART), and results may not be generalizable. Despite these limitations, this study predicts that a nearly 20-year increase in life expectancy is achievable with HAART.

MANAGEMENT

When to Start

In December 2007 the US Department of Health and Human Services (DHHS) revised its guidelines for initiation of antiretroviral therapy. The recommendation is to initiate treatment in patients with a history of an AIDS-defining illness or with CD4 lymphocyte count <350 cells/mm^3. This is a change from prior recommendations, which stated that asymptomatic persons with CD4 count between 200–350 cells/mm^3 should be offered treatment, but did not explicitly recommend treatment in this group.[9] The DHHS also recommends antiretroviral therapy regardless of CD4 lymphocyte count among (1) pregnant patients, (2) individuals with HIV-associated nephropathy, and (3) HIV/hepatitis B virus (HBV) co-infected persons, if treatment for HBV is indicated.[10]

New FDA Approved Medications

In the past several years the FDA has approved a number of new antiretroviral medications. One important development is the multi-class drug combining the nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) emtricitabine and tenofovir with the non-nucleoside reverse transcriptase inhibitor (NNRTI) efavirenz into a single pill. For appropriate patients, this pill comprises a complete antiretroviral regimen to be taken once daily.[11] Other developments include five new antiretrovirals and two new antiretroviral classes.[12–21] A detailed review of drug trials is beyond the scope of this review. However, a summary of selected major trials is presented in the Table 1.

Mallal, S, E Phillips, G Carosi, et al. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 2008;358:568–79.

Abacavir hypersensitivity is an immunologically based, potentially life-threatening condition, requiring a high degree of clinical suspicion.[22] It is estimated to affect 5–8% of patients during the first 6 weeks of therapy. Symptoms of abacavir hypersensitivity are nonspecific and often similar to other drug side effects, which may lead to false-positive diagnoses. Previous studies have shown an association of MHC class I allele HLA-B*5701 and hypersensitivity.[23,24]

This study tested the hypothesis that prospective HLA-B*5701 testing and excluding those found to be positive would reduce the incidence of abacavir hypersensitivity. This prospective, multicenter, randomized, double-blind study was performed at 265 centers in 19 countries. The study had two arms: (1) those receiving prospective screening, and if found to be HLA-B*5701 positive, excluded from abacavir-containing treatment and (2) a control group, receiving abacavir under usual care, with HLA test results reviewed after the study. Epicutaneous patch testing was used as a "gold standard" for hypersensitivity.

A total of 1,956 subjects was evaluated: 980 prospective screening and 847 control subjects. Fifty-five were excluded from the prospective screening group after testing positive for HLA-B*5701. Results showed that 109 patients out of 1,956 (5.6%) were HLA-B*5701 positive. The incidences of both clinical (0.40; 95% CI, 0.25–0.62) and immunologically confirmed (OR 0.03; 95% CI,0.00–0.18) hypersensitivity reactions were lower in the prospective screening group. In multivariate analysis, only prospective screening was a significant negative predictor of hypersensitivity reaction. There were no immunologically diagnosed hypersensitivity reactions in the screened group. Current use of a protease inhibitor (PI) or introduction of a new NNRTI were predictors of clinically diagnosed hypersensitivity, not immunologically confirmed, suggesting that the symptoms were due to another drug, and not true abacavir hypersensitivity. The positive predictive value (PPV) of HLA-B*5701 for immunologically confirmed hypersensitivity was 47.9%, with a negative predictive value (NPV) of 100%. For clinically diagnosed hypersensitivity, the PPV was 61.2% and the NPV was 95.5%. Nineteen of 49 carriers (38.8%) in the control group tolerated abacavir, suggesting that HLA B*5701 is necessary, but not sufficient for hypersensitivity.

There are a few caveats. Although it performed well in this study, the patch test is for experimental proposes only. The population studied was largely Caucasian; however, other studies suggest HLA-B*5701 testing performs well in diverse populations.[25] Cost-effectiveness may vary across different settings. Prospective HLA-B*5701 testing can reduce the risk of abacavir hypersensitivity reactions and is now recommended by the DHHS for all patients before beginning an abacavir-containing regimen.[26]

Table 1. Summary of Recently Approved Antiretroviral Medications

Medication	Trade name	Class	Selected clinical trials	Design, population	Outcome	Other notes
Efavirenz, emtricitabine, and tenofovir disoproxil fumarate	Atripla	Multi-class combination	—	—	—	-See Gallant et al. 2006 comparing NRTI backbones in efavirenz-containing regimen[1]
Etravirine (ETR)	Intelence	Nonnucleoside reverse transcriptase inhibitor (NNRTI)	*DUET-1[2] *DUET-2[3]	-Phase III, multinational, double-blind, placebo-controlled -Treatment experienced, with NNRTI resistance and at least 3 PI mutations -ETR + darunavir/ritonavir + investigator selected NRTIs with or without enfuvirtide	DUET-1 48 weeks: 60% on ETR vs. 39% on placebo achieved HIV RNA <50 copies/ml DUET-2 48 weeks: 61% on ETR vs. 41% on placebo achieved HIV RNA <50 copies	Generally well tolerated. Rash more common with ETR
Tipranavir (TPV)	Aptivus	Protease inhibitor (PI)	RESIST-1 RESIST-2 (pooled analysis)[4]	-Phase III, multinational, randomized, open-label of TPV vs. investigator selected control PI -Treatment-experienced (3-class, with at least 2 PI-based regimens), HIV RNA HIV RNA ≥1,000, genotypic evidence of some PI resistance	-48 weeks: 34% on TPV/r vs. 15% on control PIs achieved 1 $\log_{10}$ HIV RNA reduction -Time-to-treatment failure significantly longer in TPV/r group	-Gastrointestinal symptoms, elevated transaminases, cholesterol, and triglycerides more frequent in TPV/r -"Black-box" warning for intracranial hemorrhage in 2006
Darunavir (DRV)	Prezista	Protease inhibitor (PI)	ARTEMIS[5]	-Phase III, multinational randomized, open-label trial of ritonavir-boosted darunavir (DRV/r) vs. ritonavir-boosted lopinavir (LPV/r) -Treatment-naïve, and HIV RNA ≥5,000 copies/ml	-48 weeks: DRV/r was non-inferior to LPV/r at achieving HIV RNA <50 copies/ml at 48 weeks	Rates of adverse events lower in DRV/r
			POWER-1, POWER-2 (pooled analysis)[6]	-Phase IIb, multinational, open-label, randomized trials, DRV/r vs. other PIs -Treatment experienced, HIV RNA >1,000 copies/ml, at least one PI mutation, receiving a PI-containing regimen, history of PI, NRTI, and NNRTI use	48 weeks: 61% on DRV/r vs. 15% on control PIs achieved a 1 $\log_{10}$ HIV RNA reduction	Rates of adverse events were lower or comparable in the DRV/r vs. control groups
			TITAN[7]	-Phase III, multinational, open-label, randomized controlled trial of DRV/r vs. LPV/r in treatment-experienced, lopinavir-naïve persons	48 weeks: DRV/r was superior overall, at achieving HIV RNA <400 copies/ml -When analysis restricted to those with baseline LPV susceptibility, darunavir was non-inferior	More virologic failure and emergence of new resistance mutations in patients on LPV/r

(*continued on next page*)

Table 1. (continued)

Medication	Trade name	Class	Selected clinical trials	Design, population	Outcome	Other notes
Maraviroc (MVC)	Selzentry	Entry inhibitor (CCR5-co-receptor antagonist)	*MOTIVATE-1 MOTIVATE-2 (pooled analysis)[8]	-Phase III, multinational, randomized double-blind, placebo controlled trials -Three class resistance or treatment experience, R5-tropic virus, and HIV RNA ≥5,000 copies/ml -Optimized background regimen + either placebo, once daily MVC, or twice daily MVC	-48 weeks: both treatment groups showed significantly greater decreases in log_{10} transformed HIV RNA from baseline	-Safety and tolerability profile was similar across three groups -Role of MVC in treatment-naïve persons is less clear *(MERIT trial)[9]
Raltegravir (RAL)	Isentress	HIV integrase strand transfer inhibitors	Phase II[10]	-Phase II, multinational, randomized, double-blind, dose-ranging trial of RAL versus efavirenz -Treatment-naïve, HIV RNA ≥5,000 copies/ml, CD4 ≥100 cells/mm^3, no documented resistance to tenofovir, lamivudine, or efavirenz	-RAL at all doses performed similarly to efavirenz at viral suppression at 24 and 48 weeks -All RAL groups achieved HIV RNA under detectable levels more rapidly than efavirenz group	-RAL generally well-tolerated, with no dose-related toxicities seen
			BENCHMRK-1 BENCHMRK-2 (pooled analysis)[11]	-Phase III, multinational, randomized, placebo-controlled trial -Treatment experienced, HIV RNA >1,000 copies/ml and resistance to at least one medication from NRTI, NNRTI, and PI classes	48 weeks: 62% on RAL vs. 33% on placebo achieved HIV RNA <50 copies/ml	Safety and tolerability similar in both groups

Reference in abstract form

Abbreviations: DRV: darunavir; ETR: etravirine; LPV: lopinavir; MVC: maraviroc; NRTI: nucleoside reverse transcriptase inhibitor; NNRTI: non-nucleoside reverse transcriptase inhibitor; PI: protease inhibitor; r: ritonavir-boosted; RAL: raltegravir; TPR: tipranavir

[1]*Gallant et al. 2006*
[2]*Haubrich et al. 2008*
[3]*Johnson et al. 2008*
[4]*Hicks et al. 2006*
[5]*Ortiz et al. 2008*
[6]*Clotet et al. 2007*
[7]*Madruga et al. 2007*
[8]*Hardy et al. 2008*
[9]*Saag et al. 2007*
[10]*Markowitz et al. 2007*
[11]*Steigbigel et al. 2008*

HIV AND OTHER CHRONIC DISEASE

The DAD Study Group. Class of antiretroviral drugs and the risk of myocardial infarction. *N Engl J Med* 356;17, 1723–1735; 2007

Cardiovascular disease is emerging as an important cause of morbidity and mortality among persons living with HIV. There is a demonstrated association between combination antiretroviral therapy and the risk of myocardial infarction (MI).[27] The purpose of this paper was to investigate the association of exposure to PIs and NNRTIs and the risk of MI. The authors analyzed data collected from the Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) Study, an international collaboration including over 20,000 HIV-infected individuals in Australia, Europe, and the US.

This study population consisted of 23,437 HIV-infected patients. The authors calculated incidence rates of MI and determined the associations between MI and both PI and NNRTI exposure. Three hundred and forty-five patients had an MI during 94,469 person-years of observation. The incidence of MI increased from 1.53/1,000 person years in non-PI-exposed individuals to 6.01/1,000 person-years in individuals

exposed to PIs for greater than 6 years. After adjusting for exposure to other medication class and for cardiovascular risk factors other than lipids, the relative rate of MI per year of PI exposure was 1.16 (95% CI, 1.10–1.23). When additionally adjusted for lipid levels, it was 1.10 (95% CI, 1.04–1.18). The relative rate of MI per year of NNRTI exposure was 1.05 (95% CI, 0.98–1.13) and adjusting for lipid levels was 1.00 (95% CI, 0.93–1.09). While there was no evidence of this association for NNRTIs, the number of person-years of exposure to PIs was greater than that of person-years exposure to NNRTIs.

A limitation of this study is its observational nature; unmeasured confounders might contribute to the findings. Additionally, there is the possibility of "channeling bias" whereby patients considered to be at higher risk might be placed on regimens considered safer from a cardiac standpoint. Adjusting for cardiac risk factors ameliorates but cannot completely account for this. Still, the results of this study confirm a relationship between duration of combination antiretroviral use and cardiovascular disease, and suggest that PI use is associated with MI, which may be in part attributable to lipid effects.

The DAD Study Group. Use of nucleoside reverse transcriptase inhibitors and risk of myocardial infarction in HIV-infected patients enrolled in the D:A:D study: a multi-cohort collaboration. *Lancet*. 2008; 371:1417–1426.

This study, also from the D:A:D Study Group, focused on NRTIs and MI risk. It included 33,347 individuals encompassing 157,912 person-years of follow-up. Poisson regression was used to evaluate the relationship between MI and exposure to zidovudine, didanosine, stavudine, lamivudine, and abacavir. The investigators adjusted for demographic factors, calendar year, cohort, cardiovascular risk factors, and cumulative exposure to other antiretrovirals.

There were 517 patients with MI, corresponding to an event rate of 3.3 per 1,000 person-years (95% CI: 3.0–3.6). Exposure to abacavir or didanosine in the past 6 months, but not cumulative exposure, was associated with an increased risk of MI compared to those with no recent use of those medications. The relative rate of MI for recent abacavir use was 1.90 (95% CI: 1.47–2.45), and the relative rate of MI for recent didanosine use was 1.49 (1.14–1.95). There was no increased risk of MI among patients who had been previously on didanosine or abacavir and been off these medications for more than 6 months compared to those who had never received them. After adjustment for predicted 10 year risk of coronary disease,[28] recent use of abacavir and didanosine was still associated with increased rates of MI (1.89, 95% CI 1.47–2.45 for abacavir and 1.49, 95% CI 1.14–1.95 for didanosine). No association was observed between MI and cumulative exposure to zidovudine, stavudine, or lamivudine.

There are some limitations. The possibility exists that persons with higher cardiovascular risk due to other factors might have been more likely to receive abacavir or didanosine. However, after adjustment for risk factors, including predicted 10-year coronary risk, increased rates of MI persisted for abacavir and didanosine. Additionally, the risk decreased after the drugs were discontinued, suggesting the increased risk is attributable to the drugs. Data for two other commonly prescribed NRTIs, emtricitabine and tenofovir, were not presented in this analysis due to lack of sufficient follow-up time.

The findings of this study merit additional investigation, which is underway. At this time, the DHHS has not changed their recommendations on the use of antiviral agents.[29] Clinicians should take into consideration individual patient characteristics and consider all available treatment options when deciding on the appropriate antiretroviral regimen for their patients.

Patel, P., D. Hanson, P. Sullivan, et al. Incidence of types of cancer among HIV-infected persons compared with the general population in the United States, 1992–2003. Ann Intern Med. 2008; 148:728–36.

HIV-infected individuals may be at an increased risk for certain cancers. Few studies have examined cancer incidence at different stages of HIV infection or after the introduction of HAART. This study compared the incidence of non-AIDS defining cancers among HIV-infected persons at all stages of infection and during different periods of HAART availability, compared with incidence in the general population. Additionally, the authors explored risk factors for certain cancers in HIV-infected persons.

The authors compared the data from 54,780 HIV-infected persons [47,832 from the Adult and Adolescent Spectrum of HIV Disease (ASD) Project and 6,948 from the HIV Outpatient Study (HOPS)] who contributed 157,819 person-years of observation from 1992 to 2003 with 334,802,121 person-years of the general population from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. They determined standardized rate ratios (SRRs) for each and found the incidence for the following non-AIDS-defining cancers to be significantly higher in the HIV-infected population: anal (SRR 42.9; 95% CI, 34.1 to 53.3), vaginal (SRR 21.0; 95% CI, 11.2 to 35.9), Hodgkin lymphoma (SRR 14.7; 95% CI, 11.6 to 18.2), liver (SRR 7.7; 95% CI, 5.7 to 10.1), lung (3.3; 95% CI, 2.8 to 3.9), melanoma (SRR 2.6; 95% CI, 1.9 to 3.6), oropharyngeal (SRR 2.6; 95% CI, 1.9 to 3.4), leukemia (SRR 2.5; 95% CI, 1.6 to 3.8), colorectal (SRR 2.3; 95% CI, 1.8 to 2.9), and renal (SRR 1.8; 95% CI, 1.1 to 2.7). Among HIV-infected persons, incidence rates for melanoma, Hodgkin lymphoma, and colorectal, anal, and prostate cancer increased significantly over time, despite the introduction of HAART during the study period. However, anal cancer was the only type of cancer for which the *relative* incidence for HIV-infected persons compared with the general population increased over time.

In terms of risk factors, acquisition of HIV through male-male sex was associated with increased risk for Kaposi's sarcoma (KS) and non-Hodgkin lymphoma. Use of antiretroviral therapy was associated with *decreased* risk for KS, non-Hodgkin lymphoma, and cervical, lung, breast, and colorectal cancer; low CD4 nadir was associated with *increased* risk of KS, non-Hodgkin lymphoma, cervical, anal, colorectal, and lung cancers. Hepatitis B or C coinfection was associated with increased risk of liver cancer.

Limitations include (1) no formal evaluation of cancer data in these cohorts, (2) inadequate tobacco data, and (3) the cohorts are not representative of the overall HIV-infected population in the US. Still, the study suggests that HIV-infected

individuals are at an increased risk for certain cancers compared to the general population. Risks for some of these cancers may be reduced with currently available prevention strategies, and further work is needed to explore new strategies.

Acknowlegements: *Dr. Gifford is supported in part by grants from the Department of Veteran Affairs QUERI-HIV/Hepatitis Program (no. HIV 98–000), the VA HSR&D Center of Excellence, Bedford (no. HFP 90–011), and the NIMH HIV Secondary Prevention and Translation Branch, (no. 1 R01 MH076911–01A2). Dr. Sullivan is a Robert Wood Johnson Physician Faculty Scholar and is supported, in part, by the Health Resources Services Administration (1H97HA03800).*

The authors would like to thank Dr. Joel Gallant for assisting with selection of medication trials to include in this update and for his review and comments on the Table 1.

This update was presented at the meeting of the Society of General Internal Medicine in April 2008.

Conflict of Interest: *None disclosed.*

Corresponding Author: *Amina A. Chaudhry, MD MPH; Division of General Internal Medicine, Johns Hopkins University School of Medicine, 1830 East Monument Street, Room 8047A, Baltimore, MD 21287, USA (e-mail: amina.chaudhry@jhu.edu).*

REFERENCES

1. **Siegfried N, Muller M, Deeks J, et al.** HIV and male circumcision–a systematic review with assessment of the quality of studies. Lancet Infect Dis. 2005;5(3):165–73. Mar.
2. **Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A.** Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: The ANRS 1265 trial. PLoS Med. 2005;211:e298. Nov.
3. **Bailey RC, Moses S, Parker CB, et al.** Male circumcision for HIV prevention in young men in Kisumu, Kenya: A randomised controlled trial. Lancet. 2007;369:643–56.
4. **Wawer M, Kigozi G, Serwadda D, et al.** Trial of male circumcision in HIV$^+$ men, Rakai, Uganda: Effects in HIV$^+$ men and in women partners. Presented at the 15th Conference on Retroviruses and Opportunistic Infections, Boston, MA, USA. February 3–6, 2008. Abstract 33LB.
5. Joint United Nations Programme on HIV/AIDS, World Health Organization. New Data on Male Circumcision and HIV Prevention: Policy and Programme Implications. 2007. Available at: http://data.unaids.org/pub/Report/2007/mc_recommendations_en.pdf. Accessed September 29, 2008.
6. **Branson BM, Handsfield HH, Lampe MA, et al.** Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. MMWR. Centers for Disease Control. 2006;55RR-141–17.
7. **Paltiel AD, Weinstein MC, Kimmel AD, et al.** Expanded screening for HIV in the United States-an analysis of cost-effectiveness. N Engl J Med. 2005;352:586–95.
8. **Sanders GD, Bayoumi AM, Sundaram V, et al.** Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. N Engl J Med. 2005;352:570–85.
9. Panel on Antiretroviral Guidelines for Adult and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. 2006 October 10, 2006. Available at: http://aidsinfo.nih.gov/contentfiles/AdultandAdolescentGL000629.pdf. Accessed September 29, 2008.
10. Panel on Antiretroviral Guidelines for Adult and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. 2007 December 1, 2007. Available at: http://aidsinfo.nih.gov/contentfiles/AdultandAdolescentGL000721.pdf. Accessed September 29, 2008.
11. **Gallant JE, DeJesus E, Arribas JR, et al.** Tenofovir DF, emtricitabine, and efavirenz vs. zidovudine, lamivudine, and efavirenz for HIV. N Engl J Med. 2006;354:251–60. Jan 19.
12. **Haubrich R, Cahn P, Grinsztejn B, et al.** DUET-1: Week-48 results of a phase III randomized double-blind trial to evaluate the efficacy and safety of TMC125 vs placebo in 612 treatment-experienced HIV-1-infected patients. Presented at the 15th Conference on Retroviruses and Opportunistic Infections, Boston, MA, USA. February 3–6, 2008. Abstract 790.
13. **Johnson M, Campbell T, Clotet B, et al.** DUET-2: Week-48 results of a phase III randomized double-blind trial to evaluate the efficacy and safety of TMC125 vs placebo in 591 treatment-experienced HIV-1-infected patients. Presented at the 15th Conference on Retroviruses and Opportunistic Infections, Boston, MA, USA. February 3–6, 2008. Abstract 791.
14. **Hicks CB, Cahn P, Cooper DA, et al.** Durable efficacy of tipranavir-ritonavir in combination with an optimised background regimen of antiretroviral drugs for treatment-experienced HIV-1-infected patients at 48 weeks in the randomized evaluation of strategic intervention in multi-drug resistant patients with tipranavir (RESIST) studies: An analysis of combined data from two randomised open-label trials. Lancet. 2006;368:466–75.
15. **Ortiz R, Dejesus E, Khanlou H, et al.** Efficacy and safety of once-daily darunavir/ritonavir versus lopinavir/ritonavir in treatment-naive HIV-1-infected patients at week 48. AIDS. 2008;22:1389–97.
16. **Clotet B, Bellos N, Molina JM, et al.** Efficacy and safety of darunavir-ritonavir at week 48 in treatment-experienced patients with HIV-1 infection in POWER 1 and 2: A pooled subgroup analysis of data from two randomised trials. Lancet. 2007;369:1169–78.
17. **Madruga JV, Berger D, McMurchie M, et al.** Efficacy and safety of darunavir-ritonavir compared with that of lopinavir-ritonavir at 48 weeks in treatment-experienced, HIV-infected patients in TITAN: A randomised controlled phase III trial. Lancet. 2007;370:49–58.
18. **Hardy D, Reynes J, Kornourina I, et al.** Efficacy and safety of maraviroc plus optimized background therapy in treatment-experienced patients infected with CCR5-tropic HIV-1: 48-week combined analysis of the MOTIVATE studies. Presented at the 15th Conference on Retroviruses and Opportunistic Infections, Boston, MA, USA. February 3–6, 2008. Abstract 792.
19. **Saag M, Ive P, Heera J, et al.** A multicenter, randomized, double-blind, comparative trial of a novel CCR5 antagonist, maraviroc versus efavirenz, both in combination with Combivir (zidovudine/lamivudine), for the treatment of antiretroviral naive subjects infected with R5 HIV-1: Week 48 results of the MERIT study. Presented at the 4th International AIDS Society Conference on HIV Pathogenesis, Treatment and Prevention, Sydney, Australia. July 22–25, 2007. Abstract WESS104.
20. **Markowitz M, Nguyen BY, Gotuzzo E, et al.** Rapid and durable antiretroviral effect of the HIV-1 integrase inhibitor raltegravir as part of combination therapy in treatment-naive patients with HIV-1 infection: Results of a 48-week controlled study. J Acquir Immune Defic Syndr. 2007;46:125–33.
21. **Steigbigel RT, Cooper DA, Kumar PN, et al.** Raltegravir with optimized background therapy for resistant HIV-1 infection. N Engl J Med. 2008;359:339–54.
22. **Hetherington S, McGuirk S, Powell G, et al.** Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. Clin Ther. 2001;10:1603–14.
23. **Hetherington S, Hughes AR, Mosteller M, et al.** Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. Lancet. 2002;359:1121–2.
24. **Mallal S, Nolan D, Witt C, et al.** Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. Lancet. 2002;359:727–32.
25. **Saag M, Balu R, Phillips E, et al.** High sensitivity of human leukocyte antigen-b*5701 as a marker for immunologically confirmed abacavir hypersensitivity in white and black patients. Clin Infect Dis. 2008;46:1111–8.
26. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Available at: http://aidsinfo.nih.gov/contentfiles/AdultandAdolescentGL.pdf. Accessed September 29, 2008.
27. **Lundgren JD.** Combination antiretroviral therapy and the risk of myocardial infarction: The data collection on adverse events of anti-HIV drugs (DAD) study group. N Engl J Med. 2003;349:1993–2003.
28. **Anderson KM, Wilson PWF, Odell PM, Kannel WB.** An updated coronary risk profile. A statement for health professionals. Circulation. 1991;83:356–62.
29. Panel on Antiretroviral Guidelines for Adults and Adolescents. DHHS adults and adolescents antiretroviral treatment guidelines Panel's communication regarding abacavir. Available at: http://aidsinfo.nih.gov/contentfiles/ABCComm.pdf. Accessed September 29, 2008.

ELSEVIER

Alcohol 43 (2009) 65–71

ALCOHOL

Alcohol consumption and lipodystrophy in HIV-infected adults with alcohol problems

Debbie M. Cheng[a,e,*], Howard Libman[b], Carly Bridden[a], Richard Saitz[a,c], Jeffrey H. Samet[a,d]

[a]*Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA USA;*
[b]*Divisions of General Medicine and Primary Care and Infectious Diseases, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA USA;*
[c]*Department of Epidemiology and Youth Alcohol Prevention Center,*
[d]*Department of Social and Behavioral Sciences, and* [e]*Department of Biostatistics, Boston University School of Public Health, Boston University Medical Campus, 801 Massachusetts Avenue, 3rd Floor, Boston, MA 02118 USA*

Received 5 June 2008; received in revised form 19 September 2008; accepted 22 September 2008

Abstract

Lipodystrophy is a common long-term complication of HIV infection that may lead to decreased quality of life and less adherence to antiretroviral therapy (ART). A complete understanding of the etiology of HIV-associated lipodystrophy has not yet been achieved, although factors related to the virus, per se, and use of ART appear to be related. Alcohol use is common among HIV-infected patients and has biological effects on fat distribution, yet alcohol's relationship to HIV-associated lipodystrophy has not been examined. The goal of this clinical study was to assess the effect of alcohol consumption on lipodystrophy in HIV-infected adults with alcohol problems. This was a prospective study (2001–2006) of 289 HIV-infected persons with alcohol problems. The primary outcome was self-reported lipodystrophy, which was assessed at one time point (median 29 months after enrollment). Alcohol use was assessed every 6 months and classified as: abstinent at all interviews; ≥1 report of moderate drinking but no heavy drinking; 1 or 2 reports of heavy drinking; or ≥3 reports of heavy drinking. Multivariable logistic regression models were fit to the data. Fifty-two percent (150/289) of subjects reported lipodystrophy. Alcohol consumption was: 34% abstinent at all interviews; 12% ≥1 report of moderate drinking, but no heavy drinking; 34% 1–2 reports of heavy drinking; and 20% ≥3 reports of heavy drinking. Although not statistically significant, subjects with alcohol use had a higher odds of lipodystrophy (adjusted odds ratios and 95% confidence interval: ≥1 report of moderate drinking, 2.36 [0.89, 6.24]; 1–2 reports of heavy drinking, 1.34 [0.69, 2.60]; ≥3 reports of heavy drinking, 2.07 [0.90, 4.73]). Alcohol use may increase the odds of developing HIV-associated lipodystrophy among subjects with alcohol problems. However, larger studies are needed to fully elucidate the role and impact of alcohol consumption on the development of this common long-term complication of HIV infection and its treatment.

Keywords: Lipodystrophy; HIV; Alcohol consumption

Introduction

Despite dramatic improvements in morbidity and mortality of HIV-infected patients on antiretroviral therapy (ART), several important long-term complications of the disease and its treatment, including lipodystrophy syndrome, have been described. HIV-associated lipodystrophy (abnormal body fat distribution, often with hyperlipidemia and/or glucose intolerance) is common, with a prevalence ranging from 20% to 80% in patients on ART (Sattler, 2003). Between 20% and 50% of HIV-infected patients will report at least one sign of lipodystrophy within the first 2 years of starting ART (Galli et al., 2002; Heath et al., 2002). The abnormal fat distribution is typically characterized by fat accumulation in the neck, breasts, and abdomen, and fat loss in the face, buttocks, and extremities. Hyperlipidemia and glucose intolerance may contribute to the risk for cardiovascular and other atherosclerotic disease as patients age. In addition, patient's quality of life and adherence to ART may be adversely affected by lipodystrophy through decreased self-esteem and stigmatization (Blanch et al., 2004).

* Corresponding author. Tel.: +1-617-638-5989; fax: +1-617-638-6484.
E-mail address: dmcheng@bu.edu (D.M. Cheng).

0741-8329/09/$ – see front matter
doi: 10.1016/j.alcohol.2008.09.004

The cause of HIV-associated lipodystrophy is unclear, but it appears to be multifactorial in origin, involving host factors (Martinez et al., 2001b), HIV itself (Mujawar et al., 2006), and ART drug effects. The latter are particularly notable with the use of the nucleoside analogue reverse transcriptase inhibitors (NRTIs) and protease inhibitors (Caron et al., 2001). ART may have an impact on liver lipid transport, glucose homeostasis (Ben-Romano et al., 2003), and adipose tissue directly (Caron et al., 2001). The NRTIs specifically affect mitochondrial DNA replication (McComsey and Walker, 2004), which may explain the prominence of abnormal fat distribution in this syndrome because of the central role of mitochondria in fat cells and the high content of mitochondria in brown fat in particular (Brinkman et al., 1999; Villarroya et al., 2007).

Alcohol use is common among HIV-infected patients, and there are reasons to hypothesize that it may affect the development of lipodystrophy. Alcohol has multiple effects on metabolism, including changes in energy intake (Armellini et al., 1993). An intriguing link can be found between alcohol use and mitochondrial DNA replication in an uncommon disease referred to as multiple symmetrical lipomatosis (MSL). First described by Madelung in 1888 as a case series of lipomas in patients with chronic alcoholism, MSL is characterized by nonencapsulated lipomas in a symmetrical distribution, typically in the subcutaneous fat of the back of the neck and the proximal legs and arms (Brinkman et al., 1999; Enzi et al., 1977), with sparing or wasting of fat in the distal arms and legs (Coin et al., 2005). In a more recent case series, alcohol intake of over 50 mL per day was observed in 29 of 31 patients with MSL (Enzi et al., 2002). However, it is now known that many patients with MSL also have either mutations in mitochondrial DNA or abnormal mitochondrial function (Klopstock et al., 1997; Schoffer and Grant, 2006), which leads to the hypothesis that heavy alcohol use in people with reduced mitochondrial function produces an increased risk of abnormal fat growth. In addition, several investigators have suggested that HIV-associated lipodystrophy itself resembles MSL because of the similar patterns of abnormal fat distribution (Brinkman et al., 1999; Hengel et al., 1997; Teplitsky and Halabe, 1999), and Brinkman et al. (Brinkman et al., 1999) have hypothesized that mitochondrial toxicity underlies both syndromes. This association between alcohol use and MSL suggests the need to investigate whether heavy alcohol consumption contributes to the development of HIV-associated lipodystrophy.

The goal of this study was to assess the association between alcohol consumption and lipodystrophy in HIV-infected adults with alcohol problems. The relationship of these two common conditions has received little attention in this patient population. We examined this issue using data from a prospective cohort of HIV-infected persons with current or past alcohol problems.

Materials and methods

Study design and participant recruitment

Subjects were participants in HIV-Longitudinal Interrelationships of Viruses and Ethanol (HIV-LIVE), a prospective, observational cohort study of HIV-infected patients with current or past alcohol problems. Data were collected at baseline and every 6 months thereafter for up to 42 months.

Four hundred subjects were recruited from the following sources: (1) 38% ($n = 154$) from the HIV-Alcohol Longitudinal Cohort (HIV-ALC) (HIV-LIVE) study, a previous cohort study at Boston Medical Center (BMC) with identical inclusion and exclusion criteria (Samet et al., 2004); (2) 22% ($n = 88$) from the Diagnostic Evaluation Unit (Samet et al., 1995), an intake clinic for HIV-infected patients at BMC; (3) 8% ($n = 31$) from the HIV Primary Care and Specialty Clinics at Beth Israel Deaconess Medical Center (BIDMC); and (4) 32% ($n = 127$) through flyers distributed in health-care centers, homeless shelters, and drug treatment programs; advertisements in newspapers; and referrals from other HIV-LIVE subjects.

Eligibility criteria were: (1) documented HIV antibody test by enzyme-linked immunosorbent assay confirmed by Western blot; (2) ≥ 2 affirmative responses to the CAGE alcohol screening questionnaire (Buchsbaum et al., 1991; Mayfield et al., 1974) or diagnosis of lifetime alcohol abuse or dependence based on a study physician investigator clinical assessment; (3) ability to speak English or Spanish; and (4) at least one contact person. Exclusion criteria were: (1) score of <21 on the 30-item Folstein Mini-Mental State Examination (Folstein et al., 1975; Smith et al., 2006), or (2) trained interviewer assessment that the patient could not comprehend informed consent or answer the interview questions. In addition to the exclusion criteria noted for the HIV-LIVE study, the present analyses also excluded patients reporting moderate or severe isolated abdominal obesity to avoid confounding by age-related central adiposity. Enrollment began August 2001 and ended July 2003. Eligible subjects who wished to participate provided written informed consent prior to enrollment. Most interviews took place at General Clinical Research Centers. The institutional review boards of BMC and BIDMC approved this study. Additional privacy protection was secured with a Certificate of Confidentiality from the Department of Health and Human Services to protect subjects from release of research data under court order or subpoena.

Subject assessment

Subjects received an interviewer-administered assessment at baseline and 6-month intervals, conducted in English or Spanish, including questions on demographics, HIV risk behaviors, alcohol consumption, and ART use in the past 30 days. Lipodystrophy was evaluated at a single

follow-up visit using a questionnaire developed by Carr et al. (Carr et al., 2003). These questions supplemented the standard assessment and were instituted in 2004. Due to the sequential study accrual, subjects were administered the supplemental assessment at different follow-up times. Fat loss and/or fat gain in the face, front, or sides of neck, back or base of neck, arms, breasts, waist, buttocks, and legs were assessed. The severity of any reported fat gain/loss was rated as mild, moderate, or severe. The presence and location of any fat lumps was also evaluated by research associate–administered subject questionnaire. Past month alcohol consumption was assessed using a validated calendar method (Sobell and Sobell, 1996). The Composite International Diagnostic Interview Alcohol Module (Robins et al., 1988) was administered following study enrollment to determine current (past 6 months) and lifetime diagnoses of alcohol abuse and dependence.

We recorded CD4 cell counts and HIV RNA levels at each interview. Values were obtained by phlebotomy if not available from clinical records within 4 months of the interview. Hepatitis C virus (HCV) RNA was measured using commercially available assays, either by branched-chain DNA or polymerase chain reaction (PCR)-based assays. Depressive symptoms were measured using the Center for Epidemiologic Studies Depression Scale (CES-D); a cut-off of ≥23 was used to denote substantial depressive symptoms in persons with chronic diseases (Cook et al., 2002). ART use was assessed with the question, ''Have you ever taken antiviral medications for your HIV?''

Outcomes

The primary outcome was self-reported lipodystrophy since diagnosis of HIV infection. It was defined as a response of moderate or severe changes in any of the following parameters: fat loss in the face or arms; fat gain in the back or base of the neck; increase in breast size; decrease in fat on buttocks or legs; or lipomatosis. The presence of lipodystrophy was based on an objective case definition previously developed and validated by Carr et al. (2003) but excluded clinical characteristics not generally recognized as manifestations of HIV-associated lipodystrophy (i.e., fat gain in face, fat loss/gain in the front or sides of neck, fat loss in the back or base of neck, fat gain in arms, decrease in breast size, increase in fat on buttocks or legs, and decrease in waist size). Consistent with the definition by Carr et al., patients reporting ''moderate or severe isolated abdominal obesity'' were excluded from analyses to avoid confounding by age-related central adiposity. Three secondary outcomes were also examined: lipohypertrophy (defined as a response of moderate or severe fat gain in the back or base of the neck or increase in size of breast); lipoatrophy (defined as a response of moderate or severe fat loss in the face or arms, or decrease in fat on buttocks or legs); and reporting both fat gain (i.e., lipohypertrophy or lipomatosis) and fat loss (i.e., lipoatrophy) since the diagnosis of HIV infection. The last outcome is a more stringent definition of lipodystrophy as it requires a report of both fat gain and fat loss, whereas the primary definition requires a report of either fat gain or fat loss.

Primary independent variable

At each study interview, alcohol use in the past 30 days was classified as heavy, moderate, or abstinent. Heavy alcohol use was defined as >14 drinks/week or ≥5 drinks on one occasion for men <66 years old, and >7 drinks/week or ≥4 drinks on one occasion for men ≥66 years old and all women (National Institute on Alcohol Abuse and Alcoholism, 2005). Moderate alcohol use was defined as any drinking less than heavy amounts. The main independent variable was alcohol consumption, measured across all available study interviews between study enrollment and the follow-up visit when lipodystrophy was assessed. The longitudinal information collected on alcohol use was combined, and subjects were classified into one of the following four alcohol consumption categories: abstinent at all available interviews; at least one report of moderate drinking but no heavy drinking; one or two interviews in which heavy drinking was reported; and three or more interviews in which heavy drinking was reported. Based on the definition of the highest drinking category, subjects who did not complete at least three study interviews were excluded from analyses.

Potential confounding factors

The following potential confounders were included in the analyses: gender; age; race/ethnicity (black, white, other); hepatitis C RNA status (positive vs. negative); current depressive symptoms (yes vs. no); ART use (ever vs. never); cocaine use past 6 months (yes vs. no); CD4 cell count; and lifetime alcohol dependence diagnosis (no diagnosis, abuse, dependence). Covariate values for depressive symptoms, CD4 cell count, and cocaine use were taken from the follow-up interview at which lipodystrophy was assessed. All other covariates were taken from study enrollment.

Statistical analysis

Descriptive statistics were used to assess the bivariate relationship between subject characteristics and lipodystrophy for the study sample at enrollment. Chi-square and Fisher's exact tests were used as appropriate to assess the bivariate associations. We examined the relationship between alcohol consumption and lipodystrophy by fitting multivariable logistic regression models. Unadjusted and adjusted models controlling for all potential confounding factors were fit to the data. Analyses of all secondary outcomes were also conducted using logistic regression models. To minimize the potential for collinearity, we assessed correlation between pairs of independent variables

and verified that no pair of variables included in the same regression model was highly correlated (i.e., $r > 0.40$). 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).

Sample size considerations

Among the 289 study subjects, 99 reported past month abstinence at all study visits, 34 had at least one report of past month moderate drinking but no heavy drinking, 98 had one to two reports of heavy drinking during the past month, and 58 had three or more such reports of heavy drinking. Consistent with the observed proportion of 50% reporting lipodystrophy among the reference group of those abstinent at all visits, the available sample sizes allow our study 80% power to detect an odds ratio as small as 2.8 for ≥3 episodes of heavy drinking, 2.4 for 1–2 episodes of heavy drinking, and 3.7 for moderate but no heavy drinking. Power calculations were based on χ^2 tests using a significance level of 0.05.

Results

Of the 400 HIV-LIVE subjects, 326 (82%) completed the follow-up study interview in which lipodystrophy was assessed. Of the 326 subjects who completed the lipodystrophy questionnaire, 30 were excluded because of isolated abdominal obesity and seven were excluded for having fewer than three study visits at the time lipodystrophy was assessed. This resulted in a final study sample of 289 HIV-infected subjects with current or past of alcohol problems. Among the study sample, the proportion that was assessed at the baseline, 6-, 12-, 18-, 24-, 30-, 36-, and 42-month study visits was 100%, 89%, 88%, 89%, 88%, 89%, 77%, and 60%, respectively.

Characteristics of the study sample at study enrollment are shown in Table 1. The majority were male (74%) and non-white (68%). The age range of the subjects was 20.9–70.7 years, with a mean of 42.9. At study enrollment, 46% reported cocaine use in the past 12 months, 48% had substantial current depressive symptoms (CES-D ≥ 23), 50% were hepatitis C RNA positive, and 87% had ever taken ART. CD4 cell count ranged from 8 to 1809 cells/mm^3, with a median of 404. The proportions in each alcohol consumption category were as follows: 34% were abstinent at all interviews; 12% had at least one report of past month moderate drinking, but no heavy drinking; 34% had 1–2 reports of past month heavy drinking; 20% had ≥3 reports of past month heavy drinking. Among the 289 study subjects, 150 (52%) had self-reported lipodystrophy, 115 (40%) lipoatrophy, 48 (17%) lipohypertrophy, and 42 (15%) both fat gain and loss. The median follow-up time at which lipodystrophy was assessed was 29 months after study enrollment (interquartile range 20–31 months). The mean and median number of interviews that occurred between enrollment and the time that lipodystrophy was assessed was 5 (interquartile range 4–6 interviews) and the mean and median number of interviews that occurred overall (i.e., including interviews that occurred following assessment of lipodystrophy) was 7 (interquartile range 6–8 interviews).

Table 1
Characteristics of the subjects at study enrollment and bivariate associations with lipodystrophy ($N = 289$)

Variable	Number (%)	Number (%) reporting lipodystrophy by characteristic
Alcohol consumption[a]		
Abstinent at all interviews	99 (34%)	50 (51%)
≥1 Report of moderate drinking	34 (12%)	18 (53%)
1–2 Reports of heavy drinking	98 (34%)	49 (50%)
≥3 Reports of heavy drinking	58 (20%)	33 (57%)
Gender		
Male	214 (74%)	104 (49%)
Female	75 (26%)	46 (61%)
Hepatitis C RNA		
Positive	143 (49%)	79 (55%)
Negative	145 (51%)	70 (48%)
Center for Epidemiologic Studies Depression Scale*		
<23	151 (52%)	70 (46%)
≥23	138 (48%)	80 (58%)
Race		
White	92 (32%)	55 (60%)
Black	138 (48%)	69 (50%)
Other	59 (20%)	26 (44%)
Antiretroviral therapy use*		
Never	37 (13%)	12 (32%)
Ever	251 (87%)	137 (55%)
Recent cocaine use		
No	156 (54%)	83 (53%)
Yes	133 (46%)	67 (50%)
Age (years)[b]		
21–40	117 (40%)	59 (50%)
41–48	101 (35%)	46 (46%)
49–71	71 (25%)	45 (63%)
CD4 cell count[b,c]		
8–277	77 (28%)	37 (48%)
278–518	98 (36%)	52 (53%)
519–1,809	96 (35%)	52 (54%)
Lifetime alcohol diagnosis		
No diagnosis	37 (13%)	19 (51%)
Abuse	55 (19%)	25 (45%)
Dependence	193 (68%)	104 (54%)
Recent alcohol diagnosis		
No diagnosis	249 (87%)	131 (53%)
Abuse	7 (2%)	3 (43%)
Dependence	29 (10%)	14 (48%)

*$P < .05$ for bivariate association with lipodystrophy.
[a]Report is based on assessment of alcohol use in the past month at each research interview.
[b]Based on tertiles from study sample.
[c]$n = 271$.

In multiple logistic regression models controlling for potential confounders, the adjusted odds ratios (AORs) and 95% confidence intervals (CIs) for lipodystrophy were 2.07 (0.90, 4.73) for ≥3 reports of any past month heavy drinking; 1.34 (0.69, 2.60) for 1–2 reports of any past month heavy drinking; and 2.36 (0.89, 6.24) for any past month moderate but no heavy drinking (Table 2). Factors significantly associated with lipodystrophy were substantial depressive symptoms (AOR 2.32 [95% CI: 1.30, 4.15] for CES-D ≥ 23 vs. CES-D < 23), ART use ever (AOR 2.64 [95% CI: 1.10, 6.36]), and CD4 cell count (AOR 1.13 [95% CI: 1.02, 1.25] per 100 unit increase). Secondary analyses excluding depressive symptoms as a covariate produced similar results (AOR [95% CI]: ≥1 report of moderate drinking, 2.23 [0.86, 5.78]; 1–2 reports of heavy drinking, 1.46 [0.76, 2.80]; and ≥3 reports of heavy drinking, 2.14 [0.95, 4.82]). In secondary analyses of the effect of any reports of past month drinking (yes vs. no), the AOR and 95% CI for lipodystrophy was 1.63 (0.89, 3.22).

Table 2
Adjusted odds ratios for lipodystrophy based on multivariable logistic regression analyses of 289 HIV-infected subjects with current or past alcohol problems

Variable	Adjusted odds ratio (95% confidence interval)
Alcohol consumption[a]	
Abstinent at all interviews	1.00
≥1 Report of moderate drinking	2.36 (0.89, 6.24)
1–2 Reports of heavy drinking	1.34 (0.69, 2.60)
≥3 Reports of heavy drinking	2.07 (0.90, 4.73)
Gender	
Male	1.00
Female	1.83 (0.95, 3.53)
Hepatitis C RNA	
Positive	1.00
Negative	0.90 (0.50, 1.62)
Center for Epidemiologic Studies Depression Scale	
<23	1.00
≥23	2.32 (1.30, 4.15)*
Race	
White	1.00
Black	0.64 (0.34, 1.19)
Other	0.60 (0.28, 1.29)
Antiretroviral therapy use*	
Never	1.00
Ever	2.64 (1.10, 6.36)*
Recent cocaine use	
No	1.00
Yes	1.36 (0.69, 2.68)
Age (per 10 year increase)	1.42 (0.95, 2.13)
CD4 Count (per 100 cells/mm^3 increase)	1.13 (1.02, 1.25)*
Lifetime alcohol diagnosis	
No diagnosis	1.00
Abuse	0.42 (0.16, 1.10)
Dependence	0.72 (0.32, 1.61)

*$P < .05$.

[a]Report is based on assessment of alcohol use in the past month at each research interview.

The estimated associations between alcohol consumption and each of the secondary outcomes are reported in Table 3. For lipohypertrophy, subjects in each drinking category had a higher odds of reporting lipohypertrophy compared to those who were abstinent at all interviews (AOR [95% CI]: 2.51 [0.82, 7.66] for ≥3 reports of heavy drinking; 1.31 [0.53, 3.25] for 1–2 reports of heavy drinking; and 3.52 [1.07, 11.54] for moderate but no heavy drinking), with the last comparison being statistically significant. For lipoatrophy, the AORs [95% CI] were 1.77 [0.76, 4.12] for ≥3 reports of heavy drinking; 0.92 [0.46, 1.84] for 1–2 reports of heavy drinking; and 1.15 [0.44, 3.01] for moderate but no heavy drinking. For reporting both fat gain and fat loss, subjects with 1–2 reports of heavy drinking had increased odds of the outcome, but no association was statistically significant.

Discussion

Lipodystrophy syndrome is common in HIV-infected patients, is multifactorial in origin, and is associated with decreased self-esteem, stigmatization, and possibly an increased risk of cardiovascular and other atherosclerotic disease. Alcohol use is also frequent among HIV-infected persons. Based on the link between chronic alcoholism and the development of abnormal fat growth in patients with mitochondrial replication deficits, as seen in MSL, we hypothesized that alcohol use increases the risk of HIV-associated lipodystrophy. Exploration of this association has not been an active area of clinical HIV investigation. Our results suggest the potential of a clinically important relationship between alcohol and lipodystrophy among a cohort of HIV-infected subjects with alcohol problems. Subjects with multiple reports of heavy drinking had twice the odds of reporting lipodystrophy compared to those reporting abstinence, although these findings were not statistically significant. Alcohol consumption may be associated with the secondary outcome lipohypertrophy, where a statistically significant result was observed for moderate drinking compared to abstinence.

In the present study, alcohol use was prospectively examined at 6-month intervals using a comprehensive, validated instrument. In addition, we used a definition for lipodystrophy based on the objective case definition validated by Carr et al. (2003). Several factors that may be associated with HIV-associated lipodystrophy, such as ART use, depressive symptoms, and cocaine use, were accounted for in the regression analyses to minimize the potential for confounding. The findings that both ART use and depressive symptoms were associated with lipodystrophy provide face validity for the cohort and the definition used to examine the association between alcohol use and lipodystrophy. These findings are consistent with the medical literature that suggests that ART use likely plays

Table 3
Adjusted odds ratios (AORs) for the secondary outcomes lipohypertrophy, lipoatrophy, and report of both fat gain and fat loss

Alcohol consumption[a]	AOR[b] (95% confidence interval)		
	Lipohypertrophy	Lipoatrophy	Report of both fat gain and fat loss
Abstinent at all interviews	1.00	1.00	1.00
≥1 Report of moderate drinking	**3.52 (1.07, 11.54)**	1.15 (0.44, 3.01)	0.87 (0.25, 2.99)
1–2 Reports of heavy drinking	1.31 (0.53, 3.25)	0.92 (0.46, 1.84)	1.38 (0.53, 3.58)
≥3 Reports of heavy drinking	2.51 (0.82, 7.66)	1.77 (0.76, 4.12)	0.87 (0.27, 2.80)

[a]Report is based on assessment of alcohol use in the past month at each research interview.
[b]All logistic regression models adjusted for gender, age, race, hepatitis C status, depressive symptoms, antiretroviral therapy use, CD4 cell count, lifetime alcohol diagnosis, and cocaine use.

an etiologic role in the development of lipodystrophy (Ben-Romano et al., 2003; Caron et al., 2001); depressive symptoms may be a consequence of the undesirable bodily changes that define this syndrome (Blanch et al., 2004; Martinez et al., 2001a).

The primary limitation of the present study is that our definition of lipodystrophy relied solely on self-report and did not include a physical assessment by a clinician. The potential for measurement error from the self-reported outcome may have biased our estimates. However, if there is reporting bias, we would expect an under-reporting of lipodystrophy. Thus, if there was a positive association between alcohol use and lipodystrophy, such under-reporting would bias our results toward the null hypothesis of no association. A second limitation of the study was a lack of power to detect associations of the observed magnitude. This study is a secondary data analysis and was not designed to determine the impact of alcohol on lipodystrophy. However, these data provide estimates of the magnitude of the alcohol effect and will be useful in designing future studies that will more definitively assess the effect of alcohol use on HIV-associated lipodystrophy.

Another limitation of the study is the potential for misclassification of alcohol consumption status given that subjects were observed for the past 30 days rather than the past 6 months at each study visit. However, we used a validated measure for assessing alcohol consumption and one that has been shown to be correlated with assessments of longer timeframes (Carey et al., 2004; Koppes et al., 2000). There is also the potential for residual confounding due to uncontrolled confounders or misclassification of confounders, which may have biased the results of this study. Finally, the observational study design and the possibility that lipodystrophy could have occurred prior to alcohol consumption limit our ability to establish a causal relationship between alcohol consumption and development of lipodystrophy. Instead, the results describe the association between the pattern of alcohol use prior to the assessment of lipodystrophy. Of note, if subjects decreased their consumption after developing lipodystrophy, we would expect to see a negative association between drinking and lipodystrophy, not the positive association observed in the study. Nonetheless this study, novel for its evaluation of this research question, provides preliminary, hypothesis-generating data that may be useful in planning future clinical studies.

Alcohol use in HIV-infected patients with alcohol problems may increase the likelihood of developing HIV-associated lipodystrophy. However, larger studies are needed to fully elucidate its role and impact on the development of this common long-term complication of HIV infection and its treatment.

Acknowledgments

We appreciate the contributions of the research associates and data managers. Support came from the National Institute of Alcohol Abuse and Alcoholism (NIAAA) of the NIH: R01-AA13216 (Clinical Impact of HCV and Alcohol in HIV-Infected Persons); R01-AA11785 (Medication Adherence in Alcohol Abusing HIV Patients); R01-AA10870 (Enhanced Linkage of Alcohol Abusers to Primary Care); and K24-AA015674 (Impact of Alcohol Use on HIV Infection—In US and Russia). Research was conducted in part in the General Clinical Research Centers at Boston University School of Medicine, USPHS Grant MO1 RR00533 and BIDMC, USPHS Grant M01 RR01032.

References

Armellini, F., Zamboni, M., Frigo, L., Mandragona, R., Robbi, R., Micciolo, R., et al. (1993). Alcohol consumption, smoking habits and body fat distribution in Italian men and women aged 20–60 years. Eur. J. Clin. Nutr. *47*, 52–60.

Ben-Romano, R., Rudich, A., Torok, D., Vanounou, S., Riesenberg, K., Schlaeffer, F., et al. (2003). Agent and cell-type specificity in the induction of insulin resistance by HIV protease inhibitors. AIDS *17*, 23–32.

Blanch, J., Rousaud, A., Martinez, E., De Lazzari, E., Milinkovic, A., Peri, J. M., et al. (2004). Factors associated with severe impact of lipodystrophy on the quality of life of patients infected with HIV-1. Clin. Infect. Dis. *38*, 1464–1470.

Brinkman, K., Smeitink, J. A., Romijn, J. A., and Reiss, P. (1999). Mitochondrial toxicity induced by nucleoside-analogue reverse-transcriptase inhibitors is a key factor in the pathogenesis of antiretroviral-therapy-related lipodystrophy. Lancet *354*, 1112–1115.

Buchsbaum, D. G., Buchanan, R. G., Centor, R. M., Schnoll, S. H., and Lawton, M. J. (1991). Screening for alcohol abuse using CAGE scores and likelihood ratios. Ann. Intern. Med. *115*, 774–777.

Carey, K. B., Carey, M. P., Maisto, S. A., and Henson, J. M. (2004). Temporal stability of the timeline followback interview for alcohol and drug use with psychiatric outpatients. J. Stud. Alcohol *65*, 774–781.

Caron, M., Auclair, M., Vigouroux, C., Glorian, M., Forest, C., and Capeau, J. (2001). The HIV protease inhibitor indinavir impairs sterol regulatory element-binding protein-1 intranuclear localization, inhibits preadipocyte differentiation, and induces insulin resistance. Diabetes *50*, 1378–1388.

Carr, A., Emery, S., Law, M., Puls, R., Lundgren, J. D., and Powderly, W. G. (2003). An objective case definition of lipodystrophy in HIV-infected adults: a case-control study. Lancet *361*, 726–735.

Coin, A., Sergi, G., Enzi, G., Busetto, L., Pigozzo, S., Lupoli, L., et al. (2005). Total and regional body composition and energy expenditure in multiple symmetric lipomatosis. Clin. Nutr. *24*, 367–374.

Cook, J. A., Cohen, M. H., Burke, J., Grey, D., Anastos, K., Kirstein, L., et al. (2002). Effects of depressive symptoms and mental health quality of life on use of highly active antiretroviral therapy among HIV-seropositive women. J. Acquir. Immune Defic. Syndr. *30*, 401–409.

Enzi, G., Busetto, L., Ceschin, E., Coin, A., Digito, M., and Pigozzo, S. (2002). Multiple symmetric lipomatosis: clinical aspects and outcome in a long-term longitudinal study. Int. J. Obes. Relat. Metab. Disord. *26*, 253–261.

Enzi, G., Inelmen, E. M., Baritussio, A., Dorigo, P., Prosdocimi, M., and Mazzoleni, F. (1977). Multiple symmetric lipomatosis: a defect in adrenergic-stimulated lipolysis. J. Clin. Invest. *60*, 1221–1229.

Folstein, M. F., Folstein, S. E., and 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.

Galli, M., Cozzi-Lepri, A., Ridolfo, A. L., Gervasoni, C., Ravasio, L., Corsico, L., et al. (2002). Incidence of adipose tissue alterations in first-line antiretroviral therapy: the LipoICoNa Study. Arch. Intern. Med. *162*, 2621–2628.

Heath, K. V., Hogg, R. S., Singer, J., Chan, K. J., O'Shaughnessy, M. V., and Montaner, J. S. (2002). Antiretroviral treatment patterns and incident HIV-associated morphologic and lipid abnormalities in a population-based cohort. J. Acquir. Immune Defic. Syndr. *30*, 440–447.

Hengel, R. L., Watts, N. B., and Lennox, J. L. (1997). Benign symmetric lipomatosis associated with protease inhibitors. Lancet *350*, 1596.

Klopstock, T., Naumann, M., Seibel, P., Shalke, B., Reiners, K., and Reichmann, H. (1997). Mitochondrial DNA mutations in multiple symmetric lipomatosis. Mol. Cell. Biochem. *174*, 271–275.

Koppes, L. L., Kemper, H. C., Post, G. B., Snel, J., and Twisk, J. W. (2000). Development and stability of alcohol consumption from adolescence into adulthood: the Amsterdam Growth and Health Longitudinal Study. Eur. Addict. Res. *6*, 183–188.

Martinez, E., Garcia-Viejo, M. A., Blanch, L., and Gatell, J. M. (2001a). Lipodystrophy syndrome in patients with HIV infection: quality of life issues. Drug Saf. *24*, 157–166.

Martinez, E., Mocroft, A., Garcia-Viejo, M. A., Perez-Cuevas, J. B., Blanco, J. L., Mallolas, J., et al. (2001b). Risk of lipodystrophy in HIV-1-infected patients treated with protease inhibitors: a prospective cohort study. Lancet *357*, 592–598.

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

McComsey, G. A., and Walker, U. A. (2004). Role of mitochondria in HIV lipoatrophy: insight into pathogenesis and potential therapies. Mitochondrion *4*, 111–118.

Mujawar, Z., Rose, H., Morrow, M. P., Pushkarsky, T., Dubrovsky, L., Mukhamedova, N., et al. (2006). Human immunodeficiency virus impairs reverse cholesterol transport from macrophages. PLoS Biol. *4*, e365.

National Institute on Alcohol Abuse and Alcoholism. (2005). Helping Patients Who Drink Too Much: A Clinician's Guide. Washington, DC: National Institute on Alcohol Abuse and Alcoholism; Updated 2005 edition. Available at: http://pubs.niaaa.nih.gov/publications/Practitioner/CliniciansGuide2005/guide.pdf. Accessed October 24, 2007.

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., et al. (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.

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

Samet, J. H., Libman, H., LaBelle, C., Steger, K., Lewis, R., Craven, D. E., et al. (1995). A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. Arch. Intern. Med. *155*, 1629–1633.

Sattler, F. (2003). Body habitus changes related to lipodystrophy. Clin. Infect. Dis. *36*, S84–S90.

Schoffer, K., and Grant, I. (2006). Multiple lipomas, alcoholism, and neuropathy: Madelung's disease or MERRF? Muscle Nerve *33*, 142–146.

Smith, K. L., Horton, N. J., Saitz, R., and 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., and Sobell, M. B. (1996). Alcohol Timeline Followback (TFLB). In American Psychiatric Association. (Ed.), Handbook of Psychiatric Measures (pp. 477–479). Washington, DC: American Psychiatric Association.

Teplitsky, V., and Halabe, A. (1999). Fat distribution in AIDS. N. Engl. J. Med. *340*, 969–970.

Villarroya, F., Domingo, P., and Giralt, M. (2007). Lipodystrophy in HIV 1-infected patients: lessons for obesity research. Int. J. Obes. (Lond.) *31*, 1763–1776.

Table of Contents

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

The preponderance of data supporting this statement is derived from:

- (A) level 1 studies, which meet all of the evidence criteria for that study type
- (B) level 2 studies, which meet at least one of the evidence criteria for that study type
- (C) level 3 studies, which meet none of the evidence criteria for that study type or are derived from expert opinion, commentary or consensus

Study types and evidence criteria are defined at http://pier.acponline.org/criteria.html

1. Key Points 4

2. Population at Risk 4

2.1 Recognize that breast cancer incidence and death rates increase most significantly with advancing age and that most cases of breast cancer occur in women aged 40 years and older with no known predisposition (average risk), which supports routine screening in this population 4

2.2 Identify women with a family history of breast cancer, either maternal or paternal, and realize that patients with known or suspected hereditary breast cancer syndromes require tailored screening strategies and genetic counseling 5

2.3 Recognize that although the greatest risk for recurrence is within 5 years of diagnosis, breast cancer survivors remain at increased risk for recurrence or development of a new primary throughout their lifetime 6

2.4 Identify other nonfamilial markers of high risk, such as a personal history of thoracic irradiation (for Hodgkin's disease) or atypical hyperplasia or lobular neoplasia on a breast biopsy, which may justify earlier or more frequent screening 7

2.5 Recognize that despite a lower incidence of breast cancer among blacks compared with whites, mortality rates are higher in blacks than in any other racial or ethnic group, especially among young (<40 years) black women, in part due to underutilization of routine screening 7

2.6 Recognize that although breast cancer risk is associated with hormonal influences, such as early menarche, late menopause, nulliparity, or the use of certain exogenous hormones, these alone may not influence screening recommendations 8

3. Effectiveness/Harms of Screening Tests 9

3.1 Recognize that the sensitivity and specificity of identifying breast lumps with BSE are too low to justify its use in routine screening. 9

3.2 Recognize that although the sensitivity of CBE is only 54%, CBE can detect many of the cancers that are not visualized by mammography. 9

3.3 Appreciate that the sensitivity of routine screening mammography ranges from 71% to 96% and varies by a number of patient, technical, and provider

The American College of Physicians is accredited by the Accreditation Council for continuing Medical Education (ACCME) to provide continuing education for physicians. The American College of Physicians designates this module for a maximum of 1 AMA PRA Category 1 Credit(s)™. Physicians should claim only credit commensurate with the extent of their participation in the activity. **Purpose:** This activity has been developed for internists to facilitate the highest quality professional work in clinical applications, teaching, consultation, or research. Upon completion of the CME activity, participants should be able to demonstrate an increase in the skills and knowledge required to maintain competence, strengthen their habits of critical inquiry and balanced judgement, and to contribute to better patient care. **Disclosures:** *Karen M. Freund, MD, MPH[1], Pamela Ganschow, MD[1], Sarah E. Lane, MPH[1], Teresa Cheng, MD[1], Tracy A. Battaglia, MD, MPH[1], Steven E. Weinberger, MD, FACP, Acting Editor, PIER, has stock holdings in GlaxoSmithKline and Abbott.*
[1] Has no financial relationships with pharmaceutical companies, biomedical device manufacturers, or health-care related organizations.

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

factors, whereas the specificity ranges from 94% to 97%..... 10
3.4 Know that overall cancer detection rates for digital mammography are similar to those for traditional film (analog) mammography. 11
3.5 Be aware that current evidence suggests that the use of computer-aided detection systems contributes minimally to improvement in cancer detection rates, mainly in the detection of in situ rather than invasive cancer. 12
3.6 Know that among high-risk women, the sensitivity of contrast-enhanced breast MRI is higher (77% [CI, 70% to 84%]) than that of standard mammography (39% [CI, 37% to 41%])..... 13
3.7 Understand that currently there is no evidence for the use of contrast-enhanced breast MRI as a screening tool for average-risk women. ..14
3.8 Understand that currently there is no evidence to recommend the use of ultrasound to screen women at high risk for breast cancer..... 14
3.9 Realize that the major harm associated with all screening modalities is a high false-positive rate, which may lead to unnecessary diagnostic procedures..... 15
3.10 Realize that false reassurance due to false-negative mammography findings is low among women aged 50 to 65 years and increases among younger women..... 16
3.11 Know that the available evidence suggests that the radiation risk from mammography is low compared to the benefit of routine screening, but no prospective studies have been conducted..... 16
3.12 Recognize that a false-positive screening test result can significantly increase a woman's anxiety about breast cancer, but there are limited data suggesting an association with long-term effects or a negative impact on future screening behavior..... 17
3.13 Recognize that many women report pain from mammography, yet few report that the pain would impact their future screening behavior..... 18
3.14 Understand that overdiagnosis of clinically insignificant disease, mainly DCIS, is possible with screening mammography. 18

4. Effectiveness/Harms of Early Treatment..... 18
4.1 Realize that the data clearly indicate that routine screening mammography identifies breast cancer at earlier stages, when the survival time is greatest... 19
4.2 Realize that breast-conserving surgery with radiation therapy is equivalent to mastectomy in patients with early-stage disease, and, therefore, both treatment options should be explored..... 19
4.3 Recognize that overtreatment of clinically insignificant breast cancer is possible and may lead to an increase in morbidity. 20
4.4 Know that the most common complications of axillary lymph node dissection for early-stage breast cancer are lymphedema, nerve injury, and shoulder dysfunction and that the use of sentinel node biopsy has been associated with a lower risk of these postsurgical complications..... 20
4.5 Know that the complications of radiation therapy are generally mild and are decreasing with modern techniques..... 21

5. Direct Evidence that Screening Reduces Adverse Outcomes.. 21
5.1 Understand that screening mammography remains the most studied of all available breast cancer screening modalities and the only modality that is associated with a reduction in breast cancer mortality, which is estimated to be greatest among women aged 50 to 69 years..... 22
5.2 Know that there are insufficient data from randomized, controlled trials of

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

screening mammography to confirm a mortality benefit among women over age 70 23
5.3 Acknowledge that the indirect evidence associating screening CBE with a reduction in the breast cancer mortality rate comes from randomized, controlled trials that used both CBE and mammography for breast cancer detection 24
5.4 Realize that BSE alone as a screening modality does not reduce mortality from breast cancer 24

6. Timeline 25
6.1 Understand that data consistently support the use of routine mammography in women aged 50 to 70 years, but that, due to a lack of data, controversy remains regarding the optimal frequency for all women as well as the appropriate starting and stopping time for women in their 40s and those over age 70, respectively 25
6.2 Understand that screening intervals for certain high-risk patients, such as those with a positive family history, a personal history of atypical ductal hyperplasia or LCIS, or a history of thoracic radiation, are often tailored and based on expert opinion 26
6.3 Be aware that annual MRI is recommended for women over age 30 with a genetic mutation or beginning 5 to 10 years younger than the youngest family member with breast cancer when the lifetime risk is greater than 25%. 27

7. Cost-Effectiveness 27
7.1 Appreciate that screening mammography has an acceptable cost per life-year saved compared with other screening strategies. 27
7.2 Recognize that the cost-effectiveness of breast cancer screening is affected by the frequency of mammography, the age of the screened population, and whether mammography is coupled with CBE 28

8. Patient Education 29
8.1 Know that not all women are well informed about breast cancer screening 29
8.2 Know that a personal recommendation from a health care provider is the best method to increase mammography adherence; direct patient education interventions have shown a minimal or no increase in mammography adherence 29

9. Referral/Consultation 29
9.1 Consider referral to a breast specialist for certain high-risk patients, such as those with a family history of breast cancer or a personal history of atypical ductal hyperplasia, LCIS, or thoracic radiation 30

10. Guidelines 30

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

1. Key Points

1.1 Know that breast cancer is the most common non-skin cancer among women in the U.S.

1.2 Know that age alone remains the greatest risk factor for breast cancer, and most new breast cancer cases occur in women with no other known predictors.

1.3 Recognize that mammography remains the most widely studied breast cancer screening modality and the only available screening modality that has been shown to reduce mortality.

1.4 Recognize that the best evidence of routine mammographic screening efficacy exists for average-risk women aged 50 to 69 years.

1.5 Identify patients at increased risk for breast cancer who require tailored screening strategies based on personal history (i.e., thoracic irradiation) or family history (i.e., known or suspected hereditary breast cancer syndromes).

2. Population at Risk

2.1 Recognize that breast cancer incidence and death rates increase most significantly with advancing age and that most cases of breast cancer occur in women aged 40 years and older with no known predisposition (average risk), which supports routine screening in this population.

Evidence:

- Excluding nonmelanoma cancers of the skin, breast cancer is the most common female cancer (26% of all new cancers) and the second most common cause of cancer death (15% of all cancer deaths) in women in the U.S. It is estimated that in 2008 there will be 182,460 new cases of invasive breast cancer, 67,770 new cases of in situ breast cancer, and 40,480 breast cancer deaths (1; 2).
- Breast cancer is the leading cause of cancer death for women aged 20 to 59 years, with 13,368 deaths in 2005 (1).
- The cumulative lifetime risk of developing invasive breast cancer is 12.3% (or 1 in 8 women). This is considered the average risk (3; 4).
- Age is one of the most significant risk factors for breast cancer. The risk at age 70 is more than 10 times the risk at age 30 (2; 3).
- The median age at diagnosis for breast cancer from 2001 to 2005 was 61 years, and the median age at death was 69 years (4).

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

- From 2000 to 2004, 95% of new cases of breast cancer and 97% of breast cancer deaths occurred in women aged 40 years and older. The majority of new cases (approximately 70%) were in women aged 55 years and older (3).

Comments:

- More than 50% of new breast cancer cases occur in women without known predictors (5).

2.2 Identify women with a family history of breast cancer, either maternal or paternal, and realize that patients with known or suspected hereditary breast cancer syndromes require tailored screening strategies and genetic counseling.

Evidence:

- Epidemiologic data on 193 breast cancer cases accrued from a cohort of 7508 women in the first National Health and Nutrition Examination Survey Epidemiologic Follow-up Study indicate that a positive family history of breast cancer in a first-degree relative is associated with a moderate increase in the risk of breast cancer (RR, 2.6) (6).
- A cross-sectional study of 389,300 women undergoing screening mammography showed that the rates of breast cancer detection in women with a family history of breast cancer in a first-degree relative were increased and were similar to the cancer detection rates in women at average risk who were 10 years older (7).
- A meta-analysis of 74 published studies found that the magnitude of pooled risk is further stratified by certain features of the family history, including degree of relatedness to the affected family member, number of affected family members, age of the affected family member, presence of bilateral cancer in affected family members, and family history of ovarian cancer. The relative risks were as follows: one first-degree relative, 2.1 (CI, 2.0 to 2.2); one first-degree relative under age 50, 2.3 (CI, 2.2 to 2.5); one first-degree relative with bilateral breast cancer, 9.8 (CI, 2.0 to 24.9); one second-degree relative, 1.5 (CI, 1.4 to 1.6); two first-degree relatives, 3.6 (CI, 2.5 to 5.0); first-degree relative with ovarian cancer, 1.27 (CI, 0.9 to 1.8) (8).
- Data from a population-based, case-control study of women with a family history of breast cancer that included 4730 women aged 20 to 54 years with breast cancer and 4688 control subjects found that the risk for women with one or more affected relatives is always higher than that for the general population. This risk decreases and approaches the rate for the general population as the age at onset for the relative increases. Most importantly, the risk of early-onset breast cancer is highest when there is a family history of early (premenopausal) breast cancer in a first-degree relative. The risk for women with second-degree relatives with a history of breast cancer is half that of those with a first-degree relative with a history of breast cancer (9).
- Known inherited breast cancer syndromes include hereditary breast and ovarian cancer syndrome (*BRCA1/BRCA2*), Cowden's syndrome (*PTEN*),

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

Li-Fraumeni syndrome (*TP53*), and Peutz-Jeghers syndrome (*STK11*). Approximately 80% to 90% of known inherited breast cancer syndromes involve a *BRCA1* or *BRCA2* mutation (10).

- The combined prevalence of *BRCA1* and *BRCA2* mutation is 1 in 40 in the Ashkenazi Jewish population (11).
- The risk of developing breast cancer by age 70 is estimated to be 65% and 45% for *BRCA1* and *BRCA2* mutation carriers, respectively (12; 13).

Comments:

- When family history appears to be the primary risk factor, the individual risk assessment models (Cancer and Steroid Hormone Study [CASH] and BRCAPRO) may help to determine the age at which to initiate screening and the appropriate frequency and modality (9; 14).
- The American Cancer Society now recommends routine screening with annual MRI alternating every 6 months with mammography for women with a greater than 25% risk of developing breast cancer.
- The National Comprehensive Cancer Network (http://www.nccn.org/professionals/physician_gls/PDF/genetics_screening.pdf) recommends referral to a cancer genetics professional if there is a strong family history of breast and/or ovarian cancer suggestive of an inherited syndrome.

2.3 Recognize that although the greatest risk for recurrence is within 5 years of diagnosis, breast cancer survivors remain at increased risk for recurrence or development of a new primary throughout their lifetime.

Evidence:

- The National Surgical Adjuvant Breast and Bowel Project found a cumulative rate of local recurrence of 14% 20 years after breast-conserving surgery followed by radiotherapy. This rate went up to 39% if radiation therapy was not administered (15).
- An Italian study of 2233 women treated with breast-conserving surgery followed by radiation showed a local recurrence rate of 1% per year, with distant recurrences found most frequently in the second year and declining each year thereafter, such that recurrence rates were 1.5% in the fifth year and 1% in the tenth year (16).
- A review of the Eastern Oncology Cooperative Group data from 3585 women found a recurrence rate of 2% per year, with the greatest recurrence in the second year (17).
- A review of 16 cohort studies found that women with a history of breast cancer also have an increased risk of developing a new cancer in the contralateral breast. This risk of developing a second primary tumor is two to six times the risk in the general population and is even higher for patients with a hereditary cancer syndrome (18).

Comments:

- In the U.S., 85% of all patients with breast cancer are alive 5 years after diagnosis, corresponding to an estimated 2 million women (19).

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

2.4 Identify other nonfamilial markers of high risk, such as a personal history of thoracic irradiation (for Hodgkin's disease) or atypical hyperplasia or lobular neoplasia on a breast biopsy, which may justify earlier or more frequent screening.

Evidence:

- Data from the Childhood Cancer Survivor Study, which included over 6000 female survivors and 111 cases of breast cancer, showed that young female survivors treated with chest radiation are at an increased risk for breast cancer compared with women in the age-matched general population. Breast cancer risk in childhood Hodgkin's disease survivors treated with chest radiation therapy had an observed standardized incidence ratio of 26.3 compared to 4.8 among all survivors who did not receive chest irradiation (20).
- Atypical ductal hyperplasia is believed to be a precursor to invasive cancer and represents a midpoint in the histologic continuum of proliferative breast disease between usual ductal hyperplasia and DCIS. A 2005 review of the literature found that women with a history of one or more breast biopsies showing atypical hyperplasia have a four- to six-fold increase in their lifetime breast cancer risk (21).
- Data from the National Surgical Adjuvant Breast and Bowel Project show that women with lobular neoplasia on a breast biopsy have a risk of developing breast cancer in either breast of approximately 1% per year (22).

Comments:

- Based on the high risk of breast cancer in women who received radiation to the chest between the ages of 10 and 35 years, annual breast MRI screening has been recommended as an adjunct to mammography by an expert panel convened by the American Cancer Society (23).

2.5 Recognize that despite a lower incidence of breast cancer among blacks compared with whites, mortality rates are higher in blacks than in any other racial or ethnic group, especially among young (<40 years) black women, in part due to underutilization of routine screening.

Evidence:

- From 2000 to 2004, the overall incidence rate of breast cancer among black women was 118.3 in 100,000 compared to 132.5 in 100,000 among white women (unadjusted for age) (1).
- The overall mortality rate among black women with breast cancer from 2000 to 2004 was 33.8 in 100,000 compared to 25 in 100,000 among white women (2).
- Data from nine SEER databases between 1995 and 2004 showed that the age-adjusted incidence of invasive breast cancer for black women under age 40 was significantly higher than that for white women (RR, 1.16 [CI, 1.10 to 1.23]). The age-adjusted mortality rate for black women

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

under age 40 was twice that for white women. Compared to white women, black women were significantly more likely to be diagnosed with regional or distant disease, have a lower relative 5-year survival rate, and have a higher likelihood of being diagnosed with tumors with poorer prognosis (24).
- Breast cancer disparities have been attributed to a complex interplay of biologic tumor characteristics along with racial and social inequalities, leading to unequal access to timely, quality cancer care (25).

Comments:
- None.

2.6 Recognize that although breast cancer risk is associated with hormonal influences, such as early menarche, late menopause, nulliparity, or the use of certain exogenous hormones, these alone may not influence screening recommendations.

Evidence:
- Early age at menarche (<12 years) and late age at menopause (>54 years) are associated with relative risks of breast cancer of 1.5 and 2.0, respectively (26; 27).
- An analysis of over 7500 women in the first National Health and Nutrition Examination Survey found that nulliparity and late age at first full-term pregnancy (>30 years) are associated with relative risks of 1.8 (CI, 1.1 to 2.9) and 1.9 (CI, 1.1 to 3.3), respectively (6). Meanwhile, women who have their first child before age 20 have a 50% reduction in lifetime breast cancer risk compared with women who do not have children (28).
- Two meta-analyses based largely on older studies showed a small increase in the relative risk of breast cancer with oral contraceptive use. The Collaborative Group on Hormonal Factors in Breast Cancer study found a relative risk of 1.07 (CI, 1.02 to 1.13) for women who had ever used oral contraceptives compared to those who had never used them (29). A more recent meta-analysis found an increased risk of premenopausal breast cancer, with a relative risk of 1.19 (CI, 1.09 to 1.29) (30). However, the more recent Women's Contraceptive and Reproductive Experiences study found no association between past or present oral contraceptive use and breast cancer (RR, 1.0 [CI, 0.8 to 1.3]) (31).
- Data from the Women's Health Initiative, a randomized, controlled trial including 16,608 postmenopausal women, found that the use of exogenous hormones (conjugated equine estrogens, 0.625 mg/d, plus medroxyprogesterone acetate, 2.5 mg/d) is associated with a 26% excess of breast cancer (hazard ratio, 1.26 [CI, 0.83 to 1.92]) (32).

Comments:
- An individual's relative and absolute risk of developing breast cancer can be calculated using a multivariate risk prediction model, such as the Breast Cancer Risk Assessment Tool (Gail model).
- The Breast Cancer Risk Assessment Tool is most useful in women in the

absence of an extensive family history of breast or ovarian cancer and when medical counseling and entry into clinical trials are being considered.

3. Effectiveness/Harms of Screening Tests

3.1 Recognize that the sensitivity and specificity of identifying breast lumps with BSE are too low to justify its use in routine screening. (A)

Evidence:
- A review by the U.S. Preventive Services Task Force estimated a 20% to 30% sensitivity of BSE alone as a screening modality for breast cancer (33).
- The sensitivity of BSE has been found to decrease with age (41% for women aged 35 to 39 and 21% for women aged 60 to 74) (33; 34).
- In a study of lump detection in silicone breast models, the sensitivity was 40% to 89%, and the specificity was 66% to 81% (35).
- Although the accuracy of BSE has been shown in some studies to increase with formal training and with the thoroughness of the exam (36; 37), other studies have shown no improvement with additional instruction (38).

Comments:
- None.

3.2 Recognize that although the sensitivity of CBE is only 54%, CBE can detect many of the cancers that are not visualized by mammography. (B)

- See table Operating Characteristics for Breast Cancer Screening Tests.

Evidence:
- CBE had a sensitivity of 54% and a specificity of 94% in a review of pooled data from six large, controlled breast cancer screening trials in which CBE was included as part of the breast cancer screening regimen. In this review, CBE alone detected between 5.2% and 29% of breast cancers that mammography missed (39).
- In the Health Insurance Plan of Greater New York Randomized Controlled Trial, Edinburgh Randomized Trial of Breast Cancer Screening, and the Canadian National Breast Screening Study-1 and -2, the sensitivity of combined screening with mammography and CBE was higher than that of mammography alone (40; 41; 42; 43; 44).
- In a study of over 750,000 CBEs provided to low-income women in the National Breast and Cervical Cancer Early Detection Program, 5% of cancers were found in patients with abnormal CBE results recorded but mammographic results that were reported as negative or benign. Sensitivity and specificity in this group for CBE were 58.8% and 93.4%, respectively (45).
- In a study of 61,688 women aged 40 years and older, CBE detected an

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

additional 4% of invasive cancers. Sensitivity increased when CBE was added to screening mammography in women of all ages, from 6.8% in women aged 50 to 59 years with dense breasts to 1.8% in women aged 60 to 69 years with fatty breasts (46).

- Sensitivity can also be evaluated for the detection of lumps embedded in silicone breast models. CBE sensitivity as measured on silicone models ranged from 40% to 71%. CBE duration and use of correct CBE technique correlated significantly with lump detection accuracy in silicone models (39).
- Data from both clinical studies and studies using silicone models suggest that additional factors, such as age, obesity, breast size, and baseline breast nodularity, also affect CBE sensitivity (39; 47).

Comments:

- The MammaCare® method is the preferred technique for CBE and has been validated in a randomized, controlled trial. The technique consists of several components, including patient positioning so that the breast tissue is flattened against the patient's chest, examining all of the breast tissue beginning in the axilla and extending down the midaxillary line to the bra line and then moving medially in vertical strips to cover all the tissue between the clavicle and the bra line to the sternum, and using the pads of the three middle fingers to palpate each area of breast tissue in three small circles using three different pressures (superficial, intermediate, and deep) (36; 39; 48). The American Cancer Society and Centers for Disease Control and Prevention also support the use of this standardized approach to CBE (49).

3.3 Appreciate that the sensitivity of routine screening mammography ranges from 71% to 96% and varies by a number of patient, technical, and provider factors, whereas the specificity ranges from 94% to 97%.

- See table Operating Characteristics for Breast Cancer Screening Tests.
- See table Randomized, Controlled Trials of Screening Mammography.

Evidence:

- The majority of information about the operating characteristics of mammography is derived from reviews of data from eight randomized, controlled trials evaluating the effectiveness of screening mammography among 490,000 women over an average of 14 years of follow-up (5; 44; 50).
- The sensitivity of screening mammography is lower for women in their 40s compared with women aged 50 years and older (5; 51).
- The positive predictive value of mammography increases with age, from 1% to 4% for women in their 40s to 20% for women in their 70s (52). The positive predictive value varies widely across studies, depending on how a "positive mammogram" is defined (e.g., a mammogram requiring "further evaluation" or a mammogram requiring "biopsy evaluation").
- The sensitivity of screening mammography is lower in women with denser breasts (5; 51).

- The sensitivity of screening mammography is lower in women on hormone replacement therapy. A number of studies have shown that the sensitivity of screening mammography is reduced by 7% to 21% in current hormone replacement therapy users compared to never or former users, partly due to the changes in breast density resulting from hormone use (5; 53; 54; 55; 56; 57).
- The sensitivity and specificity of screening mammography are unchanged by a woman's family history of breast cancer; however, because the disease is more prevalent in women with a family history of breast cancer, the positive predictive value of mammography is higher in this population (7).
- The sensitivity of screening mammography also is affected by the quality of mammography, the number of mammographic views, the experience of the radiologist, and the choice of follow-up evaluation for abnormal test results (5).

Comments:

- Because the sensitivity of mammography is not 100%, palpable breast lumps and other breast symptoms still should be evaluated completely (e.g., biopsy) in the presence of a normal mammography result.

3.4 Know that overall cancer detection rates for digital mammography are similar to those for traditional film (analog) mammography.

- See table Operating Characteristics for Breast Cancer Screening Tests.

Evidence:

- Four large, prospective studies involving over 85,000 women compared the use of film mammography vs. digital mammography and showed little difference in the overall cancer detection rates. The Oslo I study conducted in 2000 obtained both film and digital mammography in 3683 women. Film mammography detected 28 of 31 cancers, whereas digital mammography detected 23 of 31 cancers (P=0.23) (58).
- Another study obtained both film and digital mammography in 6736 women. Of the 42 cancers detected after 1-year follow-up, 27 were detected with digital mammography and 33 were detected with film mammography (P<0.10) (59).
- The Oslo II study randomly assigned 25,263 women to either digital or film mammography and found slightly higher cancer rates for all women with digital mammography vs. film mammography (59% vs. 41%; P<0.06). Digital mammography performed slightly better in women aged 50 to 69 years, but there was no difference in women aged 40 to 49 years (60).
- The largest study, the Digital Mammographic Imaging Screening Trial, obtained both digital and film mammography in 49,760 women. The overall sensitivity of film mammography was not different from that of digital mammography (52% vs. 55%); however, the sensitivity of digital mammography was higher than that of film mammography in women under age 50 (67% vs. 44%) and women with dense breasts (57% vs.

44%) (61).

Comments:

- The advantages of digital mammography over film mammography include easier storage and retrieval, the ability to manipulate images, the potential for lower average radiation exposure, and the capacity for real-time teleradiology (sending films off site for immediate interpretation). However, these advantages come at a significant increase in cost (up to four times the cost of film mammography).
- Traditional film mammography is still a very acceptable modality for breast cancer screening. Digital mammography may provide an increase in detection rates for young women or those with dense breasts.

3.5 Be aware that current evidence suggests that the use of computer-aided detection systems contributes minimally to improvement in cancer detection rates, mainly in the detection of in situ rather than invasive cancer.

- See table Operating Characteristics for Breast Cancer Screening Tests.

Evidence:

- A large pre- and postobservational study (n=429,345 mammograms) compared the performance of mammography at 43 mammography facilities. Seven of these facilities implemented computer-aided detection during the study period from 1998 to 2002 (n=31,186 mammograms). Among these seven facilities, there was a nonsignificant increase in the sensitivity (from 80.4% to 84%) before and after implementation of computer-aided detection, a significant decrease in the specificity (from 90% to 87%) and the positive predictive value (from 4.1% to 3.2%), and a 20% increase in the biopsy rate (P<0.001). There was also a trend toward an increase in the cancer detection rate, which was limited to DCIS, did not include invasive cancers, and was not statistically significant (62).
- A study compared 10,267 mammograms, originally read in 1996, 236 of which detected cancer (83% invasive, 15% in situ) after 3 years of follow-up. At a later date, for purposes of the study, each set of films was either double read by two radiologists or single read with computer-aided detection by four radiologists. The overall rate of cancer detection with a single read plus computer-aided detection was higher (49.1%) than that for the double read without computer-aided detection (42.6%; P=0.02) (63).
- A prospective study compared radiologist-only interpretation of screening mammograms to evaluation with computer-aided detection after radiologist reading. Of the 12,860 women in the study, the radiologists recalled 830 (6.5%), whereas computer-aided detection resulted in an additional 156 patients recalled (7.9% total). Biopsy was recommended for 124 women. Detection rates for invasive cancer were 29 without computer-aided detection and 31 with computer-aided detection; DCIS was detected in 12 women without computer-aided detection and 18 women with computer-aided detection. Detection of invasive cancer was

not increased, whereas detection of DCIS was increased (64).

- Over a 3-year period, a single academic system with six sites and 24 radiologists evaluated 59,139 mammograms with computer-aided detection and 56,432 without. Rates of cancer detection were 3.49% with computer-aided detection and 3.55% without computer-aided detection (P=0.68). Recall rates were similar (11.62% vs. 11.05%) (65).

Comments:

- Computer-aided detection involves the use of a computer-based system that scans mammographic images and uses algorithms to mark areas of concern. It can be used with digital images or film images that are then digitized. The typical system places an average of three marks on a set of bilateral craniocaudal and medial lateral oblique images (64). Computer-aided detection is approved by the FDA and is covered by Medicare and many third-party insurers.

3.6 Know that among high-risk women, the sensitivity of contrast-enhanced breast MRI is higher (77% [CI, 70% to 84%]) than that of standard mammography (39% [CI, 37% to 41%]).

- See table Operating Characteristics for Breast Cancer Screening Tests.

Evidence:

- The Magnetic Resonance Imaging Screening Study obtained annual MRIs and mammograms between 1999 and 2003 for 1909 women aged 25 to 70 years with hereditary breast cancer (greater than 15% lifetime risk by Claus criteria), 358 of whom were known to have a *BRCA1* or *BRCA2* mutation. MRI sensitivity was 71% overall (70% in mutation carriers) compared with 40% (30% in mutation carriers) for mammography. Specificity was 99% for mammography and 96% for MRI (66).
- In a study of 236 women with a known *BRCA1* or *BRCA2* mutation, the sensitivity of mammography was 36% compared with 77% for MRI and 27% for ultrasound. Specificity was 93%, 95%, and 99% for MRI, ultrasound, and mammography, respectively (67).
- A prospective, multicenter, cohort study in the UK offered 649 women aged 35 to 49 years annual screening with both mammography and MRI for 2 to 7 years. All of the women were at high risk, with a strong family history of breast cancer or a known mutation. Thirty-five cancers were detected during the study. The sensitivity of mammography was lower (40%) than that of MRI (93%) and was much lower for women with a personal or family history of a *BRCA1* mutation (23% vs. 92%), in whom 13 of the cancers were identified. Specificity was 81% for MRI compared with 93% for mammography (68).
- A cohort of 529 women with a greater than 20% lifetime risk of breast cancer underwent screening with mammography, MRI, and ultrasound. After a mean follow-up of 5.3 years, 43 cancers (34 invasive) were detected. Sensitivity was 91% for MRI compared with 33% for mammography and 40% for ultrasound. The specificities for both MRI and mammography were 97% (69).

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

- A meta-analysis of these studies calculated the sensitivity of MRI to be 77% (CI, 70% to 84%) compared with 39% (CI, 37% to 41%) for mammography. The specificity of MRI was 86.3% (CI, 80.9% to 91.7%) compared with 94.7% (CI, 93.0% to 96.5%) for mammography (70).

Comments:

- Four large prospective non-randomized studies and seven smaller studies have been reported, plus one meta-analysis. All compare the sensitivity and specificity of MRI and mammography when both are used annually for screening high-risk women with hereditary breast cancer (66; 67; 68; 69; 70; 71; 72; 73; 74; 75; 76; 77).
- Three of the aforementioned studies also included screening ultrasound of the breast, which had sensitivities similar to those of mammography.
- The Magnetic Resonance Imaging Screening Study found that MRI was more sensitive in detecting invasive cancers, especially those that are small, hormone receptor negative, and nonductal. Mammography was better at detecting DCIS.
- The American Cancer Society and the National Comprehensive Cancer Network (http://www.nccn.org/professionals/physician_gls/PDF/breast-screening.pdf) have released guidelines regarding the use of breast MRI as a screening tool in high-risk women. The American Cancer Society recommendations include women with an estimated lifetime breast cancer risk of 20% or higher, as defined by prediction models that are based largely on family history, such as the BRCAPRO and Claus models (23; 78).
- The sensitivities reported for screening mammography in these studies are lower than those reported in the evidence for sensitivity of routine screening mammography and detection rates for digital mammography due to the younger ages of the populations studied.

3.7 Understand that currently there is no evidence for the use of contrast-enhanced breast MRI as a screening tool for average-risk women. Ⓒ

Evidence:

- There are no clinical trials of the use of MRI to screen for breast cancer in average-risk women.

Comments:

- The American Cancer Society specifically recommends against screening MRIs for women at average risk for breast cancer, which is defined as a lifetime risk of less than 15% (23).
- The National Comprehensive Cancer Network guidelines (http://www.nccn.org/professionals/physician_gls/PDF/breast-screening.pdf) do not recommend MRI for average-risk women (78).

3.8 Understand that currently there is no evidence to recommend the use of ultrasound to screen women at high risk for breast cancer. Ⓑ

- See table Operating Characteristics for Breast Cancer Screening Tests.

Evidence:

- In a study of 236 women with a known *BRCA1* or *BRCA2* mutation, the sensitivity of ultrasound was 27% compared with 36% for mammography and 77% for MRI (67).
- A cohort of 529 women with a greater than 20% lifetime risk of breast cancer underwent screening with mammography, MRI, and ultrasound. After a mean follow-up of 5.3 years, 43 cancers (34 invasive) were detected. Sensitivity was 91% for MRI compared with 33% for mammography and 40% for ultrasound. The specificities for both MRI and mammography were 97% (69).
- In a study of women at high risk for breast cancer (over 50% with a personal history and a 5-year risk of at least 1.7% according to the Breast Cancer Risk Assessment Tool) with dense breast tissue, the sensitivities of mammography and ultrasound each were 50% and increased to 77% when the modalities were used in combination. The specificities of mammography and ultrasound alone were higher (96% and 92%, respectively) than in combination (89%). The false-positive rate using ultrasound was higher (12%) than that of mammography (9.6%) (79).

Comments:

- Ultrasound appears to add little to screening mammography in high-risk women.

3.9 Realize that the major harm associated with all screening modalities is a high false-positive rate, which may lead to unnecessary diagnostic procedures.

Evidence:

- The majority of abnormal screening mammograms (80% to 90%) lead to false-positive test results that initiate additional diagnostic imaging and/or biopsy (5; 34; 44; 50).
- Over a decade of annual screening, a woman has a 50% cumulative risk of having at least one false-positive mammography result. Among women who do not have breast cancer, almost 20% will undergo breast biopsies during this 10-year screening interval (80).
- The risk of having a false-positive mammography result varies widely and is based on both patient factors and radiologist and/or system factors. The risk is highest in younger women because the specificity of mammography is lower and because breast cancer is less common in this population (81).
- Current evidence shows that computer-aided detection may increase the false-positive rate of mammography (62).
- The recall rates associated with digital mammography appear to be slightly higher than those for film mammography. Recall rates vary widely between studies, from 3.7% to 4.6% in the Oslo I and II studies (58; 60) to 8.5% to 12% in American trials (59; 61).
- The risk of a false-positive test result is higher with MRI than with

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

mammography. Although there is as much as a doubling of the cancer detection rate with the addition of MRI to mammography, the recall rate and biopsy rate are also twice as high (66; 68). This is due to the lower specificity of MRI, as seen in all of the large MRI screening studies.

- In the MRI screening studies, recall rates (for additional imaging) range from 8% to 17%, and biopsy rates range from 3% to 15%. The majority of patients recalled for additional imaging were able to be evaluated without the need for biopsy (66; 68; 70).
- Several studies suggest that the highest false-positive rate is in the first round of screening and that the recall rates decrease to less than 10% in subsequent rounds of screening (66; 82; 83).

Comments:
- None.

3.10 Realize that false reassurance due to false-negative mammography findings is low among women aged 50 to 65 years and increases among younger women. B

Evidence:
- Screening mammography has been found to have a false-negative rate ranging from 5% to 30%. The risk of a false-negative mammography result is higher in younger women because mammography is less sensitive in this population (5; 34; 44; 50).

Comments:
- None.

3.11 Know that the available evidence suggests that the radiation risk from mammography is low compared to the benefit of routine screening, but no prospective studies have been conducted. B

Evidence:
- Established benefit from annual mammography far outweighs the theoretical risk from radiation exposure. It has been estimated that annual mammographic screening of 100,000 women for 10 consecutive years (starting at age 40) would result in 40 lives saved and a maximum of eight breast cancer deaths occurring during their lifetime (84).
- A case-control study involving 1600 patients with breast cancer and 1600 control subjects without breast cancer who were matched for *BRCA* mutation found no association between ever having undergone screening mammography and risk of breast cancer, even among *BRCA1* and *BRCA2* carriers (85).

Comments:
- Modern mammography systems require <0.2 rad per exposure for an average-sized breast (86). This is in comparison to high-dose radiation exposure for treatment of Hodgkin's disease, which exposes the breasts to as much as 100 to 2000 rad (87).

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

3.12 Recognize that a false-positive screening test result can significantly increase a woman's anxiety about breast cancer, but there are limited data suggesting an association with long-term effects or a negative impact on future screening behavior. B

Evidence:

- Several studies have demonstrated the impact of false-positive mammography results on patient anxiety level. In a study of 8854 women undergoing mammography, 47% of those with a false-positive result reported anxiety related to the mammographic findings and concern regarding breast cancer. Anxiety levels were persistently elevated at 3 months, although they had decreased from levels measured at 3 weeks after the mammogram (88).
- Similar findings in another study of 308 women suggest that almost 20% of the women with adverse psychological effects from mammography had an impaired ability to perform daily activities (89).
- Elevated anxiety levels can continue for up to 18 months (90) and appear to persist despite an additional evaluation excluding the diagnosis of breast cancer (88; 89).
- A study in the UK found that women with false-positive mammography results had significantly greater adverse psychological consequences compared to women who had received a clear result at their initial mammogram when surveyed 1 month before their next routine breast screening appointment 3 years later (91).
- A prospective cohort study examining psychological distress in a group of Finnish women aged 50 years and older showed intrusive thinking and anxiety regarding breast cancer 12 months after a false-positive mammography result; however, 98.7% of the women with false-positive results intended to re-attend screening (92).
- In a Swedish study of 509 women recalled for further investigation after mammographic screening, there was a high prevalence of anxiety before the recall visit but no evidence of increased long-term anxiety or depression at 12 months (93).
- A cross-sectional survey of 479 women found that there was a good understanding of the possibility of false-positive test results in screening mammography and that women seem to accept false-positive results as a consequence of screening (94).
- Two systematic reviews have examined the psychological effect of mammographic screening. The first reviewed 10 studies and determined that anxiety appears to be the most common consequence of mammography, with the most significant effects on women requiring further investigation because of abnormal results (95). A second systematic review included 11 studies that assessed anxiety and 6 studies that assessed worry and found that false-positive mammography results may have persistent, small effects on some women's psychological well-being and behavior. No long-term symptoms of depression were noted in women with false-positive mammography

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

results. In the U.S., women with false-positive results as opposed to normal mammographic findings were more likely to return for routine screening (96).
- Immediate reading of screening mammograms was associated with less anxiety among women with false-positive mammography results 3 weeks after mammography compared to an educational intervention targeting coping skills (88).

Comments:
- None.

3.13 Recognize that many women report pain from mammography, yet few report that the pain would impact their future screening behavior.

Evidence:
- In a study of 954 women undergoing mammographic screening in the Netherlands, 72.9% reported mammography as painful; however, only 2.7% indicated that the pain would deter them from undergoing future mammographic screening (97).
- In a survey of 597 women, 35% of women reported discomfort and 6% reported pain during mammography. No effect on satisfaction or intention to re-attend screening was noted (98).
- In a random sample of 883 Finnish women undergoing mammographic screening, 61% reported pain or discomfort. Among those reporting pain or discomfort, there was no evidence of intent to avoid further screening; however, the study did note that the 5 women (1%) who were undecided on further screenings had experienced severe or moderate pain during mammography (99).

Comments:
- None.

3.14 Understand that overdiagnosis of clinically insignificant disease, mainly DCIS, is possible with screening mammography. 

Evidence:
- Population-based data from the Eindhoven Cancer Registry showed that the rate of DCIS increased from 3 per 100,000 person-years to 34 per 100,000 person-years over the last 20 years due to the increased use of screening mammography (100).
- Approximately 20% of breast cancers detected by screening mammography are DCIS. There is no way to predict which cases will progress to invasive cancer; however, it is estimated that only 10% to 32% of DCIS lesions will progress to invasive cancer (101).

Comments:
- None.

4. Effectiveness/Harms of Early Treatment

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

4.1 Realize that the data clearly indicate that routine screening mammography identifies breast cancer at earlier stages, when the survival time is greatest.

- See table SEER Relative Survival Rates by Stage at Diagnosis for Breast Cancer.

Evidence:

- Mammography identifies cancer at earlier stages (102; 103).
- SEER data from 1983 to 1992 show a decrease in the diagnosis of late-stage tumors.
- Treatment at an early stage directly affects survival rates (104).
- A study conducted by the Breast Cancer Surveillance Consortium between 1996 and 2000 evaluated the pathologic outcomes of 786,846 women aged 40 to 89 years after screening mammography. The majority of invasive tumors were small: 35% were between 0 mm and 10 mm in size, and 36% were between 11 mm and 20 mm in size. Furthermore, 78% were pathologically lymph node-negative tumors in comparison to the 66% prevalence observed in the SEER data of the same time period (105).
- Data on 24,740 breast cancer cases recorded in the SEER registry showed that tumor size and axillary lymph node status were two of the most important prognostic indicators (106).

Comments:

- As tumor size increases, survival decreases, regardless of lymph node status.
- As lymph node involvement increases, survival also decreases, regardless of tumor size.
- Other factors, such as patient age, race, hormone receptor status, medical comorbidities, patient preferences, and access to care, also affect survival rates.
- The overall reduction of breast cancer mortality with time supports the beneficial effect of early detection from screening (107).

4.2 Realize that breast-conserving surgery with radiation therapy is equivalent to mastectomy in patients with early-stage disease, and, therefore, both treatment options should be explored.

Evidence:

- A randomized, controlled trial in 701 patients with breast cancer measuring <2 cm in diameter and no palpable axillary lymph nodes compared Halsted radical mastectomy to "quadrantectomy" with axillary dissection and radiotherapy and found no difference in disease-free or overall survival (108).
- Several large cohort studies found no significant differences in overall survival, disease-free survival, or survival free of disease at distant sites between patients who underwent total mastectomy and those treated by lumpectomy alone or by lumpectomy plus breast irradiation with a mean

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

follow-up of 17 years. Radiation therapy following lumpectomy resulted in a significant decrease in the rate of local recurrence of breast cancer (109).

Comments:

- Choice of therapy in a particular patient is based on many factors, including tumor size, axillary node status, tumor hormone status, family history and/or genetic profile (e.g., *BRCA1*, *BRCA2*, *TP53*), patient age, patient comorbidities, patient preference, and access to care.

4.3 Recognize that overtreatment of clinically insignificant breast cancer is possible and may lead to an increase in morbidity.

Evidence:

- A systematic review of six randomized, controlled trials found that screening mammography leads to an estimated 30% increase in overtreatment. For every 2000 women undergoing screening over 10 years, only 1 woman will have her life prolonged, and 10 healthy women will be diagnosed with breast cancer and treated unnecessarily (110).
- Using incidence data from two mammographic screening trials (the Swedish Two-county Trial and the Gothenburg Trial), modeling methods found that fewer than 5% of cases diagnosed at first screen and less than 1% of cases diagnosed at subsequent screens are being overdiagnosed. Overall, approximately 1% of all cases diagnosed in screened populations were estimated to represent overdiagnosis (111).
- Data from the Netherlands show an increase in screening-detected cases of DCIS in women aged 50 to 74 years since the introduction of screening and a decline in incidence at around age 80. Modeling estimated that 3% of the total incidence would otherwise not have been diagnosed clinically (112).

Comments:

- Because the reported incidence of death from breast cancer in patients diagnosed with DCIS is less than 2%, no prospective data currently exist to determine whether there are observed small differences.
- The question of whether DCIS is diagnosed too frequently or treated too aggressively in the U.S. depends on whether these practices result in better outcomes. The outcome of greatest interest, of course, is breast cancer mortality, but because the reported incidence of death from breast cancer in patients diagnosed with DCIS is less than 2%, it will be difficult to detect differences between large populations in which there are multiple variables in addition to the method of diagnosis and treatment that might account for any observed small differences.

4.4 Know that the most common complications of axillary lymph node dissection for early-stage breast cancer are lymphedema, nerve injury, and shoulder dysfunction and that the use of sentinel node biopsy has been associated with a lower risk of these postsurgical complications.

Evidence:

- A randomized, controlled trial compared sentinel node biopsy with axillary lymph node dissection in the management of patients with early-stage breast cancer. Patients were randomly assigned to either standard treatment with axillary lymph node dissection (n=405) or sentinel node biopsy (n=424). At 18 months, the patients who underwent axillary lymph node dissection had experienced more arm swelling (14% vs. 7%) or numbness (19.0% vs. 8.7%) compared to those who underwent sentinel node biopsy. Axillary lymph node dissection was also associated with a decrease in quality of life compared to sentinel node biopsy (113).
- Another randomized, controlled trial in 298 patients with early-stage breast cancer compared axillary lymph node dissection with sentinel node biopsy followed by axillary lymph node dissection if necessary. A significant reduction in postoperative arm swelling, rate of seroma formation, numbness, and loss of sensitivity to light touch and pinprick was observed in the sentinel node biopsy group (114).
- A prospective, multicenter study in Switzerland compared sentinel node biopsy alone with sentinel node biopsy and completion axillary lymph node dissection. A total of 659 patients with early-stage breast cancer were included in the study, 449 of whom underwent sentinel node biopsy alone, and 210 of whom underwent sentinel node biopsy and completion axillary lymph node dissection. Sentinel node biopsy was associated with a lower incidence of lymphedema (3.5% vs. 19.1%), impaired shoulder range of motion (3.5% vs. 11.3%), shoulder/arm pain (8.1% vs. 21.1%), and numbness (10.9% vs. 37.7%) compared to sentinel node biopsy and completion axillary lymph node dissection. Median follow-up was 31 months for patients undergoing sentinel node biopsy alone and 29.5 months for those undergoing sentinel node biopsy and completion axillary lymph node dissection (115).

Comments:

- None.

4.5 Know that the complications of radiation therapy are generally mild and are decreasing with modern techniques. A

Evidence:

- Common short-term side effects of radiation therapy include fatigue and skin erythema (116).
- Late complications of radiation therapy are rare with modern delivery dosing and techniques but include pulmonary fibrosis, brachial plexopathy, lymphedema, ischemic heart disease, and an increased risk of second malignancies (117).

Comments:

- None.

5. Direct Evidence that Screening Reduces Adverse

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

Outcomes

5.1 Understand that screening mammography remains the most studied of all available breast cancer screening modalities and the only modality that is associated with a reduction in breast cancer mortality, which is estimated to be greatest among women aged 50 to 69 years.

Evidence:

- Eight randomized, controlled trials of screening mammography involving over 475,000 women have provided breast cancer mortality data with up to 20 years of follow-up. Results indicate relative risks for breast cancer mortality ranging from 0.76 to 1.02 among those screened (5; 40; 118; 119; 120; 121; 122; 123; 124; 125; 126; 127).
- A controversial systematic review in 2000 and 2001 excluded three trials based on "poor" or "flawed" designs and found no reduction in breast cancer mortality associated with screening mammography for women in any age category (RR, 0.97 [CI, 0.82 to 1.14]). In 2006, an updated systematic review with additional data found a relative risk of 0.80 (CI, 0.73 to 0.88) for six trials combined (110; 128).
- A Swedish study expanded an earlier analysis and compared breast cancer mortality in the prescreening and postscreening periods among women aged 40 to 69 years in six counties and women aged 50 to 69 years in a seventh county. After adjustment for selection bias, there was a mortality reduction from 44% to 39% among women who underwent screening (129).
- A systematic review of eight screening trials conducted by the U.S. Preventive Services Task Force in 2002 found that the use of routine mammography was associated with an overall relative risk of 0.84 for mortality from breast cancer (5). Reductions in breast cancer mortality associated with screening mammography were observed for women aged 39 to 74 years, with greater reductions seen in women aged 50 to 70 years compared to women aged 40 to 49 years (RR, 22% vs. 15%). The decrease in mortality after 14 years of follow-up for women who began screening in their 40s (RR, 0.85 [CI, 0.73 to 0.99]) was lower than that seen for women aged 50 years and older (RR, 0.78 [CI, 0.70 to 0.87]) (5).
- Whether the smaller mortality benefit in women in their 40s is due to screening that occurs before or after the age of 50 is less clear (the "age creep" phenomenon). The UK Age trial, which involved 161,000 women, was one of the first trials to address this issue. The trial compared breast cancer mortality rates among women aged 39 to 41 years who were randomly assigned to either annual mammographic screening or usual care. At 10.7 years of follow-up, there was a nonsignificant trend toward a mortality benefit of mammographic screening (RR, 0.83 [CI, 0.66 to 1.04]) (130).
- The difference in mortality benefit for younger women in their 40s can also be examined by looking at the number needed to screen to prevent

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

one woman from dying of breast cancer. Based on relative risk estimates from the U.S. Preventive Services Task Force review of eight randomized, controlled screening trials, the number needed to screen for women in their 40s (approximately 1800 women undergoing mammography for 14 years) is over twice the estimated number needed to screen (850 women) for women in their 50s. In addition, over half of the women would have at least one false-positive mammography result, requiring additional imaging or biopsy, during this time (5).

- Breast cancer screening also was associated with a reduction in all-cause mortality in an analysis of four randomized, controlled trials conducted in Sweden. These trials followed a combined total of 247,010 women for approximately 16 years, during which time the relative risk for overall mortality was 0.98 (CI, 0.96 to 1.00) (118).

Comments:

- There are no controlled trials evaluating the impact of the newer screening imaging modalities (MRI, ultrasound, and digital mammography), mainly because the ability to study their effects in the absence of screening mammography is not possible.
- Mammography may be less effective for women in their 40s than for older women because the incidence of breast cancer and the overall accuracy of mammography are lower in younger women. If benefit is considered in terms of cumulative years of life saved rather than simply lives saved, this would reflect the benefit of averting premature deaths in younger women who are raising children and/or are active in the workforce. Nonetheless, these smaller benefits need to be weighed against an increase in the risks of mammography in this same age group, such as higher rates of false-positive test results and overdiagnosis of DCIS.

5.2 Know that there are insufficient data from randomized, controlled trials of screening mammography to confirm a mortality benefit among women over age 70. A

Evidence:

- Only two randomized, controlled trials (the Malmö Mammographic Screening Trial and the Swedish Two-county Trial) included women between the ages of 65 and 74 years. When their data are pooled, the summary relative risk in women aged 65 to 74 years is 0.78 (CI, 0.62 to 0.99) (118).

Comments:

- The results of these studies may not be generalizable to the general population based on the extensive comorbidity in these age groups. Women with breast cancer and more than three comorbid medical conditions are 20 times more likely to die of a cause other than breast cancer within 3 years. The effects of comorbidity were independent of age, disease stage, tumor size, histologic type, type of treatment, and race (131).

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

5.3 Acknowledge that the indirect evidence associating screening CBE with a reduction in the breast cancer mortality rate comes from randomized, controlled trials that used both CBE and mammography for breast cancer detection. B

Evidence:

- The Canadian National Breast Screening Study-2 included over 39,000 women aged 50 to 59 years who were randomly assigned to annual screening with mammography and CBE or CBE alone. At 13-year follow-up, there was no impact on breast cancer mortality with the addition of annual mammographic screening to CBE, with 107 deaths in the combined screening group and 105 deaths in the CBE-only group. A total of 622 invasive and 71 in situ cancers were found in the combined screening group, and 610 invasive and 16 in situ cancers were identified in the CBE-only group. Mammography was able to detect a cancer 2.1 years earlier than CBE alone but with no impact on survival. CBE performance in the study was standardized and was longer in duration than most clinical examinations done in practice. There were approximately three times as many biopsies and more diagnostic tests done in the combined screening group (43).
- The Canadian National Breast Screening Study-1 randomly assigned 50,430 women aged 40 to 49 years to screening with mammography and CBE or no screening. CBE detected 59% of the cancers, 32% of which were detected by CBE alone and 27% of which were detected by combined screening. Although there was a more favorable size distribution for cancer detected by mammography alone compared to that detected by CBE, there was no difference in breast cancer mortality after 11 to 16 years of follow-up (42).
- In the National Breast and Cervical Cancer Early Detection Program, data on over 750,000 CBEs done in low-income women in the U.S. showed that CBE detected 5.1% of cancers that were not found on mammography (45).

Comments:

- The independent contribution of CBE and mammography to the reduction in mortality from breast cancer due to screening is unknown, but it is likely that CBE contributes modestly.
- Several organizations, including the American Cancer Society, the American Medical Association, and the Canadian Task Force on Preventive Health Care, continue to recommend combined CBE and mammography for breast cancer screening. The U.S. Preventive Services Task Force recommends mammography every 1 to 2 years beginning at age 40, with or without CBE.
- CBE may also have a role in women who choose not to partake in mammographic screening programs.

5.4 Realize that BSE alone as a screening modality does not reduce mortality from breast cancer. A

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

Evidence:

- In a randomized, controlled trial designed to evaluate the feasibility of BSE in China, there was no reduction in mortality from breast cancer. A total of 266,064 women were randomly assigned to a BSE instruction group or a control group. BSE performance was monitored closely for the 5-year duration of the trial. The total duration of follow-up was 10 years. Compared to women in the control group, women in the BSE instruction group did not have cancers detected at an earlier stage. More breast biopsies were done in the BSE instruction group compared to the control group (132).
- A Russian study randomly assigned 120,310 women aged 40 to 64 years to a BSE program or a control group and found no differences in the stage of breast cancer detection or mortality after 10 years of follow-up. It was noted, however, that more women in the BSE group sought advice on suspected breast lesions than those in the control group (133).
- A meta-analysis of the effect of regular BSE on breast cancer mortality or rates of advanced breast cancer included 20 observational studies and three clinical trials. There was no difference in the death rate in studies in which cancer was detected by BSE (pooled RR, 0.9 [CI, 0.72 to 1.12]), and none of the trials of BSE training showed a lower mortality rate in the BSE group (pooled RR, 1.01 [CI, 0.92 to 1.12]) (134).

Comments:

- The two existing randomized, controlled trials evaluating BSE were conducted in settings without other breast cancer screening modalities, such as mammography. Thus, these trials indicated that BSE confers no breast cancer mortality benefit compared to no screening at all. However, no randomized trial has evaluated whether BSE adds additional mortality benefit when used in conjunction with other screening modalities, such as mammography and CBE.
- The U.S. Preventive Services Task Force states that there is insufficient evidence to recommend for or against performing or teaching BSE.
- The American Cancer Society recommends that providers discuss BSE with their patients and provide appropriate instruction in BSE if the patient chooses this option.
- BSE should be performed using the MammaCare® method.
- Women should be counseled to be aware of changes in their breasts and seek medical advice if they are concerned.

6. Timeline

6.1 Understand that data consistently support the use of routine mammography in women aged 50 to 70 years, but that, due to a lack of data, controversy remains regarding the optimal frequency for all women as well as the appropriate starting and stopping time for women in their 40s and those over age 70, respectively.

Evidence:

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

- Data from a U.S. Preventive Services Task Force review of six clinical trials show that for women over 50, screening mammography can lead to a 22% reduction in breast cancer mortality (5).
- Looking at trial data from women aged 40 to 49 years, a meta-analysis showed a significantly beneficial trend over time (135). The U.S. Preventive Services Task Force review also found a 15% decrease in breast cancer mortality in this age group (5).
- Only two of the major randomized, controlled trials of mammography included women over age 65; these data suggest that screening is beneficial for women up to age 74 (118).
- Optimal frequency of mammographic screening in all age groups is uncertain. Clinical trials evaluating screening mammography used intervals of approximately 18 months, ranging from 12 to 33 months (5; 50).
- There are well-documented age differences in growth rates of primary breast cancer, such that women under age 50 have the shortest tumor volume doubling time (80 days) compared with women aged 50 to 70 years (157 days) and those over age 70 (188 days). To observe a beneficial effect of screening in women under age 50, more frequent screening than in the older age group is necessary (136).
- A framework has been developed positing that patients with life expectancies of less than 5 years are unlikely to derive any survival benefit from cancer screening (137).

Comments:

- The interval between the time when a tumor develops and when it becomes clinically significant is thought to increase with age (the sojourn time). This may mean that early detection is most beneficial in younger women, and less beneficial in older age groups.

6.2 Understand that screening intervals for certain high-risk patients, such as those with a positive family history, a personal history of atypical ductal hyperplasia or LCIS, or a history of thoracic radiation, are often tailored and based on expert opinion.

- See table Breast Cancer Screening Guidelines by Screening Modality.

Evidence:

- The optimal age to begin screening women with a family history of breast cancer is unknown (138).
- Current recommendations for women with an inherited predisposition to breast cancer include annual mammography and breast MRI starting at age 30 (23; 78).
- It is currently recommended that women with a history of atypical ductal hyperplasia or LCIS begin annual mammographic screening at the time of diagnosis, regardless of age, if not already initiated. Such patients should also consider risk reduction strategies, such as chemoprevention (78).
- It is currently recommended that women with a history of thoracic radiation exposure begin annual mammography 8 to 10 years after

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

radiation therapy or at age 40, whichever occurs first (78).

Comments:
- Currently there is no evidence for the use of early or alternative breast cancer screening regimens in women with other traditional risk factors for breast cancer, such as early menarche, late menopause, or dense breasts on mammography.
- Women with an inherited predisposition to breast cancer should consider medical and surgical risk reduction strategies for management of breast and other associated cancer risks.
- Women with a history of atypical ductal hyperplasia or LCIS also should consider risk reduction strategies, such as chemoprevention (78). See module Chemoprevention of Breast Cancer.

6.3 Be aware that annual MRI is recommended for women over age 30 with a genetic mutation or beginning 5 to 10 years younger than the youngest family member with breast cancer when the lifetime risk is greater than 25%. (B)

Evidence:
- It is currently recommended that women with genetic mutations begin screening with annual mammography and MRI at age 25 (23; 78).
- It is currently recommended that women with a family history of premenopausal cancer but no identified genetic mutations begin screening 5 to 10 years younger than the youngest family member with breast cancer (78).
- If the lifetime risk of breast cancer is 20% to 25%, annual MRI is recommended in addition to annual mammography (23; 78).

Comments:
- The optimal age to initiate breast cancer screening in high-risk women is unknown.
- Additional information on genetic/familial high-risk assessment for breast cancer is available from the National Comprehensive Cancer Network (http://www.nccn.org/professionals/physician_gls/PDF/genetics_screening.pdf).

7. Cost-Effectiveness

7.1 Appreciate that screening mammography has an acceptable cost per life-year saved compared with other screening strategies. (B)

Evidence:
- Many studies have been conducted to evaluate the cost-effectiveness of screening mammography, with the cost per life-year saved ranging from $18,800 to $20,200 (139; 140).
- A Markov model-based study evaluating the cost-effectiveness of four mammographic screening schedules that varied in terms of patient age and frequency found that all were within a generally accepted range of

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

cost per life-year saved compared with other screening programs (less than $40,000 per life-year saved) (140). The most cost-effective schedule involved annual mammographic screening from age 40 to 49 years followed by biennial mammographic screening from age 50 to 79 years, which resulted in a marginal cost per life-year saved of $16,100 (141).
- Age-targeted use of digital mammography appears to be cost effective; however, breast density-targeted digital mammography is not (142).

Comments:
- None.

7.2 Recognize that the cost-effectiveness of breast cancer screening is affected by the frequency of mammography, the age of the screened population, and whether mammography is coupled with CBE. (B)

Evidence:
- A study evaluating the cost-effectiveness of screening mammography using a Markov model noted that the marginal cost per life-year saved is lowest in women aged 50 to 69 years and higher in women aged 40 to 49 years due to the lower incidence of breast cancer in the younger age group. Cost-effectiveness is further decreased in women aged 80 to 84 years despite the increased breast cancer rate in the elderly because of the concurrent shorter life expectancy (143).
- A systematic literature review concluded that extending biennial mammographic screening for women after age 65 up through age 75 or 80 would be at a cost of $34,000 to $88,000 per life-year gained. The most cost-effective approach to screening older women was one that targeted healthy women rather than those with competing life-shortening illnesses (144).
- A retrospective analysis of costs incurred in current U.S. breast cancer screening programs recommending annual to biennial mammography for women aged 40 years and older determined that over 10 years, 947.5 million quality-adjusted life-years resulted at a cost of $166 billion over the screened women's lifetimes (145). Among the screening scenarios examined, annual mammography from age 40 to 80 was the most expensive, costing $58,000 per additional quality-adjusted life-year gained compared with alternative strategies with longer intervals between screenings. Cost-effectiveness was sensitive to factors affecting quality of life, such as false-positive test results and mammography-related pain.
- Annual mammography and CBE for women aged 50 to 79 years reduces total costs by 35% as compared to biennial mammography and annual CBE (146).

Comments:
- Published cost-effectiveness estimates vary and are influenced by differences in study methodologies, assumptions, and population

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

characteristics.

8. Patient Education

8.1 Know that not all women are well informed about breast cancer screening. B

Evidence:

- Many women have inaccurate knowledge, attitudes, and beliefs surrounding breast cancer screening. A survey of rural women concluded that, overall, a significant proportion of women had low levels of breast cancer screening knowledge and inaccurate beliefs regarding screening. In particular, minority women tended to have the lowest levels of knowledge regarding breast cancer screening (147).
- Several interventions have been shown to improve women's knowledge of breast cancer screening, including mailed educational pamphlets (148).

Comments:

- Population surveys and persistent disparities in breast cancer screening rates indicate that significant groups of women remain poorly informed about the benefits of screening.

8.2 Know that a personal recommendation from a health care provider is the best method to increase mammography adherence; direct patient education interventions have shown a minimal or no increase in mammography adherence. B

Evidence:

- Published studies indicate that a physician recommendation remains critical in encouraging mammography compliance for all women, regardless of age and ethnic group (149; 150).
- Several randomized, controlled trials of brief, office-based patient education have shown minimal or no benefit (all ORs <1.8) (151; 152; 153; 154; 155).
- One study showed a moderate benefit of multiple educational interventions tailored to individual patients (OR, 1.93 to 3.55) (156).
- An effect size synthesis of nine randomized, controlled trials specifically educating patients on their personalized cancer risk showed a weak benefit (OR, 1.31 [CI, 0.98 to 1.77]) on subsequent screening adherence (157).
- Interventions to increase physician recommendations, including audit and physician reminders, were found to have a stronger effect on mammography adherence than patient education interventions (158).

Comments:

- None.

9. Referral/Consultation

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

9.1 Consider referral to a breast specialist for certain high-risk patients, such as those with a family history of breast cancer or a personal history of atypical ductal hyperplasia, LCIS, or thoracic radiation. (A)

Evidence:

- Patients at high risk for breast cancer include those with a family history of breast cancer in a first-degree relative, especially if the cancer was premenopausal, male breast cancer, or bilateral breast cancer (8); those with a family history of two or more second-degree relatives (8); and those with a personal history of atypical ductal hyperplasia or LCIS (21; 22).

Comments:

- None.

10. Guidelines

American Cancer Society, 2008

American Cancer Society guidelines for the early detection of cancer

These guidelines recommend discussing monthly BSE beginning at age 20 (patient's choice), performing CBE every 3 years from age 20 to 39 years and yearly after age 40, and obtaining mammography yearly beginning at age 40. Women at high risk (greater than 20% lifetime risk) should undergo mammography and MRI yearly starting at age 30. Women at moderately increased risk (15% to 20%) should discuss the option of MRI with a yearly mammogram

American Cancer Society, 2007

American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography

These guidelines recommend discussing the option of MRI in addition to yearly mammography with women at moderately increased risk (15% to 20%).

American College of Obstetricians and Gynecologists, 2003

Breast cancer screening

This guideline recommends CBE yearly as part of the physical exam for all women and mammography every 1 to 2 years for women aged 40 to 49 years and yearly after age 50. BSE is recommended despite a lack of definitive data. No information is provided for MRI.

American College of Physicians, 2007

Screening mammography for women 40 to 49 years of age: a clinical practice guideline from the American College of Physicians

This guideline recommends that the decision to obtain mammography in women aged 40 to 49 years be made on an individualized basis using risk assessment and shared decision-making. For women aged 40 to 49 years who do not wish to partake in shared decision-making, mammography should be done every 1 to 2 years. No information is provided for MRI.

Canadian Task Force on Preventive Health Care, 2001

Preventive health care, 2001 update: screening mammography among women aged 40-49 years at average risk of breast cancer

Based on this review, mammography every 1 to 2 years is recommended for women over age 50, and CBE with mammography every 1 to 2 years is recommended for women aged

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes 2009-10-28

50 to 69 years. CBE with mammography is not recommended for women aged 40 to 49 years, and not enough evidence was found to recommend for or against mammography every 12 to 18 months for average-risk women aged 40 to 49 years. BSE is not recommended. No information is provided for MRI.

National Comprehensive Cancer Network, 2008

Breast cancer screening and diagnosis guidelines (http://www.nccn.org/professionals/physician_gls/PDF/breast-screening.pdf)
These guidelines recommend periodic BSE beginning at age 20 and CBE and mammography yearly for women aged 40 years and older.

Genetic/familial high-risk assessment: breast and ovarian (http://www.nccn.org/professionals/physician_gls/PDF/genetics_screening.pdf)
These guidelines supply criteria for recommending further genetic evaluation for breast cancer.

U.S. Preventive Services Task Force, 2002

Screening for breast cancer: recommendations and rationale
The U.S. Preventive Services Task Force recommends obtaining mammography every 1 to 2 years beginning at age 40, with or without CBE. Not enough evidence was found to recommend for or against CBE alone or to recommend for or against teaching or performing BSE. No information is provided for MRI.

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

(1) Studies that meet all of the evidence criteria for that study type

(2) Studies that meet at least one of the criteria for that study type

(3) Studies that meet none of the evidence criteria for that study type or are derived from expert opinion, commentary, or consensus

Study types and evidence criteria are defined at http://pier.acponline.org/criteria.html

The number in parentheses at the end of the reference citations identify PubMed abstracts, which can be found on the National Library of Medicine's web site http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

References

1 (3) Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, et al. Cancer statistics, 2008. CA Cancer J Clin. 2008;58:71-96. (PMID: 18287387)

2 (3) American Cancer Society. Cancer Facts & Figures 2008. Atlanta: American Cancer Society; 2008. [Full Text (http://www.cancer.org/downloads/STT/2008CAFFfinalsecured.pdf)]

3 (3) American Cancer Society. Breast Cancer Facts & Figures 2007-2008. Atlanta: American Cancer Society, Inc. [Full Text (http://www.cancer.org/downloads/STT/BCFF-Final.pdf)]

4 (3) Surveillance Epidemiology and End Results (SEER), National Cancer Institute. SEER Stat Fact Sheets: Cancer of the Breast. Available at: http://seer.cancer.gov/statfacts/html/breast.html. Accessed 24 June 2008.

5 (1) Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2002;137:347-60. (PMID: 12204020) [Full Text]

6 (3) Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Proportion of breast cancer cases in the United States explained by well-established risk factors. J Natl Cancer Inst. 1995;87:1681-5. (PMID: 7473816)

7 (1) Kerlikowske K, Carney PA, Geller B, Mandelson MT, Taplin SH, Malvin K, et al. Performance of screening mammography among women with and without a first-degree relative with breast cancer. Ann Intern Med. 2000;133:855-63. (PMID: 11103055) [Full Text]

8 (1) Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. Int J Cancer. 1997;71:800-9. (PMID: 9180149)

9 (2) Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. Cancer. 1994;73:643-51. (PMID: 8299086)

10 (1) Garber JE, Offit K. Hereditary cancer predisposition syndromes. J Clin Oncol. 2005;23:276-92. (PMID: 15637391)

11 (2) Tonin P, Weber B, Offit K, Couch F, Rebbeck TR, Neuhausen S, et al. Frequency of recurrent BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer families. Nat Med. 1996;2:1179-83. (PMID: 8898735)

12 (1) Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet. 2003;72:1117-30. (PMID: 12677558)

13 (2) Andrieu N, Easton DF, Chang-Claude J, Rookus MA, Brohet R, Cardis E, et al. Effect of chest X-rays on the risk of breast cancer among BRCA1/2 mutation carriers in the international BRCA1/2 carrier cohort study: a report from the EMBRACE, GENEPSO, GEO-HEBON, and IBCCS Collaborators' Group. J Clin Oncol. 2006;24:3361-6. (PMID: 16801631)

14 (2) Berry DA, Iversen ES Jr, Gudbjartsson DF, Hiller EH, Garber JE, Peshkin BN, et al. BRCAPRO validation, sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. J Clin Oncol. 2002;20:2701-12. (PMID: 12039933)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD;
Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

15 Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med. 2002;347:1233-41. (PMID: 12393820)

16 Veronesi U, Marubini E, Del Vecchio M, Manzari A, Andreola S, Greco M, et al. Local recurrences and distant metastases after conservative breast cancer treatments: partly independent events. J Natl Cancer Inst. 1995;87:19-27. (PMID: 7666458)

17 Saphner T, Tormey DC, Gray R. Annual hazard rates of recurrence for breast cancer after primary therapy. J Clin Oncol. 1996;14:2738-46. (PMID: 8874335)

18 Chen Y, Thompson W, Semenciw R, Mao Y. Epidemiology of contralateral breast cancer. Cancer Epidemiol Biomarkers Prev. 1999;8:855-61. (PMID: 10548312)

19 Kattlove H, Winn RJ. Ongoing care of patients after primary treatment for their cancer. CA Cancer J Clin. 2003;53:172-96. (PMID: 12809410)

20 Kenney LB, Yasui Y, Inskip PD, Hammond S, Neglia JP, Mertens AC, et al. Breast cancer after childhood cancer: a report from the Childhood Cancer Survivor Study. Ann Intern Med. 2004;141:590-7. (PMID: 15492338) [Full Text]

21 Arpino G, Laucirica R, Elledge RM. Premalignant and in situ breast disease: biology and clinical implications. Ann Intern Med. 2005;143:446-57. (PMID: 16172443) [Full Text]

22 Fisher ER, Land SR, Fisher B, Mamounas E, Gilarski L, Wolmark N. Pathologic findings from the National Surgical Adjuvant Breast and Bowel Project: twelve-year observations concerning lobular carcinoma in situ. Cancer. 2004;100:238-44. (PMID: 14716756)

23 Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin. 2007;57:75-89. (PMID: 17392385)

24 Baquet CR, Mishra SI, Commiskey P, Ellison GL, DeShields M. Breast cancer epidemiology in blacks and whites: disparities in incidence, mortality, survival rates and histology. J Natl Med Assoc. 2008;100:480-8. (PMID: 18507200)

25 Haynes MA, Smedley BD, eds. The Unequal Burden of Cancer: An Assessment of NIH Research and Programs for Ethnic Minorities and the Medically Underserved. Washington, DC: National Academy Press; 1999.

26 Clemons M, Goss P. Estrogen and the risk of breast cancer. N Engl J Med. 2001;344:276-85. (PMID: 11172156)

27 Hulka BS. Epidemiologic analysis of breast and gynecologic cancers. Prog Clin Biol Res. 1997;396:17-29. (PMID: 9108587)

28 MacMahon B, Cole P, Lin TM, Lowe CR, Mirra AP, Ravnihar B, et al. Age at first birth and breast cancer risk. Bull World Health Organ. 1970;43:209-21. (PMID: 5312521)

29 Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet. 1996;347:1713-27. (PMID: 8656904)

30 Kahlenborn C, Modugno F, Potter DM, Severs WB. Oral contraceptive use as a risk factor for premenopausal breast cancer: a meta-analysis. Mayo Clin Proc. 2006;81:1290-302. (PMID: 17036554)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD;
Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

31 Marchbanks PA, McDonald JA, Wilson HG, Folger SG, Mandel MG, Daling JR, et al. Oral contraceptives and the risk of breast cancer. N Engl J Med. 2002;346:2025-32. (PMID: 12087137)

32 Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA. 2002;288:321-33. (PMID: 12117397)

33 O'Malley MS, Fletcher SW. US Preventive Services Task Force. Screening for breast cancer with breast self-examination. A critical review. JAMA. 1987;257:2196-203. (PMID: 3550165)

34 Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst. 1993;85:1644-56. (PMID: 8105098)

35 Fletcher SW, O'Malley MS, Pilgrim CA, Gonzalez JJ. How do women compare with internal medicine residents in breast lump detection? A study with silicone models. J Gen Intern Med. 1989;4:277-83. (PMID: 2788213)

36 Fletcher SW, O'Malley MS, Earp JL, Morgan TM, Lin S, Degnan D. How best to teach women breast self-examination. A randomized controlled trial. Ann Intern Med. 1990;112:772-9. (PMID: 2184711)

37 Newcomb PA, Weiss NS, Storer BE, Scholes D, Young BE, Voigt LF. Breast self-examination in relation to the occurrence of advanced breast cancer. J Natl Cancer Inst. 1991;83:260-5. (PMID: 1994055)

38 von Georgi R, Thele F, Hackethal A, Münstedt K. Can an instruction video or palpation aid improve the effectiveness of breast self-examination in detecting tumors? An experimental study. Breast Cancer Res Treat. 2006;97:167-72. (PMID: 16322886)

39 Barton MB, Harris R, Fletcher SW. The rational clinical examination. Does this patient have breast cancer? The screening clinical breast examination: should it be done? How? JAMA. 1999;282:1270-80. (PMID: 10517431)

40 Shapiro S. Periodic screening for breast cancer: the HIP Randomized Controlled Trial. Health Insurance Plan. J Natl Cancer Inst Monogr. 1997:27-30. (PMID: 9709271)

41 Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, et al. The Edinburgh randomised trial of breast cancer screening: results after 10 years of follow-up. Br J Cancer. 1994;70:542-8. (PMID: 8080744)

42 Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. Ann Intern Med. 2002;137:305-12. (PMID: 12204013) [Full Text]

43 Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. J Natl Cancer Inst. 2000;92:1490-9. (PMID: 10995804)

44 Shen Y, Zelen M. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. J Clin Oncol. 2001;19:3490-9. (PMID: 11481355)

45 Bobo JK, Lee NC, Thames SF. Findings from 752,081 clinical breast examinations reported to a national screening program from 1995 through 1998. J Natl Cancer Inst. 2000;92:971-6. (PMID: 10861308)

46 Oestreicher N, Lehman CD, Seger DJ, Buist DS, White E. The incremental contribution of clinical breast examination to invasive cancer detection in a mammography screening program. AJR Am J Roentgenol. 2005;184:428-32. (PMID: 15671358)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

47 Oestreicher N, White E, Lehman CD, Mandelson MT, Porter PL, Taplin SH. Predictors of sensitivity of clinical breast examination (CBE). Breast Cancer Res Treat. 2002;76:73-81. (PMID: 12408378)

48 Saunders KJ, Pilgrim CA, Pennypacker HS. Increased proficiency of search in breast self-examination. Cancer. 1986;58:2531-7. (PMID: 3768844)

49 Saslow D, Hannan J, Osuch J, Alciati MH, Baines C, Barton M, et al. Clinical breast examination: practical recommendations for optimizing performance and reporting. CA Cancer J Clin. 2004;54:327-44. (PMID: 15537576)

50 Mushlin AI, Kouides RW, Shapiro DE. Estimating the accuracy of screening mammography: a meta-analysis. Am J Prev Med. 1998;14:143-53. (PMID: 9631167)

51 Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA. 1996;276:33-8. (PMID: 8667536)

52 Kerlikowske K, Grady D, Barclay J, Sickles EA, Eaton A, Ernster V. Positive predictive value of screening mammography by age and family history of breast cancer. JAMA. 1993;270:2444-50. (PMID: 8230621)

53 Laya MB, Larson EB, Taplin SH, White E. Effect of estrogen replacement therapy on the specificity and sensitivity of screening mammography. J Natl Cancer Inst. 1996;88:643-9. (PMID: 8627640)

54 Greendale GA, Reboussin BA, Sie A, Singh HR, Olson LK, Gatewood O, et al. Effects of estrogen and estrogen-progestin on mammographic parenchymal density. Postmenopausal Estrogen/Progestin Interventions (PEPI) Investigators. Ann Intern Med. 1999;130:262-9. (PMID: 10068383) [Full Text]

55 Marugg RC, van der Mooren MJ, Hendriks JH, Rolland R, Ruijs SH. Mammographic changes in postmenopausal women on hormonal replacement therapy. Eur Radiol. 1997;7:749-55. (PMID: 9166577)

56 Kavanagh AM, Cawson J, Byrnes GB, Giles GG, Marr G, Tong B, et al. Hormone replacement therapy, percent mammographic density, and sensitivity of mammography. Cancer Epidemiol Biomarkers Prev. 2005;14:1060-4. (PMID: 15894653)

57 Evans A. Hormone replacement therapy and mammographic screening. Clin Radiol. 2002;57:563-4. (PMID: 12096852)

58 Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading—Oslo I study. Radiology. 2003;229:877-84. (PMID: 14576447)

59 Lewin JM, D'Orsi CJ, Hendrick RE, Moss LJ, Isaacs PK, Karellas A, et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. AJR Am J Roentgenol. 2002;179:671-7. (PMID: 12185042)

60 Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program—the Oslo II Study. Radiology. 2004;232:197-204. (PMID: 15155893)

61 Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005;353:1773-83. (PMID: 16169887)

62 Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med. 2007;356:1399-409. (PMID: 17409321)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD;
Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

63 Gilbert FJ, Astley SM, McGee MA, Gillan MG, Boggis CR, Griffiths PM, et al. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. Radiology. 2006;241:47-53. (PMID: 16990670)

64 Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. Radiology. 2001;220:781-6. (PMID: 11526282)

65 Gur D, Sumkin JH, Rockette HE, Ganott M, Hakim C, Hardesty L, et al. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. J Natl Cancer Inst. 2004;96:185-90. (PMID: 14759985)

66 Kriege M, Brekelmans CT, Boetes C, Besnard PE, Zonderland HM, Obdeijn IM, et al. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. N Engl J Med. 2004;351:427-37. (PMID: 15282350)

67 Warner E, Plewes DB, Hill KA, Causer PA, Zubovits JT, Jong RA, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. JAMA. 2004;292:1317-25. (PMID: 15367553)

68 Leach MO, Boggis CR, Dixon AK, Easton DF, Eeles RA, Evans DG, et al. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). Lancet. 2005;365:1769-78. (PMID: 15910949)

69 Kuhl CK, Schrading S, Leutner CC, Morakkabati-Spitz N, Wardelmann E, Fimmers R, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. J Clin Oncol. 2005;23:8469-76. (PMID: 16293877)

70 Warner E, Messersmith H, Causer P, Eisen A, Shumak R, Plewes D. Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer. Ann Intern Med. 2008;148:671-9. (PMID: 18458280) [Full Text]

71 Warner E, Plewes DB, Shumak RS, Catzavelos GC, Di Prospero LS, Yaffe MJ, et al. Comparison of breast magnetic resonance imaging, mammography, and ultrasound for surveillance of women at high risk for hereditary breast cancer. J Clin Oncol. 2001;19:3524-31. (PMID: 11481359)

72 Trecate G, Vergnaghi D, Manoukian S, Bergonzi S, Scaperrotta G, Marchesini M, et al. MRI in the early detection of breast cancer in women with high genetic risk. Tumori. 2006;92:517-23. (PMID: 17260493)

73 Hartman AR, Daniel BL, Kurian AW, Mills MA, Nowels KW, Dirbas FM, et al. Breast magnetic resonance image screening and ductal lavage in women at high genetic risk for breast carcinoma. Cancer. 2004;100:479-89. (PMID: 14745863)

74 Lehman CD, Blume JD, Weatherall P, Thickman D, Hylton N, Warner E, et al; International Breast MRI Consortium Working Group. Screening women at high risk for breast cancer with mammography and magnetic resonance imaging. Cancer. 2005;103:1898-905. (PMID: 15800894)

75 Sardanelli F, Podo F, D'Agnolo G, Verdecchia A, Santaquilani M, Musumeci R, et al. Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT study): interim results. Radiology. 2007;242:698-715. (PMID: 17244718)

76 Lehman CD, Isaacs C, Schnall MD, Pisano ED, Ascher SM, Weatherall PT, et al Cancer yield of mammography, MR, and US in high-risk women: prospective multi-institution breast cancer screening study. Radiology. 2007;244:381-8. (PMID: 17641362)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

77 Hagen AI, Kvistad KA, Maehle L, Holmen MM, Aase H, Styr B, et al. Sensitivity of MRI versus conventional screening in the diagnosis of BRCA-associated breast cancer in a national prospective series. Breast. 2007;16:367-74. (PMID: 17317184)

78 National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology™. Breast Cancer Screening and Diagnosis Guidelines. V.1.2008. Available at: http://www.nccn.org/professionals/physician_gls/PDF/breast-screening.pdf (http://www.nccn.org/professionals/physician_gls/PDF/breast-screening.pdf). Accessed 27 Feb 2008.

79 Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Böhm-Vélez M, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. JAMA. 2008;299:2151-63. (PMID: 18477782)

80 Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med. 1998;338:1089-96. (PMID: 9545356)

81 Christiansen CL, Wang F, Barton MB, Kreuter W, Elmore JG, Gelfand AE, et al. Predicting the cumulative risk of false-positive mammograms. J Natl Cancer Inst. 2000;92:1657-66. (PMID: 11036111)

82 Warner E, Causer PA. MRI surveillance for hereditary breast-cancer risk. Lancet. 2005;365:1747-9. (PMID: 15910935)

83 Kuhl CK, Schmutzler RK, Leutner CC, Kempe A, Wardelmann E, Hocke A, et al. Breast MR imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results. Radiology. 2000;215:267-79. (PMID: 10751498)

84 Feig SA, Hendrick RE. Radiation risk from screening mammography of women aged 40-49 years. J Natl Cancer Inst Monogr. 1997:119-24. (PMID: 9709287)

85 Narod SA, Lubinski J, Ghadirian P, Lynch HT, Moller P, Foulkes WD, et al. Screening mammography and risk of breast cancer in BRCA1 and BRCA2 mutation carriers: a case-control study. Lancet Oncol. 2006;7:402-6. (PMID: 16648044)

86 Kopans DB. Mammography and radiation risk. In: American College of Radiology. Radiation Risk: A Primer. Reston, VA: American College of Radiology; 1996.

87 Goss PE, Sierra S. Current perspectives on radiation-induced breast cancer. J Clin Oncol. 1998;16:338-47. (PMID: 9440762)

88 Barton MB, Morley DS, Moore S, Allen JD, Kleinman KP, Emmons KM, et al. Decreasing women's anxieties after abnormal mammograms: a controlled trial. J Natl Cancer Inst. 2004;96:529-38. (PMID: 15069115)

89 Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med. 1991;114:657-61. (PMID: 2003712)

90 Gram IT, Lund E, Slenker SE. Quality of life following a false positive mammogram. Br J Cancer. 1990;62:1018-22. (PMID: 2257206)

91 Brett J, Austoker J. Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. J Public Health Med. 2001;23:292-300. (PMID: 11873891)

92 Aro AR, Pilvikki Absetz S, van Elderen TM, van der Ploeg E, van der Kamp LJ. False-positive findings in mammography screening induces short-term distress - breast cancer-specific concern prevails longer. Eur J Cancer. 2000;36:1089-97. (PMID: 10854941)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD;
Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

93 Lampic C, Thurfjell E, Bergh J, Sjödén PO. Short- and long-term anxiety and depression in women recalled after breast cancer screening. Eur J Cancer. 2001;37:463-9. (PMID: 11267855)

94 Schwartz LM, Woloshin S, Sox HC, Fischhoff B, Welch HG. US women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. BMJ. 2000;320:1635-40. (PMID: 10856064)

95 Steggles S, Lightfoot N, Sellick SM. Psychological distress associated with organized breast cancer screening. Cancer Prev Control. 1998;2:213-20. (PMID: 10093635)

96 Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med. 2007;146:502-10. (PMID: 17404352) [Full Text]

97 Keemers-Gels ME, Groenendijk RP, van den Heuvel JH, Boetes C, Peer PG, Wobbes TH. Pain experienced by women attending breast cancer screening. Breast Cancer Res Treat. 2000;60:235-40. (PMID: 10930111)

98 Rutter DR, Calnan M, Vaile MS, Field S, Wade KA. Discomfort and pain during mammography: description, prediction, and prevention. BMJ. 1992;305:443-5. (PMID: 1392955)

99 Aro AR, Absetz-Ylöstalo P, Eerola T, Pamilo M, Lönnqvist J. Pain and discomfort during mammography. Eur J Cancer. 1996;32A:1674-9. (PMID: 8983273)

100 van Steenbergen LN, Voogd AC, Roukema JA, Louwman WJ, Duijm LE, Coebergh JW, et al. Screening caused rising incidence rates of ductal carcinoma in situ of the breast. Breast Cancer Res Treat. 2008;115:181-3. (PMID: 18516674)

101 Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, Cutter G, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. J Natl Cancer Inst. 2002;94:1546-54. (PMID: 12381707)

102 Kosary CL, Ries LAG, Miller BA, Hankey BF, Harras A, Edwards BK, eds. SEER Cancer Statistics Review, 1973-1992: Tables and Graphs. DHHS Publication No. NIH 96-2789. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health; 1995.

103 Miller BA, Feuer EJ, Hankey BF. Recent incidence trends for breast cancer in women and the relevance of early detection: an update. CA Cancer J Clin. 1993;43:27-41. (PMID: 8422603)

104 Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: Incidence - SEER 17 Regs Public-Use, Nov 2005 Sub (1973-2003 varying), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2006, based on the November 2005 submission.

105 Weaver DL, Rosenberg RD, Barlow WE, Ichikawa L, Carney PA, Kerlikowske K, et al. Pathologic findings from the Breast Cancer Surveillance Consortium: population-based outcomes in women undergoing biopsy after screening mammography. Cancer. 2006;106:732-42. (PMID: 16411214)

106 Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer. 1989;63:181-7. (PMID: 2910416)

107 Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005;353:1784-92. (PMID: 16251534)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD;
Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

108 Veronesi U, Saccozzi R, Del Vecchio M, Banfi A, Clemente C, De Lena M, et al. Comparing radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in patients with small cancers of the breast. N Engl J Med. 1981;305:6-11. (PMID: 7015141)

109 Fisher B, Anderson S, Redmond CK, Wolmark N, Wickerham DL, Cronin WM. Reanalysis and results after 12 years of follow-up in a randomized clinical trial comparing total mastectomy with lumpectomy with or without irradiation in the treatment of breast cancer. N Engl J Med. 1995;333:1456-61. (PMID: 7477145)

110 Olsen O, Gøtzsche PC. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2001:CD001877. (PMID: 11687128)

111 Duffy SW, Agbaje O, Tabar L, Vitak B, Bjurstam N, Björneld L, et al. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. Breast Cancer Res. 2005;7:258-65. (PMID: 16457701)

112 de Koning HJ, Draisma G, Fracheboud J, de Bruijn A. Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. Breast Cancer Res. 2006;8:202. (PMID: 16524452)

113 Fleissig A, Fallowfield LJ, Langridge CI, Johnson L, Newcombe RG, Dixon JM, et al. Postoperative arm morbidity and quality of life. Results of the ALMANAC randomised trial comparing sentinel node biopsy with standard axillary treatment in the management of patients with early breast cancer. Breast Cancer Res Treat. 2006;95:279-93. (PMID: 16163445)

114 Purushotham AD, Upponi S, Klevesath MB, Bobrow L, Millar K, Myles JP, et al. Morbidity after sentinel lymph node biopsy in primary breast cancer: results from a randomized controlled trial. J Clin Oncol. 2005;23:4312-21. (PMID: 15994144)

115 Langer I, Guller U, Berclaz G, Koechli OR, Schaer G, Fehr MK, et al. Morbidity of sentinel lymph node biopsy (SLN) alone versus SLN and completion axillary lymph node dissection after breast cancer surgery: a prospective Swiss multicenter study on 659 patients. Ann Surg. 2007;245:452-61. (PMID: 17435553)

116 Truong PT, Olivotto IA, Whelan TJ, Levine M; Steering Committee on Clinical Practice Guidelines for the Care and Treatment of Breast Cancer. Clinical practice guidelines for the care and treatment of breast cancer: 16. Locoregional postmastectomy radiotherapy. CMAJ. 2004;170:1263-73. (PMID: 15078851)

117 Højris I, Andersen J, Overgaard M, Overgaard J. Late treatment-related morbidity in breast cancer patients randomized to postmastectomy radiotherapy and systemic treatment versus systemic treatment alone. Acta Oncol. 2000;39:355-72. (PMID: 10987233)

118 Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet. 2002;359:909-19. (PMID: 11918907)

119 Tabár L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF, et al. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. Radiol Clin North Am. 2000;38:625-51. (PMID: 10943268)

120 Bjurstam N, Björneld L, Duffy SW, Smith TC, Cahlin E, Eriksson O, et al. The Gothenburg breast screening trial: first results on mortality, incidence, and mode of detection for women ages 39-49 years at randomization. Cancer. 1997;80:2091-9. (PMID: 9392331)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

121 Frisell J, Lidbrink E, Hellström L, Rutqvist LE. Followup after 11 years—update of mortality results in the Stockholm mammographic screening trial. Breast Cancer Res Treat. 1997;45:263-70. (PMID: 9386870)

122 Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmö Mammographic Screening Program. J Natl Cancer Inst Monogr. 1997:63-7. (PMID: 9709278)

123 Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. CMAJ. 1992;147:1459-76. (PMID: 1423087)

124 Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. CMAJ. 1992;147:1477-88. (PMID: 1423088)

125 Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study: update on breast cancer mortality. J Natl Cancer Inst Monogr. 1997:37-41. (PMID: 9709273)

126 Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. Lancet. 1999;353:1903-8. (PMID: 10371567)

127 Nyström L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Rydén S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. Lancet. 1993;341:973-8. (PMID: 8096941)

128 Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2006:CD001877. (PMID: 17054145)

129 Duffy SW, Tabár L, Chen HH, Holmqvist M, Yen MF, Abdsalah S, et al. The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties. Cancer. 2002;95:458-69. (PMID: 12209737)

130 Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet. 2006;368:2053-60. (PMID: 17161727)

131 Satariano WA, Ragland DR. The effect of comorbidity on 3-year survival of women with primary breast cancer. Ann Intern Med. 1994;120:104-10. (PMID: 8256968) [Full Text]

132 Thomas DB, Gao DL, Ray RM, Wang WW, Allison CJ, Chen FL, et al. Randomized trial of breast self-examination in Shanghai: final results. J Natl Cancer Inst. 2002;94:1445-57. (PMID: 12359854)

133 Semiglazov VF, Sagaidak VN, Moiseyenko VM, Mikhailov EA. Study of the role of breast self-examination in the reduction of mortality from breast cancer. The Russian Federation/World Health Organization Study. Eur J Cancer. 1993;29A:2039-46. (PMID: 8280499)

134 Hackshaw AK, Paul EA. Breast self-examination and death from breast cancer: a meta-analysis. Br J Cancer. 2003;88:1047-53. (PMID: 12671703)

135 Glasziou PP. Meta-analysis adjusting for compliance: the example of screening for breast cancer. J Clin Epidemiol. 1992;45:1251-6. (PMID: 1432006)

136 Peer PG, van Dijck JA, Hendriks JH, Holland R, Verbeek AL. Age-dependent growth rate of primary breast cancer. Cancer. 1993;71:3547-51. (PMID: 8490903)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD;
Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

137 Walter LC, Covinsky KE. Cancer screening in elderly patients: a framework for individualized decision making. JAMA. 2001;285:2750-6. (PMID: 11386931)

138 Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2008: a review of current American Cancer Society guidelines and cancer screening issues. CA Cancer J Clin. 2008;58:161-79. (PMID: 18443206)

139 Lindfors KK, Rosenquist CJ. The cost-effectiveness of mammographic screening strategies. JAMA. 1995;274:881-4. (PMID: 7674501)

140 Rosenquist CJ, Lindfors KK. Screening mammography beginning at age 40 years: a reappraisal of cost-effectiveness. Cancer. 1998;82:2235-40. (PMID: 9610704)

141 Holland RR, Ellis CA, Geller BM, Plante DA, Secker-Walker RH. Life expectancy estimation with breast cancer: bias of the declining exponential function and an alternative to its use. Med Decis Making. 1999;19:385-93. (PMID: 10520676)

142 Tosteson AN, Stout NK, Fryback DG, Acharyya S, Herman BA, Hannah LG, et al. Cost-effectiveness of digital mammography breast cancer screening. Ann Intern Med. 2008;148:1-10. (PMID: 18166758) [Full Text]

143 Rosenquist CJ, Lindfors KK. Screening mammography in women aged 40-49 years: analysis of cost-effectiveness. Radiology. 1994;191:647-50. (PMID: 8184041)

144 Mandelblatt J, Saha S, Teutsch S, Hoerger T, Siu AL, Atkins D, et al. The cost-effectiveness of screening mammography beyond age 65 years: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2003;139:835-42. (PMID: 14623621) [Full Text]

145 Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective cost-effectiveness analysis of screening mammography. J Natl Cancer Inst. 2006;98:774-82. (PMID: 16757702)

146 Shen Y, Parmigiani G. A model-based comparison of breast cancer screening strategies: mammograms and clinical breast examinations. Cancer Epidemiol Biomarkers Prev. 2005;14:529-32. (PMID: 15734983)

147 Paskett ED, Tatum C, Rushing J, Michielutte R, Bell R, Foley KL, et al. Racial differences in knowledge, attitudes, and cancer screening practices among a triracial rural population. Cancer. 2004;101:2650-9. (PMID: 15505784)

148 Haakenson CP, Vickers KS, Cha SS, Vachon CM, Thielen JM, Kircher KJ, et al. Efficacy of a simple, low-cost educational intervention in improving knowledge about risks and benefits of screening mammography. Mayo Clin Proc. 2006;81:783-91. (PMID: 16770979)

149 Fox SA, Siu AL, Stein JA. The importance of physician communication on breast cancer screening of older women. Arch Intern Med. 1994;154:2058-68. (PMID: 8092910)

150 Fox SA, Stein JA. The effect of physician-patient communication on mammography utilization by different ethnic groups. Med Care. 1991;29:1065-82. (PMID: 1943268)

151 Legler J, Meissner HI, Coyne C, Breen N, Chollette V, Rimer BK. The effectiveness of interventions to promote mammography among women with historically lower rates of screening. Cancer Epidemiol Biomarkers Prev. 2002;11:59-71. (PMID: 11815402)

152 Costanza ME, Stoddard AM, Luckmann R, White MJ, Spitz Avrunin J, Clemow L. Promoting mammography: results of a randomized trial of telephone counseling and a medical practice intervention. Am J Prev Med. 2000;19:39-46. (PMID: 10865162)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD;
Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

153 Rakowski W, Ehrich B, Goldstein MG, Rimer BK, Pearlman DN, Clark MA, et al. Increasing mammography among women aged 40-74 by use of a stage-matched, tailored intervention. Prev Med. 1998;27:748-56. (PMID: 9808807)

154 Davis TC, Berkel HJ, Arnold CL, Nandy I, Jackson RH, Murphy PW. Intervention to increase mammography utilization in a public hospital. J Gen Intern Med. 1998;13:230-3. (PMID: 9565385)

155 Clover KA, Redman S, Forbes JF, Sanson-Fisher RW, Dickinson JA. Promotion of attendance for mammographic screening through general practice: a randomised trial of two strategies. Med J Aust. 1992;156:91-4. (PMID: 1736084)

156 Champion V, Maraj M, Hui S, Perkins AJ, Tierney W, Menon U, et al. Comparison of tailored interventions to increase mammography screening in nonadherent older women. Prev Med. 2003;36:150-8. (PMID: 12590989)

157 Edwards AG, Evans R, Dundon J, Haigh S, Hood K, Elwyn GJ. Personalised risk communication for informed decision making about taking screening tests. Cochrane Database Syst Rev. 2006:CD001865. (PMID: 17054144)

158 Mandelblatt J, Kanetsky PA. Effectiveness of interventions to enhance physician screening for breast cancer. J Fam Pract. 1995;40:162-71. (PMID: 7654272)

159 Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. BMJ. 1988;297:943-8. (PMID: 3142562)

160 Tabar L, Fagerberg G, Duffy SW, Day NE. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. J Epidemiol Community Health. 1989;43:107-14. (PMID: 2512366)

161 Tabar L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A, et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. Cancer. 1995;75:2507-17. (PMID: 7736395)

New References of Interest

162 Buist DS, Anderson ML, Reed SD, Aiello Bowles EJ, Fitzgibbons ED, Gandara JC, et al. Short-term hormone therapy suspension and mammography recall: a randomized trial. Ann Intern Med. 2009;150:752-65. (PMID: 19487710) [Full Text]

163 Moser K, Patnick J, Beral V. Inequalities in reported use of breast and cervical screening in Great Britain: analysis of cross sectional survey data. BMJ. 2009;338:b2025. (PMID: 19531549)

164 Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. BMJ. 2009;339:b2587. (PMID: 19589821)

165 Esserman L, Shieh Y, Thompson I. Rethinking screening for breast cancer and prostate cancer. JAMA. 2009;302:1685-92. (PMID: 19843904)

Screening for Breast Cancer

Author(s):
Teresa Cheng, MD; Karen M. Freund, MD, MPH; Sarah E. Lane, MPH; Pamela Ganschow, MD; Tracy A. Battaglia, MD, MPH

Status:
Editorial changes
2009-10-28

Glossary

BSE	breast self-examination
CBE	clinical breast examination
CI	confidence interval
DCIS	ductal carcinoma in situ
FDA	Food and Drug Administration
LCIS	lobular carcinoma in situ
MRI	magnetic resonance imaging
OR	odds ratio
RR	relative risk
SEER	Surveillance, Epidemiology, and End Results (program)

Operating Characteristics for Breast Cancer Screening Tests

Test	Gold Standard	Sensitivity (%)	Specificity (%)	Likelihood Ratio Positive	Likelihood Ratio Negative	Notes
CBE	Diagnosis of cancer	54	94	10.6	0.47	39
Mammography	Diagnosis of cancer	71-96	94-97			1; 5; 44
Computer-aided detection		84	87			62
MRI	Diagnosis of cancer	77 (70-84)	86.3 (80.9-91.7)			From studies of high-risk women aged 25 to 70 y (66; 67; 68; 69; 70)
Ultrasound	Diagnosis of cancer	27-50	92	12		From studies of high-risk women aged 25 to 70 y (67; 69; 70; 79)

CBE = clinical breast examination; MRI = magnetic resonance imaging.

Randomized, Controlled Trials of Screening Mammography

Trial	Year Began	Number of Participants	Age at Enrollment (y)	Number of Cancers	Estimated Sensitivity of All Rounds of Mammography	Sensitivity of First-round Screening at 1-year Intervals	Sensitivity of First-round Screening at 2-year Intervals	Reference(s)
Health Insurance Plan of Greater New York Randomized Controlled Trial	1963	30,239 cases 30,256 controls	40-64	173	39			40
Malmö Mammographic Screening Trial	1976-8	21,088 cases 21,195 controls	45-69 45-49 50-59 0-69 70-74	227	61	92 73 71 85 81		122; 159
Swedish Two-county Trial	1977	77,080 cases 55,985 controls	40-74 40-49 50-59 60-69 70-74	82 137 220 112		95 81 96 95 98	86	118; 119; 160; 161
Stockholm Mammographic Screening Trial	1981	40,318 cases 19,943 controls	40-64 40-49 50-59 60-64	45 95 48	64 9	86	68 53 75 69	121
Canadian National Breast Screening Study-1	1980	25,214 cases 25,216 controls	40-49	286	61	77	56	123; 125

Table Continued...

Randomized, Controlled Trials of Screening Mammography

Trial	Year Began	Number of Participants	Age at Enrollment (y)	Number of Cancers	Estimated Sensitivity of All Rounds of Mammography	Sensitivity of First-round Screening at 1-year Intervals	Sensitivity of First-round Screening at 2-year Intervals	Reference(s)
Canadian National Breast Screening Study-2	1980	19,711 cases 19,694 controls	50-59	347	66	88	56	43; 124

SEER Relative Survival Rates by Stage at Diagnosis for Breast Cancer

Survival Interval	Stage at Diagnosis (%)			
	Localized	Regional	Distant	Unstaged
Time zero	100.0	100.0	100.0	100.0
1-y	100.0	98.1	67.0	84.5
2-y	100.0	93.4	49.2	75.7
3-y	99.2	88.4	37.5	68.5
4-y	98.4	84.0	29.9	63.1
5-y	97.6	80.0	24.6	59.1
6-y	96.7	76.7	21.1	56.2
7-y	95.9	73.6	18.6	53.1
8-y	95.1	71.3	16.7	50.9
9-y	94.2	69.0	15.1	48.7
10-y	93.5	67.0	13.6	47.4

SEER = Surveillance, Epidemiology, and End Results (program).

Adapted from 104.

Breast Cancer Screening Guidelines by Screening Modality

Association/Society	BSE	CBE	Mammography	MRI
American Cancer Society				
Recommended?	Discuss (individual choice)	Yes	Yes	When lifetime risk >20%
Frequency	Monthly	a. Every 3 y b. Yearly	Yearly	Yearly
Age (y)	20+	a. 20-39 b. 40+	40+	20+
National Cancer Institute				
Recommended?	Not specific	Not specific	Yes	Not specific
Frequency			Every 1-2 y	
Age (y)			40+	
U.S. Preventive Services Task Force				
Recommended?	Not enough evidence	Not enough evidence	Yes	No, insufficient evidence to recommend in high-risk women
Frequency			1-2 y	
Age (y)			40+	
American College of Physicians				
Recommended?			a. Yes, and woman wants to partake in shared decision-making b. Yes, and woman does not want to partake in shared decision-making	Not specific
Frequency			a. Individualized via risk assessment and shared decision-making b. 1-2 y	
Age (y)			40-49	
American Medical Association				
Recommended?	Yes	Yes	Yes	No specific recommendations for high-risk women
Frequency		Yearly	Yearly	

Table Continued...

Breast Cancer Screening Guidelines by Screening Modality

Association/Society	BSE	CBE	Mammography	MRI
Age (y)		40+	40+	
American College of Obstetricians and Gynecologists				
Recommended?	Not specific	Yes	Yes	
Frequency		Yearly	a. Every 1-2 y b. Yearly	Not specific
Age (y)			a. 40-49 b. 50+	
Canadian Task Force on Preventive Health Care				
Recommended?	No	Yes	Yes	Not specific
Frequency		Every 1-2 y	Every 1-2 y	
Age (y)		50+	50+	

BSE = breast self-examination; CBE = clinical breast examination; MRI = magnetic resonance imaging.

MMWR
Morbidity and Mortality Weekly Report
www.cdc.gov/mmwr

Weekly April 3, 2009 / Vol. 58 / No. 12

Sociodemographic Differences in Binge Drinking Among Adults — 14 States, 2004

Binge drinking, defined in this study as consuming five or more alcoholic drinks on one occasion,* was responsible for 43,731 (54.9%) of the estimated 79,646 alcohol-attributable deaths each year in the United States during 2001–2005.† *Healthy People 2010* calls for reducing the prevalence of binge drinking among adults from the 16.6% baseline in 1998 to 6.0% (*1*). An overarching goal of *Healthy People* is to eliminate health disparities among different segments of the population.§ To assess binge drinking by sex, age group, race/ethnicity, education level, and income level, CDC analyzed data from an optional module of the 2004 Behavioral Risk Factor Surveillance System (BRFSS) survey, the most recent data available on binge drinking prevalence, frequency, and intensity (i.e., the number of drinks consumed per binge episode). This report summarizes the results of that analysis, which indicated that the prevalence of binge drinking was more common among men (24.3%), persons aged 18–24 years (27.4%) and 25–34 years (24.4%), whites (17.5%), and persons with household incomes ≥$50,000 (17.4%). However, after adjusting for sex and age, the highest average number of binge drinking episodes during the preceding 30 days was reported by binge drinkers whose household income was <$25,000. (4.9), and the highest average number of drinks per binge episode was reported by non-Hispanic blacks (8.4) and Hispanics (8.1). These findings underscore the need to implement effective population-based prevention strategies (e.g., increasing alcohol excise taxes) and develop effective interventions targeted at groups at higher risk.

BRFSS conducts annual state-based, random-digit–dialed telephone surveys of the noninstitutionalized U.S. civilian population aged ≥18 years, collecting data on health conditions and health risk behaviors, including binge drinking. In 2004, an optional survey module with additional questions on binge drinking was administered in 14 states.¶ Binge drinking was defined as having consumed five or more alcoholic drinks on one or more occasions during the preceding 30 days. For this report, responses to questions regarding the prevalence, frequency, and intensity of binge drinking were analyzed, beginning with the 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?" Those who acknowledged at least one occasion were then asked, "During the most recent occasion when you had five or more alcoholic beverages, about how many beers, including malt liquor, did you drink? ...about how many glasses of wine, including wine coolers, hard lemonade, or hard cider, did you drink? ...about how many drinks of liquor, including cocktails, did you have?" After excluding persons with missing or incomplete information, data from 62,684 respondents in the 14 states were used

* In 2006, the Behavioral Risk Factor Surveillance System definition of binge drinking for women changed from five alcoholic drinks to four drinks on one occasion.

† Estimated using the Alcohol-Related Disease Impact (ARDI) database. Available at http://apps.nccd.cdc.gov/ardi.

§ Including differences that occur by sex, race/ethnicity, education, income, disability, geographic location, or sexual orientation.

¶ California, Delaware, Idaho, Maine, Michigan, Minnesota, Montana, Nevada, New Hampshire, New Mexico, North Dakota, Virginia, Wisconsin, and Wyoming.

INSIDE

305 Tobacco Use Among Students Aged 13–15 Years — Baghdad, Iraq, 2008
308 Progress Toward Interruption of Wild Poliovirus Transmission — Worldwide, 2008
312 Use of Northern Hemisphere Influenza Vaccines by Travelers to the Southern Hemisphere
313 QuickStats

DEPARTMENT OF HEALTH AND HUMAN SERVICES
CENTERS FOR DISEASE CONTROL AND PREVENTION

The *MMWR* series of publications is published by the Coordinating Center for Health Information and Service, Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services, Atlanta, GA 30333.

Suggested Citation: Centers for Disease Control and Prevention. [Article title]. MMWR 2009;58:[inclusive page numbers].

for analysis. Response rates for each state were calculated using Council of American Survey and Research Organizations (CASRO) guidelines. Response rates ranged from 39.0% (California) to 63.2% (Minnesota) (median: 54.1%, and cooperation rates ranged from 59.9% (California) to 86.9% (Minnesota)(median: 74.9%).**

The prevalence of binge drinking was calculated by dividing the total number of respondents who reported at least one binge drinking episode during the preceding 30 days by the total number of BRFSS respondents in the 14 states. Analysis by state was not performed because of multiple subgroups with fewer than 50 respondents. The frequency of binge drinking was calculated by averaging the number of episodes reported by all binge drinkers during the preceding 30 days. The intensity of binge drinking (i.e., number of drinks per binge episode) was calculated by averaging the number of drinks consumed by binge drinkers during their most recent episode. All data were weighted to produce population-based estimates according to age-, race-, and sex-specific state population counts and to the respondent's probability of selection. Data were adjusted to the standard age and sex distribution of 2004 BRFSS respondents to provide estimates for race/ethnicity, education level, and annual household income level. Statistical significance was determined by pairwise linear contrasts of the estimates (*2*).

In 2004, the overall unadjusted prevalence of binge drinking among adults in the 14 states was 15.9% (Table 1). Binge drinking prevalence among men (24.3%) was three times higher than among women (7.9%). Men who reported binge drinking also reported a significantly higher average number of binge drinking episodes during the preceding 30 days (4.6) than women (2.9) and a significantly higher number of drinks consumed during their most recent binge episode (8.3 versus 6.9). Binge drinking prevalence decreased with advancing age, from 27.4% among respondents aged 18–24 years to 3.7% among respondents aged ≥65 years. In contrast, among binge drinkers, respondents aged ≥65 years reported the highest average number of binge drinking episodes during the preceding 30 days (6.8). The number of drinks consumed during the most recent binge decreased with advancing age, from 9.8 among adults aged 18–24 years to 6.4 among those aged ≥65 years.

The age- and sex-adjusted prevalence of binge drinking among non-Hispanic whites (17.5%) was significantly higher than the prevalence for Hispanics (14.4%) and non-Hispanic blacks (10.9%) (Table 2). Overall, among binge drinkers, the frequency of binge drinking episodes and the number of drinks

** The response rate is the percentage of persons who completed interviews among all eligible persons, including those who were not successfully contacted. The cooperation rate is the percentage of persons who completed interviews among all eligible persons who were contacted.

TABLE 1. Unadjusted percentage of persons reporting binge drinking, number of binge drinking episodes during the preceding 30 days, and average number of drinks consumed during the most recent binge drinking episode, by sex and age group — Behavioral Risk Factor Surveillance System, 14 states,* 2004

Characteristic	Prevalence		Average no. of binge drinking episodes,† preceding 30 days		Average no. of drinks consumed during most recent binge drinking episode†	
	%	(95% CI§)	No.	(95% CI)	No.	(95% CI)
Overall	**15.9**	**(15.2–16.6)**	**4.2**	**(3.9–4.4)**	**8.0**	**(7.7–8.2)**
Sex						
Men	24.3	(23.1–25.6)	4.6	(4.3–4.9)	8.3	(8.0–8.6)
Women	7.9	(7.3–8.5)	2.9	(2.7–3.1)	6.9	(6.6–7.3)
Age group (yrs)						
18–24	27.4	(24.6–30.4)	4.7	(4.0–5.3)	9.8	(9.1–10.4)
25–34	24.4	(22.5–26.4)	3.4	(3.1–3.8)	8.0	(7.6–8.4)
35–44	17.3	(15.9–18.8)	4.0	(3.5–4.4)	7.3	(7.0–7.6)
45–64	10.9	(10.1–11.9)	4.4	(3.9–4.9)	6.9	(6.6–7.1)
≥65	3.7	(3.0–4.6)	6.8	(4.6–9.1)	6.4	(5.4–7.3)

* California, Delaware, Idaho, Maine, Michigan, Minnesota, Montana, Nevada, New Hampshire, New Mexico, North Dakota, Virginia, Wisconsin, and Wyoming.
† Among the 8,381 respondents who reported binge drinking.
§ Confidence interval.

consumed during the most recent binge episode were similar among racial/ethnic populations; however, non-Hispanic blacks and Hispanics reported a higher intensity of binge drinking (8.4 and 8.1 drinks per binge episode, respectively) than whites (6.9).

College graduates had significantly lower age- and sex-adjusted prevalence of binge drinking (14.5%) than high school graduates or those with some college or technical school (both 17.1%) (Table 2). Respondents who did not graduate from high school reported the lowest binge drinking prevalence (14.2%) but, along with high school graduates, the highest frequency of binge drinking episodes (4.6) and the highest number of drinks consumed in the most recent episode (7.8). In contrast, binge drinking prevalence increased with income level and was highest among respondents with annual household incomes ≥$50,000 (17.4%) (Table 2). However, the number of drinks consumed per episode was significantly lower among respondents whose household income was ≥$35,000 compared with those whose household income was <$25,000.

Reported by: *JL Cremeens, PhD, D Nelson, MD, TS Naimi, MD, RD Brewer, MD, WS Pearson, PhD, Div of Adult and Community Health, National Center for Chronic Disease Prevention and Health Promotion; PR Chavez, PhD, EIS Officer, CDC.*

Editorial Note: Binge drinking is a risk factor for numerous adverse health and social outcomes, including alcohol poisoning, hypertension, acute myocardial infarction, sexually transmitted infections, unintended pregnancy, fetal alcohol syndrome, sudden infant death syndrome, suicide, interpersonal violence, and motor vehicle crashes (*3*). This report indicates that binge drinking is common among U.S. adults, especially among whites, males, persons aged 18–34 years, and those with household incomes ≥$50,000. These sociodemographic characteristics stand in contrast to characteristics for many other health risk factors (e.g., smoking and obesity), where prevalence tends to be higher among minorities and persons with lower education and income (*4*).

The findings in this report highlight the need for assessing the frequency and intensity of binge drinking among binge drinkers in addition to the prevalence of binge drinking in the general population. These additional measures are important because the risk for adverse outcomes (e.g., alcoholic liver disease or traffic fatalities) increases with the frequency of binge drinking and with the amount consumed per binge episode. Furthermore, reductions in the frequency and intensity of binge drinking generally might be expected to occur before reductions in the prevalence of binge drinking.

One plausible reason why binge drinking is more prevalent among whites and persons at higher income levels is that, unlike smoking, binge drinking has not been widely recognized as a health risk, subjected to intense prevention efforts, and socially stigmatized (*5*). The differences in binge drinking among population segments also likely reflects cultural factors and differences in state and local laws (*6*) that affect the price, availability, and marketing of alcoholic beverages. Finally, the increase in prevalence of binge drinking with increasing income levels likely reflects the fact that persons with higher household incomes have more disposable income available to spend on alcohol.

The findings in this report are subject to at least three limitations. First, the 14 states that administered the optional binge drinking module are not necessarily representative of all 50 states; therefore, the results cannot be generalized to the entire U.S. population. Second, BRFSS data are self-reported; alcohol consumption generally, and excessive drinking in particular, are underreported in surveys because of recall

TABLE 2. Age- and sex-adjusted* percentage of adults reporting binge drinking, number of binge drinking episodes during the preceding 30 days, and average number of drinks consumed during the most recent binge drinking episode, by race/ethnicity, education level, and income level — Behavioral Risk Factor Surveillance System (BRFSS), 14 states,† 2004

Characteristic	Prevalence		Average no. of binge drinking episodes,† preceding 30 days		Average number of drinks consumed during most recent binge drinking episode†	
	%	(95% CI§)	No.	(95% CI)	No.	(95% CI)
Race/Ethnicity						
White, non-Hispanic	17.5	(16.8–18.2)	3.9	(3.7–4.2)	6.9	(6.7–7.0)
Black, non-Hispanic	10.9	(8.7–13.6)	4.5	(3.3–5.7)	8.4	(7.0–9.8)
Hispanic	14.4	(12.6–16.4)	3.6	(2.4–4.8)	8.1	(7.3–8.9)
American Indian/Alaska Native	13.4	(10.1–17.5)	4.5	(3.3–5.6)	7.7	(7.1–8.3)
Other**	8.8	(6.9–11.3)	4.0	(3.1–4.9)	7.5	(6.8–8.2)
Education level						
Less than high school diploma	14.2	(12.2–16.5)	4.6	(3.7–5.5)	7.8	(7.2–8.5)
High school diploma	17.1	(15.9–18.3)	4.6	(4.0–5.2)	7.6	(7.3–7.9)
Some college or technical school	17.1	(15.8–18.4)	3.6	(3.3–3.9)	7.0	(6.8–7.3)
College graduate	14.5	(13.3–15.8)	3.3	(2.9–3.7)	6.5	(6.2–6.9)
Annual household income level						
<$15,000	13.7	(11.3–16.4)	4.9	(4.0–5.7)	7.7	(7.2–8.3)
$15,000 to <$25,000	14.3	(12.6–16.1)	4.9	(3.9–5.9)	8.0	(7.4–8.6)
$25,000 to <$35,000	16.5	(14.7–18.4)	4.3	(3.7–4.9)	7.2	(6.9–7.4)
$35,000 to <$50,000	16.7	(15.3–18.3)	4.0	(3.4–4.6)	6.8	(6.6–7.1)
≥$50,000	17.4	(16.2–18.7)	3.5	(3.1–4.0)	6.9	(6.6–7.1)

* Age and sex adjusted to the standard distribution of all 2004 BRFSS respondents.
† California, Delaware, Idaho, Maine, Michigan, Minnesota, Montana, Nevada, New Hampshire, New Mexico, North Dakota, Virginia, Wisconsin, Wyoming.
§ Among the 8,381 respondents who reported binge drinking.
¶ Confidence interval.
** Asians/Pacific Islanders and persons with mixed or unreported race/ethnicity.

bias, social desirability response bias, and nonresponse bias (*7*). Finally, in 2005, BRFSS changed the definition of binge drinking for women from five or more drinks per occasion to four or more drinks per occasion; the prevalence of binge drinking among women would have been higher using the new definition (*8*).

These findings support the need to implement effective population-based strategies (e.g., increasing alcohol excise taxes, limiting the number of retail outlets where alcohol is sold in a particular geographic area, and maintaining and enforcing age 21 years as the minimum age for legal drinking) (*9,10*) to prevent binge drinking. In addition, the frequency and intensity of binge drinking should be routinely monitored to guide the development and evaluation of culturally appropriate binge drinking prevention and intervention strategies for groups at greater risk.

References

1. US Department of Health and Human Services. Reduce the proportion of persons engaging in binge drinking of alcoholic beverages. Objective 26-11c. Healthy people 2010 (conference ed, in 2 vols). Washington, DC: US Department of Health and Human Services; 2000. Available at http://www.healthypeople.gov/document/pdf/volume2/26substance.pdf.
2. Ott L. An introduction to statistical methods and data analysis. 4th ed. Belmont, CA: Duxbury Press; 1988:888.
3. Naimi TS, Brewer RD, Molded A, Denny C, Ferula MK, Marks JS. Binge drinking among US adults. JAMA 2003;289:70–5.
4. Kantilla S, Gregg WE, Cheng I, et al. Socioeconomic status and trends in disparities in 4 major risk factors for cardiovascular disease among US adults, 1971–2002. Arch Intern Med 2006;166:2348–55.
5. Brewer RD, Swan MH. Binge drinking and violence. JAMA 2005; 294:616–8.
6. National Institute on Alcohol Abuse and Alcoholism. Alcohol Policy Information System. Available at http://www.alcoholpolicy.niaaa.nih.gov.
7. Stockwell T, Donath S, Cooper-Stanbury M, Chikritzhs T, Catalano P, Mateo C. Under-reporting of alcohol consumption in household surveys: a comparison of quantity-frequency, graduated-frequency and recent recall. Addiction 2004;99:1024–33.
8. National Institute on Alcohol Abuse and Alcoholism. NIAAA Council approves definition of binge drinking. NIAAA Newsletter; winter 2004:3. Available at http://pubs.niaaa.nih.gov/publications/newsletter/winter2004/newsletter_number3.pdf.
9. Babor T, Caetano R, Casswell S, et al. Alcohol: no ordinary commodity—research and public policy. Oxford and London, England: Oxford University Press; 2003.
10. CDC. Guide to community preventive services. Preventing excessive alcohol use. Available at http://www.thecommunityguide.org/alcohol/index.html.

Saved by the Nose: Bystander-Administered Intranasal Naloxone Hydrochloride for Opioid Overdose

Maya Doe-Simkins, MPH, Alexander Y. Walley, MD, MSc, Andy Epstein, RN, MPH, and Peter Moyer, MD, MPH

Administering naloxone hydrochloride (naloxone) during an opioid overdose reverses the overdose and can prevent death. Although typically delivered via intramuscular or intravenous injection, naloxone may be delivered via intranasal spray device. In August 2006, the Boston Public Health Commission passed a public health regulation that authorized an opioid overdose prevention program that included intranasal naloxone education and distribution of the spray to potential bystanders. Participants were taught by trained nonmedical needle exchange staff. After 15 months, the program provided training and intranasal naloxone to 385 participants who reported 74 successful overdose reversals. Problems with intranasal naloxone were uncommon. Overdose prevention education with distribution of intranasal naloxone is a feasible public health intervention to address opioid overdose. *Am J Public Health.* 2009;99:788–791. doi:10.2105/AJPH.2008.146647.

KEY FINDINGS

- Needle-exchange participants have experienced and witnessed high rates of overdoses.
- Needle-exchange participants can successfully recognize an overdose and use intranasal naloxone to reverse potentially fatal opioid overdoses.
- With the support and regulation of the local public health authority, overdose prevention programs can provide training and distribute intranasal naloxone without a direct clinical health care provider–patient encounter.
- Overdose prevention programs that include the distribution of intranasal naloxone by nonmedical personnel are feasible for city public health departments.

RATES OF OPIOID OVERDOSE

have increased since the early 1990s because of lower-cost, higher-purity heroin and prescription opioid abuse.[1–5] In Massachusetts, from 1990 to 2006, annual opioid overdose–related fatalities increased over 6-fold, from 94 to 637.[6,7] In response, the Boston Public Health Commission (BPHC) passed a regulation that authorized the development of an overdose prevention program with naloxone distribution through its mobile needle-exchange program. This program is innovative, because it includes the distribution of intranasal naloxone by trained, nonmedical public health workers to potential overdose bystanders for administration to overdose victims. Legal and regulatory barriers to implementation are detailed in the box on page 791.

Naloxone, an opioid antagonist, reverses opioid overdose by displacing opioid agonists, such as heroin or oxycodone, from opioid receptors. It is the standard treatment used by medical personnel. It has no abuse potential, and its only contraindication is a prior allergic reaction, which is rare.[8] Although typically administered intravenously or intramuscularly, it can be administered intranasally.[9–13] Strong interest in overdose prevention training and access to naloxone exists among potential overdose bystanders, including family members[14] and drug-using partners.[15] Overdose prevention programs with naloxone distribution that train and distribute naloxone to people who are likely to witness an overdose have been successfully implemented in several communities, including Chicago,[16,17] New York,[18,19] San Francisco,[20] Baltimore,[15,21] and New Mexico.[8] A 6-program study demonstrated that trained bystanders were similarly skilled as medical experts in recognizing opioid overdose situations, and when naloxone was indicated.[22]

The BPHC started an overdose prevention program with intranasal naloxone distribution as a result of the successful experience of the city's emergency medical services use of the nasal spray as a prehospital treatment for opioid overdose; the concept was also seen as an attractive option because intranasal delivery of the drug eliminates the risks of needlestick injuries and needle disposal. BPHC implemented the program through the needle-exchange program because program participants were considered particularly likely to witness overdoses.

PROGRAM CURRICULUM

All participating needle-exchange program staff—2 nurses and 4 nonmedical public health workers—completed 8 hours of didactic training, a knowledge test, and at least 4 supervised bystander-training sessions. Both the staff training and bystander training were adapted from existing program curricula from other cities that primarily used needle-based naloxone.[8,14,17–21]

The 15-minute bystander training included techniques in overdose prevention. Staff completed a checklist (available as a supplement to the online article at http://www.ajph.org) to ensure participant comprehension. Overdose prevention kits included instructions; 2 luer-lock, prefilled

syringes with 2 mg/2 mL naloxone hydrochloride; and the mucosal atomization device. Participants were instructed to deliver 1 mL (1 mg) to each nostril of the overdose victim. Because most opioid agonists have a longer half-life than naloxone, if overdose symptoms returned, victims could be treated with the second dose.

DATA COLLECTION AND ANALYSIS

From September 2006 to December 2007, during each bystander training, staff completed an enrollment form, recording respondents' demographics and overdose risk factors. When participants returned to the needle-exchange program, staff completed a form detailing overdoses witnessed, use of naloxone, and whether additional doses were needed. Data were maintained in a Microsoft Access 2003 database (Microsoft Corp, Redmond, WA). We compared enrollment data from participants who reported overdose reversals with those who did not with the *t* test of means and the χ^2 or Fisher exact test. We used SAS version 9.1 (SAS Institute, Cary, NC) for all tests of comparison.

DISCUSSION AND EVALUATION

Over 15 months, the program provided education and intranasal naloxone to 385 potential bystanders. At enrollment (Table 1), heroin was the most frequently used drug, followed by cocaine, methadone, benzodiazepines, and alcohol. Opioids were used on a mean of 24.1 of the last 30 days. Cocaine was the drug used most commonly in combination with heroin. Among 224 (64%) who reported a previous overdose, the median number of lifetime overdoses was 2, and among the 303 (92%) who had witnessed an overdose, the median number of lifetime witnessed overdoses was 5.

Follow-up contact was made at least once with 278 (72%) participants, 222 of whom reported no overdoses witnessed and no need for additional doses of naloxone. Among the 57 participants who requested additional doses, 7 had the naloxone lost, stolen, or confiscated, and 50 administered naloxone while observing an overdose (Figure 1). Among the 50 participants (13%) who reported reversing an overdose, 74 successful reversals were reported. Except for mean age (43 vs 39 years; $P < .05$), there were no significant differences between those participants who reported reversing an overdose and those who did not (data not shown). Emergency medical personnel were involved in 21 of the 74 (28%) reported overdoses and were not involved in 39 (53%) reported overdoses. Involvement by emergency medical personnel was not reported in the remainder (data available as a supplement to the online article at http://www.ajph.org). Two previous studies of naloxone distribution programs have reported similar rates of emergency medical personnel involvement (10% to 31%).[20,23]

Among follow-up contacts, problems were uncommon. During 4 overdoses, bystanders could not connect the mucosal atomization device to the syringe, although each resulted in successful reversal. Two administered naloxone nasally directly from the syringe, 1 injected the naloxone intramuscularly, and 1 did not administer naloxone, but delivered rescue breathing and physical stimulation until Boston Emergency Medical Services arrived. Two bystanders reported that naloxone induced withdrawal symptoms, but, in both cases, the victim did not use additional opioids to alleviate symptoms. Two bystanders observed the naloxone wearing off: 1 readministered it after 90 minutes, and 1 reported that the victim became resedated after 20 minutes, when Boston Emergency Medical Services assumed care. Two people had naloxone confiscated at a homeless shelter, 1 reported being expelled from a residential drug treatment program for having the substance, and 3 reported negative interactions with emergency medical personnel,

TABLE 1—Selected Characteristics of Participants (N = 385) in an Overdose Prevention Program With Intranasal Naloxone Distribution: Boston, MA, September 2006–December 2007

	Sample Total	No. (%) or Mean ±SD
Age, y	377	39.6 ± 11
Women	381	129 (34)
Race/Ethnicity	374	
White		245 (66)
Hispanic		81 (22)
Black		45 (12)
Other		3 (1)
HIV status	219	
Positive		26 (12)
Negative		193 (88)
HCV status	246	
Positive		159 (65)
Negative		87 (35)
Days opioids used	351	24.1 ± 10.7
Substance used in the last 30 d	385	
Heroin		273 (71)
Methadone		149 (39)
Buprenorphine		11 (3)
Other opioids		60 (16)
Cocaine		155 (40)
Benzodiazepines		118 (31)
Alcohol		88 (23)
Heroin and cocaine		125 (33)
Heroin and benzos		98 (26)
Heroin and alcohol		69 (18)
Heroin, benzos, alcohol		35 (9)
Clonodine		26 (7)
History of nonfatal overdose	349	
Had a nonfatal overdose		225 (65)
Nonfatal overdoses experienced, median (interquartile range)		2 (1–5)
Nonfatal overdose treated with naloxone		146 (69)[a]
Lifetime witnessed overdose	329	
Had witnessed an overdose		303 (92)
Overdoses witnessed, median (interquartile range)		5 (3–15)

[a]The percentage represents the percentage of respondents who had a nonfatal overdose and answered the question about whether naloxone had been used (n = 212).

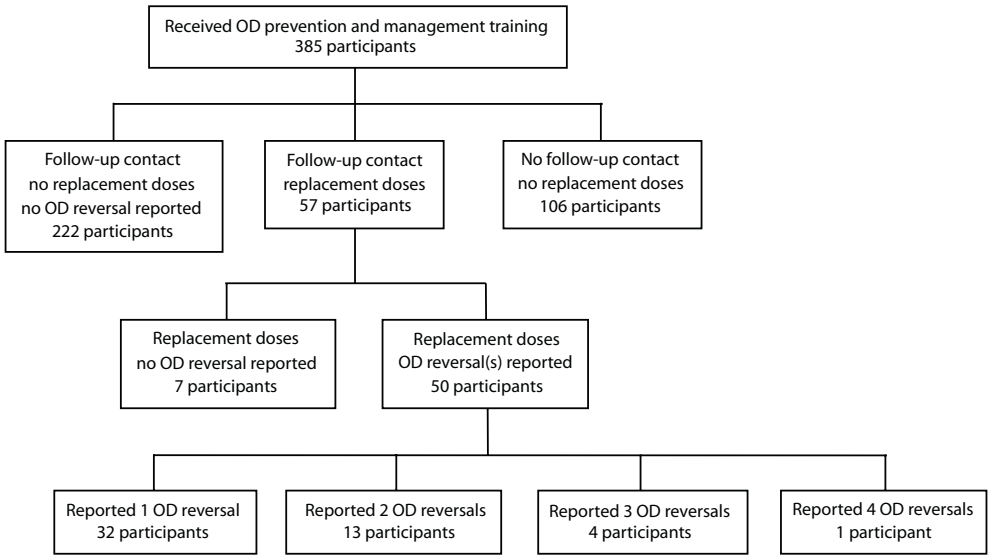


FIGURE 1—Flow chart of follow-up of 385 potential bystanders who received overdose (OD) prevention and management training: Boston, MA, September 2006–December 2007.

none of which resulted in arrest (8 reported positive interactions).

Of the 74 reported reversals, 4 reports were of bystanders not initially enrolled in the program who used intranasal naloxone obtained from peers who were enrolled. Thus, there was some peer-to-peer overdose knowledge and skill transfer beyond the program.

The BPHC overdose-prevention naloxone distribution program was implemented without substantial additional funding. Space, printing costs, and staff time were provided by the existing needle-exchange program. Naloxone kits cost approximately $25.

NEXT STEPS

The BPHC naloxone distribution program is a feasible, successful program that includes distribution of intranasal naloxone by nonmedical staff. The Massachusetts Department of Public Health has identified overdose prevention as a major focus area for new public health initiatives and has expanded the program to 5 additional sites that target needle-exchange participants, staff at substance abuse treatment programs, homeless shelters, and families and friends of opioid users. ■

About the Authors

At the time of the study, Maya Doe-Simkins was with the Boston Public Health Commission AHOPE Needle Exchange program, Boston, MA. Alexander Y. Walley is with the Boston University School of Medicine, Boston, and the Massachusetts Department of Public Health's overdose-prevention pilot program, Boston. Andy Epstein is with the Massachusetts Department of Public Health, Boston. Peter Moyer is with the Department of Emergency Medicine, Boston University School of Medicine, Boston, and Boston Emergency Medical Services, Boston Police and Fire Departments, Boston.

Requests for reprints should be sent to Alexander Y. Walley, MD, Section of General Internal Medicine, Boston Medical Center, 801 Massachusetts Ave, 2nd Floor, Boston, MA 02118 (e-mail: awalley@bu.edu).

This article was accepted September 2, 2008.

Contributors

M. Doe-Simkins and A. Y. Walley, as co-first authors, jointly wrote the first draft and led subsequent revisions of the article. M. Doe-Simkins managed the data collection and assembly of the dataset and was a lead trainer of participants and other staff. A. Y. Walley performed the data analysis and led the institutional review board application. A. Epstein led the development and implementation of the project. P. Moyer provided medical supervision and direction. All authors helped to conceptualize ideas, develop the project, interpret findings, and review drafts of the article.

Acknowledgments

This project would not have been possible without the AHOPE needle-exchange staff and the administration of the Boston Public Health Commission (BPHC). The program benefited from ongoing input and feedback from the participants. We would like to specifically thank John Auerbach, former director of BPHC, and current Commissioner of Health, Adam Butler, a lead trainer and manager of the AHOPE needle-exchange program, and John Townsend, General Counsel of the BPHC.

Human Participant Protection

This study was approved as an exempt study by the Boston University Medical Center institutional review board.

References

1. Paulozzi LJ, Budnitz DS, Xi Y. Increasing deaths from opioid analgesics in the United States. *Pharmacoepidemiol Drug Saf.* 2006;15:618–627.

2. Centers for Disease Control and Prevention. Unintentional opiate overdose deaths—King County, Washington, 1990–1999. *MMWR Morb Mortal Wkly Rep.* 2000;49:636–640.

3. Centers for Disease Control and Prevention. Heroin overdose deaths—Multnomah County, Oregon, 1993-1999. *MMWR Morb Mortal Wkly Rep.* 2000;49:633–636.

4. Shah NG, Lathrop SL, Reichard RR, Landen MG. Unintentional drug overdose death trends in New Mexico, USA, 1990–2005: combinations of heroin, cocaine, prescription opioids and alcohol. *Addiction.* 2008;103:126–136.

5. Paulozzi LJ. Opioid analgesic involvement in drug abuse deaths in American metropolitan areas. *Am J Public Health.* 2006;96:1755–1757.

6. Fernandez W, Hackman H, McKeown L, Anderson T, Hume B. Trends in opioid-related fatal overdoses in Massachusetts, 1990–2003. *J Subst Abuse Treat.* 2006;31:151–156.

7. Massachusetts Department of Public Health Bureau of Substance Abuse Services. Fatal opioid overdose trends continue. *Prevention News* [serial online]. 2008;2. Available at: http://170.63.97.68/Eeohhs2/docs/dph/substance_abuse/prevention_newsletter_2008spring.rtf. Accessed September 21, 2008.

8. Baca CT, Grant KJ. Take-home naloxone to reduce heroin death. *Addiction.* 2005;100:1823–1831.

9. Ashton H, Hassan Z. Intranasal naloxone in suspected opioid overdose. *Emerg Med J.* 2006;23:221–223.

10. Barton ED, Colwell CB, Wolfe T, et al. Efficacy of intranasal naloxone as a needleless alternative for treatment of opioid overdose in the prehospital setting. *J Emerg Med.* 2005;29:265–271.

11. Kelly AM, Kerr D, Dietze P, Patrick I, Walker T, Koutso giannis Z. Randomised trial of intranasal versus intramuscular naloxone in prehospital treatment for suspected opioid overdose. *Med J Aust.* 2005;182:24–27.

12. Loimer N, Hofmann P, Chaudhry HR. Nasal administration of naloxone is as effective as the intravenous route in opiate addicts. *Int J Addict.* 1994;29:819–827.

13. Barton ED, Ramos J, Colwell C, Benson J, Baily J, Dunn W. Intranasal administration of naloxone by paramedics. *Prehosp Emerg Care.* 2002;6:54–58.

14. Strang J, Manning V, Mayet S, et al. Family carers and the prevention of heroin overdose deaths: Unmet training need and overlooked intervention opportunity of resuscitation training and supply of naloxone. *Drugs Educ Prev Policy*; 2008;15:211–218.

15. Sherman SG, Gann DS, Tobin KE, Latkin CA, Welsh C, Bielenson P. "The life they save may be mine": diffusion of overdose prevention information from a city sponsored programme [published online ahead of print May 23, 3008]. *Int J Drug Policy.* PMID: 18502635.

16. Bigg D. Data on take home naloxone are unclear but not condemnatory. *BMJ.* 2002;324:678.

17. Maxwell S, Bigg D, Stanczykiewicz K, Carlberg-Racich S. Prescribing naloxone to actively injecting heroin users: a program to reduce heroin overdose deaths. *J Addict Dis.* 2006;25:89–96.

18. Galea S, Worthington N, Piper TM, Nandi VV, Curtis M, Rosenthal DM. Provision of naloxone to injection drug users as an overdose prevention strategy: early evidence from a pilot study in New York City. *Addict Behav.* 2006;31:907–912.

19. Piper TM, Rudenstine S, Stancliff S, et al. Overdose prevention for injection drug users: lessons learned from naloxone training and distribution programs in New York City. *Harm Reduct J.* 2007;4:3.

20. Seal KH, Thawley R, Gee L, et al. Naloxone distribution and cardiopulmonary resuscitation training for injection drug users to prevent heroin overdose death: a pilot intervention study. *J Urban Health.* 2005;82:303–311.

21. Tobin KE, Sherman SG, Beilenson P, Welsh C, Latkin CA. Evaluation of the Staying Alive programme: training injection drug users to properly administer naloxone and save lives [published online ahead of print April 21, 2008]. *Int J Drug Policy.* PMID: 18434126.

22. Green TC, Heimer R, Grau LE. Distinguishing signs of opioid overdose and indication for naloxone: an evaluation of six overdose training and naloxone distribution programs in the United States. *Addiction.* 2008;103:979–989.

23. Dettmer K, Saunders B, Strang J. Take home naloxone and the prevention of deaths from opiate overdose: two pilot schemes. *BMJ.* 2001;322(7291): 895–896.

24. Davis JE. Self-injectable epinephrine for allergic emergencies. *J Emerg Med.* 2008. Epub ahead of print January 31, 2008. Available at: http://www.sciencedirect.com. Accessed September 21, 2008.

25. Sporer KA, Kral AH. Prescription naloxone: a novel approach to heroin overdose prevention. *Ann Emerg Med.* 2007;49(2):172–177.

26. Burris S, Norland J, Edlin BR. Legal aspects of providing naloxone to heroin users in the United States. *Int J Drug Policy.* 2001;12:237–248.

Legal and Regulatory Barriers to Implementing an Overdose Prevention Program With Intranasal Naloxone Distribution by Nonmedical Personnel

Barrier	Response
Nonmedical personnel are not authorized to distribute prescription medication and are not authorized to administer a prescription medication to a person who has not been prescribed the medication.	▪ The standard of care for the use of naloxone has for decades included use by prehospital personnel in nonclinical settings operating under standing orders from physicians who are neither on-site nor directly supervising. ▪ Other life saving prescription medications, such as epinephrine injectors for anaphylactic shock,[24] and other devices, such as automated external defibrillators, are used by bystanders and nonmedical personnel. ▪ Other states, such as New Mexico, New York and Connecticut, have addressed this by passing laws that limit the liability of medical and nonmedical personnel who administer and distribute potentially lifesaving medication.[25] ▪ A study of 6 programs that train bystanders to recognize and respond to opioid overdose by using naloxone has demonstrated that trained potential bystanders are similarly skilled as medical experts in recognizing opioid overdose situations and when naloxone is indicated.[22] ▪ A local public health regulation was passed by BPHC, the City of Boston's board of health, identifying the overdose-prevention naloxone distribution program as an official public health program and assuming liability for the work of medical and non-medical personnel involved in the program. ▪ Under the medical license of the Medical Director of Boston Emergency Medical Services, potential bystanders received a standard curriculum about overdose prevention with instructions and demonstration of how to properly use the medication. Receipt of this curriculum was documented by BPHC staff.
Intranasal delivery of naloxone is an off-label method.	▪ Prescriptions drugs may be and are routinely given for any indication not explicitly prohibited by law.[25,26] ▪ While no large scale randomized clinical trials have been conducted, intranasal naloxone has been evaluated in several research studies, with little evidence of adverse events.[9–13] A small randomized trial comparing intranasal with intramuscular delivery of naloxone used by emergency personnel demonstrated that intranasal delivery had a longer time to clinical response (8 minutes vs 6 minutes), but less agitation or irritation (2% vs 13%).[11] ▪ Intranasal naloxone is a first-line treatment for opioid overdose among emergency medical personnel in the local Boston community.

Journal of Substance Abuse Treatment 37 (2009) 286–291

Journal of Substance Abuse Treatment

Regular article

HIV screening among substance-abusing veterans in care

Nameeta M. Dookeran, (M.D., M.Sc.)[a,b,*], James F. Burgess, Jr., (Ph.D.)[c,d], Candice C. Bowman, (Ph.D., R.N.)[e], Matthew Bidwell Goetz, (M.D.)[f,g], Steven M. Asch, (M.D., M.P.H.)[h,i], Allen L. Gifford, (M.D.)[a,d,j]

[a]*VA HSR&D Center for Health Quality, Outcomes and Economic Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA 01730, USA*
[b]*Brigham and Women's Hospitalist Service, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA*
[c]*Center for Organization, Leadership and Management Research, VA Boston Healthcare System, Boston, MA 02130, USA*
[d]*Department of Health Policy and Management, Boston University School of Public Health, Boston, MA 02118, USA*
[e]*Center for the Study of Healthcare Provider Behavior, VA Greater Los Angeles Healthcare System, Sepulveda, CA 91343, USA*
[f]*Infectious Diseases Section and Department of Medicine, VA Greater Los Angeles Healthcare System, Los Angeles, CA 90073, USA*
[g]*The David Geffen School of Medicine at UCLA, Los Angeles, CA 90024, USA*
[h]*HIV Hepatitis Quality Enhancement Research Initiative, VA Greater Los Angeles Healthcare System, Los Angeles, CA 90073, USA*
[i]*Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA*
[j]*Section of General Internal Medicine, Boston University School of Medicine, Boston, MA 02118, USA*

Received 23 September 2008; received in revised form 3 February 2009; accepted 2 March 2009

Abstract

Calls for screening for HIV infection among individuals with substance use disorders, including alcohol use, are increasing. We investigated HIV screening and its predictors in the Veterans Health Administration (VA) system among such individuals in care. Our primary outcome was retrospective evidence of screening for HIV infection, adjusting for patient demographics and important comorbid disease. Of the 371,749 sample patients with histories of substance use disorders using VA services, 20% had evidence of HIV screening. Screening was lowest among those with alcohol use disorders alone (11%) and highest among those treated in substance use programs (28%) or receiving inpatient care (28%). The findings suggest a low recognition of substance use disorders (especially alcohol use) as risk factors for HIV. Quality improvement initiatives to increase risk factor recognition and screening among patients with substance use disorders will yield benefits in the fight against HIV.

Keywords: HIV screening; HIV testing; HIV risk; Alcohol use disorders; Substance use disorders

1. Introduction

Screening for HIV in individuals with drug and alcohol use disorders is a longstanding and broadly endorsed quality of care recommendation (Branson et al., 2006; Centers for Disease Control and Prevention [CDC], 2001; National Institute of Alcohol Abuse and Alcoholism [NIAAA], 2002). Too often, however, HIV screening of substance users is not done or is done too late (Branson et al., 2006; Owens et al., 2007; Samet et al., 1999, 2001). In data from a large managed care organization, more than half of HIV-infected individuals already had advanced disease when diagnosed (suggested by CD4 cell counts <350), although 26% had HIV risk factors documented more than 1 year before diagnosis (Klein et al., 2003).

Although risky injection drug use practices directly increase the risk of HIV transmission, noninjection drug use also indirectly increases risk substantially by increasing disinhibition that leads to other directly risky behaviors (Neaigus et al., 2001; Strathdee & Sherman 2003). Among heroin and/or cocaine users in a state cross-sectional survey, the HIV seroprevalence was found to be similar in injection and noninjection drug users (in the 12%–17% range; Des

* Corresponding author. Brigham and Women's Hospitalist Service, 75 Francis St, PBB-B-428, Boston, MA 02115, USA. Tel.: +1 617 732 6649; fax: +1 617 278 6906.
E-mail address: ndookeran@partners.org (N.M. Dookeran).

0740-5472/09/$ – see front matter
doi:10.1016/j.jsat.2009.03.003

Jarlais et al., 2007). Alcohol abuse has also been associated with HIV risky behaviors and sexually transmitted diseases (Cook & Clark, 2005; Ehrenstein et al., 2004; Rees et al., 2001; Stein et al., 2000). Screening of high-risk groups is essential to improving access to effective treatment and reducing HIV transmission (DiClemente et al., 2002; Marks et al., 2005; Palella et al., 2003; Paltiel et al., 2005; Yeni et al., 2004). However, most studies examining HIV screening rates among substance users have been in relatively small or geographically limited samples, and few have focused on screening among patients with alcohol use disorders (Liddicoat et al., 2004; Owens et al., 2007; Samet et al., 1999).

The U.S. Veterans Health Administration (VA) health care system provides care yearly to approximately 5 million military veterans, including many with substance use disorders who are the focus of this study (Justice et al., 2006). Our primary objective was to determine the rate of HIV screening (i.e., HIV status assessment, as defined by having been tested for or diagnosed with HIV during a defined period) retrospectively among veterans with substance use disorders seen for VA care nationally. We also sought to determine whether HIV screening rates differed either by type of substance use disorder (alcohol as opposed to illicit drugs, i.e., cocaine, opiate and/or amphetamine) or by use of different medical care treatment venues (primary care [PC], emergency department [ED], outpatient substance use clinic/program, and inpatient care).

2. Methods

2.1. Study population

Patients were selected from the 4,974,979 in the VA Medical Inpatient and Outpatient Data Sets of the National Patient Care Database who used specified VA health care services within a 12-month identification period from June 1, 2004, to May 31, 2005. Health care services use was defined as at least (a) one outpatient care visit (defined by the presence of at least 1 of 75 clinic codes for visits to PC and medicine subspecialty clinics, general mental health and substance use clinics/programs, and EDs) or (b) one inpatient care stay (defined by presence of any data in the VA Medical Inpatient Data Set, which included data on residential substance use programs and domiciliary stays).

Patients seen during the 12-month identification period were included in the analysis if they had a diagnosis of one or more substance use disorders in the past and if they used VA health care services in the identification period at least once after diagnosis to ensure there was clinical opportunity to assess for HIV. Substance use disorder histories were identified retrospectively based on presence of *International Classification of Diseases, Ninth Revision* (*ICD-9*) diagnostic codes within a 68-month look-back period (October 1, 1999, through May 31, 2005). Patients were included if they had at least 2 documented inpatient/outpatient *ICD-9* diagnostic codes for one or more of the following: (a) alcohol (8 codes) or illicit drug abuse/dependence (31 codes) or (b) a medical condition indicating substance abuse/dependence, for example, "Alcohol-related Liver Disease" (20 codes). These criteria led to a final national sample of 371,749 patients (7% of all VA patients seen in the identification period).

2.2. Main independent variables and covariates

The main study independent variable identified the category of substance use (alcohol use, illicit drug use, or both) using *ICD-9* codes within the 68-month look-back period. Other independent categorical variables focused on use of different health care services in the 12-month identification period, namely, PC/ED use, use of outpatient substance use clinics/programs, and receipt of inpatient care. Covariates included categorical variables for patient demographics and type of health insurance (if any, e.g., Medicaid, Medicare, or private) within the 12-month identification period, and for presence of hepatitis B, hepatitis C, and sexually transmitted disease diagnoses by *ICD-9* codes within the 68-month look-back period.

2.3. Primary outcome measure

HIV screening was defined as any retrospective evidence of clinical HIV status assessment. As such, using the VA Decision Support System laboratory file for HIV tests, we defined patients as having been screened for HIV if they had any evidence in the VA electronic medical record of ever having been tested for or diagnosed with HIV within the 68-month look-back period. The definition was intentionally as inclusive as possible, given data availability. Patients were classified as tested if they had at least one HIV laboratory test of any kind (HIV antibody or plasma HIV RNA concentration). They were classified as diagnosed with HIV if they had any of four possible *ICD-9* codes for HIV/AIDS.

2.4. Statistical analyses

We performed bivariate analyses to explore the association between our primary outcome measure, HIV screening, and each of the key independent variables and covariates. To determine the most significant predictors of HIV screening, we estimated a generalized linear mixed model accounting for main VA health care facility used as a random effect and other potential confounders as fixed effects.

Given the generally consistent trends in HIV screening rates on comparison of bivariate and multivariate model analyses, in our results below, we chose to highlight the absolute HIV screening rates found in bivariate analyses. Details on the generalized linear mixed model analysis can be found on the VA Center for Health Quality Outcomes & Economic Research (CHQOER) Web site (Dookeran et al., 2009).

This research was approved by institutional review boards at Boston University and the Edith Nourse Rogers Memorial VA Hospital.

3. Results

3.1. Patient characteristics

Among the 371,749 veterans with substance use disorders in this study, the two most common disorders were those of alcohol use (96%) and cocaine use (36%) (Table 1). More than half of the study population (57%) had histories of alcohol use disorders alone; 39% had histories of both illicit drug and alcohol use disorders (Table 2). Most patients (69%) had been seen in PC at least twice in the 12-month identification period. About one quarter (26%) had been in an outpatient substance use clinic or program, and 29% had received inpatient care.

Veterans with substance abuse in our study were mostly male (97%), older (54% more than 54 years old), White (51%), and not currently married (72%). Most had housing (69%), lived within 30 miles of their primary VA care facility (59%), and reported incomes lower than twice the poverty line (55%). Most (71%) had no health insurance other than VA membership and almost a quarter (22%) had hepatitis C infection.

3.2. Overall HIV screening rates among patients with substance use disorders

Only 20% of all veterans in care with a substance use disorder history had any retrospective evidence of HIV screening in VA records. Patients with both illicit drug and alcohol use disorders were screened the most (32%), followed by those with illicit drug use disorders alone (25%), and those with alcohol use disorders alone (11%) (Table 2).

Table 1
Retrospective HIV screening rates among veteran patients with different types of substance use disorders

Substance use disorder history (October 1999–May 2005)	*n*	% screened for HIV
Any alcohol use		
Yes	357,576	19
No	14,173	25
Any illicit drug use		
Cocaine use		
Yes	132,354	33
No	239,395	12
Opiate use		
Yes	67,723	35
No	304,026	16
Amphetamine use		
Yes	24,586	37
No	347,163	18

3.3. HIV screening rates by patient characteristics

Although Table 2 shows that patients with any ED visits had the highest HIV screening rates (24% and 25%), in multivariate mixed modeling (detailed results available online), HIV screening was most likely among those who had had at least two PC visits as well as any ED visits (Dookeran et al., 2009). Table 2 also shows higher HIV screening rates among patients who had been in any outpatient substance use programs and among those who had received any inpatient care. Patients who were younger, female, Black, or not currently married, and patients who lacked housing or lived within 30 miles of the VA, also had higher HIV screening rates.

4. Discussion

In our nationwide retrospective study of a veteran population, the most striking finding was that only 20% of patients with illicit drug and/or alcohol use disorders in VA care between 2004 and 2005 had evidence of ever having being screened for HIV. We found this despite longstanding quality of care guidelines recommending HIV screening in substance abusers and the fact that the VA, the largest integrated health care delivery system in the United States, is a leader in performance and overall quality of care (Asch et al., 2004; Branson et al., 2006; NIAAA, 2002; Oliver, 2007). We found evidence of HIV screening in only 11% of patients with alcohol use disorders alone and in 25% to 28% of those with illicit drug use disorders. Even among those with more access to care (greater use of PC and ED services, use of substance use programs, or any inpatient care), fewer than one third had evidence of HIV screening. These findings support the need for more widespread interventions to expand routine voluntary HIV screening nationally, within and outside of the VA (Goetz, Hoang, et al., 2008; Sanders et al., 2005).

Our findings are from an exceptionally comprehensive sample: more than 350,000 veterans with substance abuse from VA facilities throughout the United States. In comparison, among veterans with illicit drug use disorders at four selected VA medical centers, 48% had evidence of testing (Owens et al., 2007). The lower percentage we found nationally is quite similar to results from two national surveys showing that 22% and 27% of individuals with HIV risk behaviors self-reported HIV testing, respectively (CDC, 2004). The higher but still suboptimal HIV screening rates at selected VA sites could be because the facilities are more urban, more academic in focus, or may serve patients with different socioeconomic profiles.

There are many potential obstacles to HIV testing. In our study, even among those with the highest utilization of PC and the ED, only 24% had been screened for HIV. In PC and other clinical settings in which substance abusers receive care, the focus may be on addressing more urgent and

Table 2
Retrospective HIV screening rates by demographic and other characteristics of veteran patients with substance use disorders

Population characteristics	*n*	% screened for HIV
Substance use disorder history (October 1999–May 2005)		
Illicit drug use and alcohol use	146,822	32
Illicit drug use alone	14,173	25
Alcohol use alone	210,754	11
PC/ED visits (June 2004–May 2005)[a]		
At least 2 PC, any ED visits	116,799	24
At least 2 PC, no ED visits	138,950	14
Less than 2 PC, any ED visits	57,141	25
Less than 2 PC, no ED visits	58,859	18
Any outpatient substance use program (June 2004–May 2005)		
Yes	96,268	28
No	275,481	17
Any inpatient care (June 2004–May 2005)		
Yes	107,166	28
No	264,583	16
Age group (years)		
<35	8,703	23
35–44	36,357	27
45–54	125,554	27
55–64	143,582	17
65>	57,553	6
Gender		
Female	12,171	29
Male	359,578	19
Race/ethnicity		
Black	94,735	29
Hispanic	20,143	27
Pacific Islander/American Indian/Other	10,776	23
White	190,950	16
Not reported	55,145	12
Marital status		
Currently married	104,873	12
All others[a]	266,876	23
Lack of housing		
Yes	115,479	32
No	256,270	14
Distance of residence from VA facility		
Live more than 30 miles away	151,073	16
Live within 30 miles[b]	220,676	22
Income group		
Above 2× poverty line[c]	102,794	16
Up to 2× poverty line	74,366	18
Up to poverty line	128,250	22
No income reported[d]	66,339	22
Insurance		
None or not recorded	264,591	22
Medicaid	3,493	29
Medicare	68,029	14
Private/Other	35,636	13
Sexually transmitted disease history (October 1999–May 2005)		
Yes	13,179	44
No	358,570	19
Hepatitis B history (October 1999–May 2005)		
Yes	15,258	45
No	356,491	19
Hepatitis C history (October 1999–May 2005)		
Yes	80,898	38
No	290,851	14

[a] Includes married in the past, never married, and unknown marital status.
[b] Includes 129 patients for whom distance was not reported.
[c] Poverty line approximately $10,000 per year.
[d] Service-connected veterans are not asked about income.

emergent substance use and medical presenting complaints, thus pushing HIV testing lower in clinical priorities. There may also be logistical and administrative barriers to performing HIV testing, such as limited time and resources for obtaining the written, informed consent required in the VA system during the period of our study.

Low HIV screening rates also may reflect less understanding of HIV risk among patients with certain substance use disorders and sociodemographic characteristics not often thought of as associated with HIV. For example, lower screening rates in patients with alcohol use disorders may be related to patients and/or providers not recognizing the increased HIV risk (and the need for screening) in these patients (Cook & Clark, 2005; Ehrenstein et al., 2004; Rees et al., 2001; Stein et al., 2000). The high prevalence of heavy alcohol use in both veterans and nonveterans (7.5% and 6.5%, respectively) accentuates this concern (National Survey on Drug Use and Health, 2005). Similarly, the VA user population is older, and although cultural norms may suggest that HIV risk-taking—and therefore HIV infection—is rare in older adults, there are in fact growing numbers of HIV-infected older persons in the United States. HIV testing of older adults is cost-effective on a societal level and is clearly indicated (Paul et al., 2007; Sanders et al., 2008).

Although we believe our study design identified most cases of HIV screening, our results may have been influenced by several factors. Patients may have been tested earlier than the 1999–2005 period during which we looked for evidence of HIV screening. Additionally, in this analysis, we looked at rates of ever having been screened rather than tracking changes in HIV screening by time. Thus, it is possible screening rates may have improved in the latter years of the study period with increasing uptake of guidelines. Our data did not allow us to identify patients who refused HIV testing. To the extent that test refusal accounts for low screening rates, this may suggest need for better public education about the benefits of early HIV diagnosis.

We were also unable to identify veteran patients screened for HIV outside of the VA. We have no direct data on how many veteran HIV tests were missed because of this. However, regional VA data (southern Nevada and California) collected from an electronic clinical reminder found that less than 2% of the patients in PC with identified HIV risk who were not tested, reported previous tests outside the VA as the reason (Goetz, Hoang, et al., 2008). Finally, to conduct a study of this size, scope, and inclusiveness, we used the presence of two *ICD-9* diagnostic codes rather than direct clinical assessments to identify substance use disorders in VA patients (Miller et al., 2004).

Substantial efforts are underway to increase low rates of HIV testing within the VA, including improved targeted approaches toward those with substance use disorders and other HIV risk factors, and implementation of more routine HIV screening as now recommended by the CDC. Targeting those at high risk continues to be important to counsel patients about risk behaviors, increase acceptance of screening, and pursue more frequent screening in those at highest risk (Branson et al., 2006; Millen et al., 2008). Prioritizing those at highest behavioral risk can be an efficient use of limited health care dollars (Holtgrave, 2007). Although behavioral health services and substance abuse services in the United States are often fragmented, poorly reimbursed, and delivered in settings where systematic quality assessment and improvement are difficult, the VA is an exception to this (Anaya et al., 2004; Goetz, Bowman, et al., 2008; Rubenstein et al., 2000; Yano et al., 2007). Our results support greater integration of HIV screening services into care for patients with substance use disorders, including the many veterans in care with alcohol abuse. There is a need to create more venues for screening (such as walk-in testing) and to better standardize, streamline, and increase the capacity for testing (Goetz, Hoang, et al., 2008; Valdiserri, Rodriguez, & Holodniy, 2008). HIV testing in substance abusers may be streamlined by use of rapid HIV testing instead of conventional enzyme immunoassay in some clinical settings (Branson, 2003; Morbidity and Mortality Weekly Report [MMWR], 2007).

To conclude, our nationwide retrospective study found that only 20% of patients with substance use disorders receiving care in the VA had HIV status assessed, despite 2001 CDC and 2002 NIAAA guidelines recommending HIV screening at the time of the study. Although new CDC guidelines call for expansion to opt out routine HIV testing, there is still great need for more aggressive measures to improve the recognition of those at high risk, such as substance abusers, and to minimize the barriers to HIV screening within these groups. Otherwise, efforts to increase testing rates generally may focus on more compliant patients in low-risk groups. Greatest priority should be placed on integrating HIV screening into the care of patients with substance use disorders in many clinical settings, increasing capacity for HIV screening, as well as facilitating long-term continuity of care for these vulnerable populations.

Acknowledgments

This research was supported in part by the VA QUERI-HIV/Hepatitis Program, by the Health Services Research and Development Grant IIR 03-089-2, and by the VA Physician Fellowship in Health Services Research. We gratefully acknowledge the statistical support provided by Yuexin Chen, M.S. A portion of this research was presented in poster presentations at the 2007 New England Regional and National Society of General Internal Medicine Conferences.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsat.2009.03.003.

References

Anaya, H. D., Yano, E. M., et al. (2004). Early adoption of human immunodeficiency virus quality improvement in Veterans Affairs medical centers: Use of organizational surveys to measure readiness to change and adapt interventions to local priorities. *American Journal of Medical Quality*, *19*, 137−144.

Asch, S. M., McGlynn, E. A., et al. (2004). Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. *Annals of Internal Medicine*, *141*, 938−945.

Branson, B. M. (2003). Point-of-care rapid tests for HIV antibodies. *Journal of Laboratory Medicine*, *27*, 288−295.

Branson, B. M., Handsfield, H. H., et al. (2006). Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. *Morbidity & Mortality Weekly Report Recommendations and Reports*, *55*, 1−17 (quiz CE1-4).

CDC. (2001). Centers for Disease Control and Prevention (CDC): Revised guidelines for HIV counseling, testing, and referral. *Morbidity & Mortality Weekly Report Recommendations & Reports*, *50*, 1−57 (quiz CE1-19a1-CE6-19a1).

CDC. (2004). Centers for Disease Control and Prevention (CDC): Number of persons tested for HIV—United States, 2002. *Morbidity & Mortality Weekly Report*, *53*, 1110−1113.

Cook, R. L., & Clark, D. B. (2005). Is there an association between alcohol consumption and sexually transmitted diseases? A systematic review. *Sexually Transmitted Diseases*, *32*, 156−164.

Des Jarlais, D. C., Arasteh, K., et al. (2007). Convergence of HIV seroprevalence among injecting and non-injecting drug users in New York City. *AIDS*, *21*, 231−235.

DiClemente, R. J., Wingood, G. M., et al. (2002). Prevention interventions for HIV positive individuals. *Sexually Transmitted Infections*, *78*, 393−395.

Dookeran, N. M., Burgess, J. F., et al. (2009). HIV screening among substance abusing veterans in care: Appendix table for *Journal of Substance Abuse Treatment* manuscript. Retrieved March 30, 2009, from http://www.chqoer.research.va.gov/Patient_Center_Care.asp.

Ehrenstein, V., Horton, N. J., et al. (2004). Inconsistent condom use among HIV-infected patients with alcohol problems. *Drug and Alcohol Dependence*, *73*, 159−166.

Goetz, M. B., Bowman, C., et al. (2008). 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: QUERI Series. *Implement Sci*, *3*, 16.

Goetz, M. B., Hoang, T., et al. (2008). A system-wide intervention to improve HIV testing in the Veterans Health Administration. *Journal of General Internal Medicine*.

Holtgrave, D. R. (2007). Costs and consequences of the US Centers for Disease Control and prevention's recommendations for opt-out HIV testing. *PLoS Medicine*, *4*, e194.

Justice, A. C., Erdos, J., et al. (2006). The Veterans Affairs Healthcare System: A unique laboratory for observational and interventional research. *Medical Care*, *44*(8 Suppl 2), S7−S12.

Klein, D., Hurley, L. B., et al. (2003). Review of medical encounters in the 5 years before a diagnosis of HIV-1 infection: Implications for early detection. *Journal of Acquired Immune Deficiency Syndromes*, *32*, 143−152.

Liddicoat, R. V., Horton, N. J., et al. (2004). Assessing missed opportunities for HIV testing in medical settings. *Journal of General Internal Medicine*, *19*, 349−356.

Marks, G., Crepaz, N., et al. (2005). Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States: Implications for HIV prevention programs. *Journal of Acquired Immune Deficiency Syndromes*, *39*, 446−453.

Miller, D. R., Safford, M. M., et al. (2004). Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes Care*, *27*(Suppl 2), B10−B21.

Millen, J. C., Arbelaez, C., et al. (2008). Implications and impact of the new US Centers for Disease Control and Prevention HIV Testing Guidelines. *Current Infectious Diseases Report*, *10*, 157−163.

MMWR. (2007). Rapid HIV testing in outreach and other community settings—United States, 2004–2006. *MMWR - Morbidity & Mortality Weekly Report*, *56*, 1233−1237.

Neaigus, A., Miller, M., et al. (2001). Sexual transmission risk among noninjecting heroin users infected with human immunodeficiency virus or hepatitis C virus. *Journal of Infectious Diseases*, *184*, 359−363.

NIAAA. (2002). National Institute of Alcohol Abuse and Alcoholism (NIAAA): Alcohol and HIV/AIDS (2002). *Alcohol Alert, National Institute of Alcohol Abuse and Alcoholism*, *57*.

NSDUH. (2005). The National Survey on Drug Use and Health (NSDUH) Report. Substance use, dependence, and treatment among veterans. Retrieved September 3rd, 2007, from http://www.oas.samhsa.gov/2k5/vets/vets.htm.

Oliver, A. (2007). The Veterans Health Administration: An American success story? *Milbank Quarterly*, *85*, 5−35.

Owens, D. K., Sundaram, V., et al. (2007). HIV testing of at risk patients in a large integrated health care system. *Journal of General Internal Medicine*, *22*, 315−320.

Palella, F. J., Deloria-Knoll, M., et al. (2003). Survival benefit of initiating antiretroviral therapy in HIV-infected persons in different CD4+ cell strata. *Annals of Internal Medicine*, *138*, 620−626.

Paltiel, A. D., Weinstein, M. C., et al. (2005). Expanded screening for HIV in the United States—An analysis of cost-effectiveness. *New England Journal of Medicine*, *352*, 586−595.

Paul, S. M., Martin, R. M., et al. (2007). Changing trends in human immunodeficiency virus and acquired immunodeficiency syndrome in the population aged 50 and older. *Journal of American Geriatrics Society*, *55*, 1393−1397.

Rees, V., Saitz, R., et al. (2001). Association of alcohol consumption with HIV sex- and drug-risk behaviors among drug users. *Journal of Substance Abuse Treatment*, *21*, 129−134.

Rubenstein, L. V., Mittman, B. S., et al. (2000). From understanding health care provider behavior to improving health care: The QUERI framework for quality improvement. *Quality Enhancement Research Initiative Med Care*, *38*(6 Suppl 1), I129−I141.

Samet, J. H., Freedberg, K. A., et al. (2001). Understanding delay to medical care for HIV infection: The long-term non-presenter. *AIDS*, *15*, 77−85.

Samet, J. H., Mulvey, K. P., et al. (1999). HIV testing in substance abusers. *American Journal of Drug and Alcohol Abuse*, *25*, 269−280.

Sanders, G. D., Bayoumi, A. M., et al. (2005). Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. [see comment]. *New England Journal of Medicine*, *352*, 570−585.

Sanders, G. D., Bayoumi, A. M., et al. (2008). Cost-effectiveness of HIV screening in patients older than 55 years of age. *Annals of Internal Medicine*, *148*, 889−903.

Stein, M. D., Hanna, L., et al. (2000). Alcohol use patterns predict high-risk HIV behaviors among active injection drug users. *Journal of Substance Abuse Treatment*, *18*, 359−363.

Strathdee, S. A., & Sherman, S. G. (2003). The role of sexual transmission of HIV infection among injection and non-injection drug users. *Journal of Urban Health*, *80*(4 Suppl 3), iii7−iii14.

Valdiserri, R. O., Rodriguez, F., & Holodniy, M. (2008). Frequency of HIV screening in the Veterans Health Administration: Implications for early diagnosis of HIV infection. *AIDS Education and Prevention*, *20*, 258−264.

Yano, E. M., Simon, B. F., et al. (2007). The evolution of changes in primary care delivery underlying the Veterans Health Administration's quality transformation. *American Journal of Public Health*, *97*, 2151−2159.

Yeni, P. G., Hammer, S. M., et al. (2004). Treatment for adult HIV infection: 2004 recommendations of the International AIDS Society—USA Panel. *Journal of the American Medical Association*, *292*, 251−265.

ORIGINAL ARTICLE

Physicians' pain management confidence versus competence

Mark A. Douglass, PharmD
Gail M. Sanchez, PharmD, BCPS
Daniel P. Alford, MD, MPH
Gail Wilkes, RNC, MS, AOCN
Jeffrey L. Greenwald, MD

ABSTRACT

Objective: *To assess awareness of existing pain management guidelines and compare physicians' confidence versus competence in selected pain management skills.*

Design: *Prospective survey study.*

Setting: *A large urban tertiary medical center.*

Patients, participants: *All Department of Medicine interns, senior residents, and attending physicians were sent a questionnaire; the overall response rate was 30 percent (91/304).*

Interventions: *The questionnaire assessed physicians' awareness of the institution's pain management guidelines, their self-reported comfort level (confidence) with, and a knowledge assessment (competence) of three pain management skills (managing chronic-continuous pain, equianalgesic dose conversion, and managing breakthrough pain) using validated, standardized case vignettes.*

Main outcome measures: *A comparison of physicians' confidence with their competence in these pain management skills.*

Results: *A total of 23 percent (21/91) of the respondents reported an awareness of the institution's pain management guidelines. Interns were significantly less confident than senior residents in all three pain management skills ($p < 0.001$, 0.006, 0.02) but nonsignificantly more competent in two of three skills (chronic-continuous pain, dose conversion). Attendings were generally more confident and nonsignificantly more competent than senior residents in all three pain management skills.*

Conclusions: *The underutilization of the pain management guidelines illustrates that the mere existence of these resources as a means of ensuring optimal pain management is insufficient. Creative pain management educational initiatives are needed to address the disparity between physician confidence and competence.*

Key words: pain management, physician knowledge, physician confidence

INTRODUCTION

It is estimated that persistent pain will lead to some 40 million physician visits annually, costing the US economy approximately $100 billion/year.[1] Unrelieved pain can also result in more frequent clinic visits, increased frequency of hospital admissions, lengthened hospital stays, prolonged recovery time, lost productivity, severe psychosocial sequelae, and poor quality of life.[2]

In an effort to improve pain management practices at our institution and address the Joint Commission's standards for pain assessment and management, a comprehensive, multidisciplinary, pain management guideline for the treatment of acute and chronic pain was developed by the Boston Medical Center Department of Pharmacy and approved by the hospital's Pharmacy and Therapeutics Committee. This collaborative effort was implemented several years prior to this study and is readily available along with other clinical guidelines on the hospital's internal Web site. Despite the implementation of this institutional guideline, the inpatient acute care medicine service has scored low on a national patient satisfaction survey regarding pain management (ie, Press Ganey survey) compared with similar hospitals.[3]

Although the need to improve pain management practices is apparent, recognizing and adequately treating a patient's pain remains challenging for many healthcare providers. Several barriers exist

that may prevent patients from receiving adequate control of their pain: reluctance to administer opioid medications for nonmalignant pain; misconceptions regarding analgesic side effects; and an unclear understanding of tolerance, physical dependence, or addiction.[4] Furthermore, concerns over regulatory scrutiny by the state medical board may limit the willingness of the physician to use opioids in patients with continuous, unrelieved, moderate to severe pain.[5] A survey by Loder et al. revealed that physicians tend to report a relatively high comfort level with assessing, evaluating, and treating conditions associated with pain, though they are less comfortable with specific concepts such as opioid titration, managing patients on patient controlled analgesia (PCA) pumps, and performing equianalgesic conversion calculations.[6]

The evaluation of pain management attitudes, beliefs, and knowledge deficits among practice experienced healthcare providers has been well documented, especially in the setting of cancer patients.[7-10] However, there is a lack of published data comparing pain management confidence and competence between medical trainees and attending physicians in an acute care academic medical setting. Although pain management guidelines exist on our hospital intranet, the Department of Medicine has no formal pain management curriculum, and there exists a perceived need to improve physicians' pain management competence at our institution. Therefore, the objectives of this study were to evaluate physician confidence and competence regarding three selected pain management skills and assess the use of the hospitals pain management guidelines.

METHODS

To assess and evaluate physicians' confidence and competence regarding effective pain management practices at our medical center, we developed a comprehensive questionnaire that was electronically mailed to all interns (early in their first year of training), senior medical residents, and clinical attending physicians at our medical center. To maximize the questionnaire response rate, the electronic survey was sent on two occasions over one month. During this time period, there were no formal pain medicine educational activities provided at our institution. This study was approved by the Boston Medical Center Institutional Review Board.

The questionnaire was three pages in length and contained several sections. First, physicians were asked to rate their general knowledge on a five point Likert scale (none, minimal, adequate, above average, superior) and experience (none, little, moderate, considerable, extensive) with managing pain in hospitalized patients. Respondents were also asked to rate their perceived comfort level with eight core pain management skills, which have been identified in previously published reports.[6,9-11] They are: pain management in the elderly patient population, pain management in a patient with a history of substance abuse, managing patients with chronic-continuous pain, opioid titration, equianalgesic dose conversion, breakthrough dosing calculations for a patient with continuous pain, PCA prescribing, and prevention/management of opioid side effects. Physicians identified their comfort level as: uncomfortable, somewhat uncomfortable, somewhat comfortable, and comfortable.

The study investigators chose three of the eight core pain management skills as they covered areas of greatest perceived difficulties at our medical center among medical trainees. They included managing chronic continuous pain, equianalgesic dose conversion, and breakthrough pain dosing. Physicians were then provided with three case vignette exercises which had been previously developed at our institution and tested for validity and reliability.[12,13] Each exercise corresponded to one of the three selected core pain management skills and assessed the physician's competence in this area. To ascertain the utilization of the comprehensive system wide pain management resources, physicians were also asked about their awareness of, and how frequently they have used the hospital's pain management guidelines.

The Fisher's exact test was used to compare the self-reported confidence and competence measures for each pain management skill independently between physician groups and for the pair wise comparison of confidence and competence within physician groups. To preserve the overall level of significance at $p = 0.05$, a Bonferroni correction threshold of $p = 0.025$ was applied for the individual competence and confidence measures between physician groups, treating each of the three pain management skills independently. A correction of $p = 0.017$ was applied for the pair wise comparison of competence and confidence within physician groups.

RESULTS

All medical physician staffs received the questionnaire (n = 304). Overall, 30 percent (91/304) of the physicians completed the survey and returned the electronic questionnaire which included 60 percent (34/57) of the medical interns, 21 percent (30/145) of the senior medical residents, and 26 percent (27/102) of the attending physicians.

Compared with the attending physicians and senior medical residents, the interns reported a lower level of pain management knowledge and experience (Table 1). Only 23 percent of the physicians surveyed were aware of the hospital's pain management guidelines. Of those subjects who were aware of these guidelines, only 45 percent reported ever utilizing them.

The senior residents reported being less confident than attending physicians but more confident than interns with most of the pain management skills. Compared with attending physicians, the senior residents reported a similar degree of comfort toward managing patients with chronic-continuous pain ($p = 0.36$), performing equianalgesic conversions ($p = 1.0$), and determining appropriate breakthrough dosing ($p = 0.42$). In contrast, the interns were significantly less confident than the senior residents with the same pain management competencies; 12 vs 69 percent ($p < 0.001$), 26 vs 62 percent ($p = 0.006$), and 23 vs 53 percent ($p = 0.02$), respectively.

There were few significant differences noted between physician groups on the competency assessment (Figure 1). The attending physicians scored higher than the senior residents on all three core pain management skills: managing patients with chronic-continuous pain ($p = 0.1$), performing equianalgesic conversions ($p = 0.79$), and determining appropriate breakthrough dosing ($p = 0.04$). The senior residents scored lower than the interns on managing chronic continuous pain and performing equianalgesic dose conversion ($p = 0.1$, $p = 0.29$, respectively) but not on determining appropriate breakthrough dosing ($p = 0.53$), though none was statistically significant.

An inverse relationship between senior medical residents self-reported confidence level and competency assessment was noted for both the chronic continuous pain and equianalgesic conversion competencies (69 vs 23 percent, $p = 0.006$ and 62 vs 33 percent, $p = 0.04$, respectively). In contrast, a linear relationship was observed between the interns confidence level and competency assessment for the same pain management skills [(12 vs 45 percent, $p = 0.004$); (26 vs 50 percent, $p = 0.07$)]. Although the majority of attending

Table 1. Physicians' general pain management knowledge and experience

Survey item	Rating	New medical interns (percent) N = 34	Senior medical residents (percent) N = 30	Attending physicians (percent) N = 27
Self-reported knowledge of acute and chronic pain management	None	0	0	0
	Minimal	25 (74)	6 (20)	1 (4)
	Adequate	6 (18)	21 (70)	14 (56)
	Above average	3 (8)	3 (10)	10 (37)
	Superior	0	0	1 (3)
Experience managing acute care medicine patient pain	None	1 (3)	0	0
	Little	20 (59)	1 (3)	2 (7)
	Moderate	11 (32)	18 (60)	14 (52)
	Considerable	1 (3)	9 (30)	9 (33)
	Extensive	1 (3)	2 (7)	2 (8)

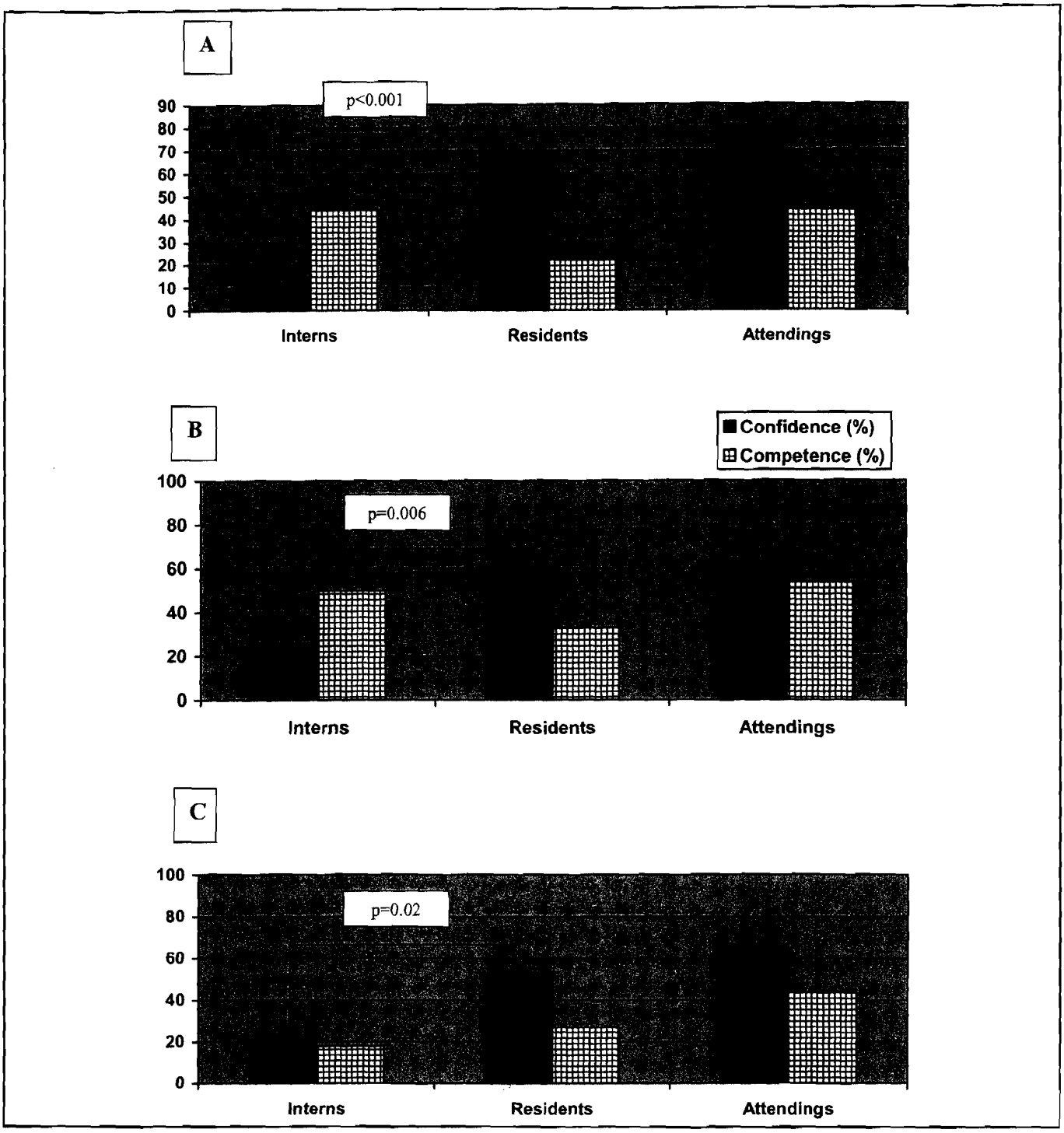


Figure 1. Competence vs confidence according to pain management skill. (A) Chronic continuous pain. (B) Equianalgesic conversion. (C) Breakthrough dosing.

physicians were confident managing chronic-continuous pain (81 percent) and determining appropriate breakthrough dosing (67 percent), fewer than half of them (44 percent) scored correctly on both of the competency assessments ($p = 0.01$ and $p = 0.17$, respectively).

DISCUSSION

The results of our study highlight the need for further comprehensive pain management education efforts targeting all acute care medicine physicians at our institution. We were not surprised to find that

inexperienced interns would be less confident with their pain management skills compared with their senior colleagues. However, the unexpected disparity between the senior residents' overconfidence and their relatively poor skill competency warrants further examination. Despite these findings, this phenomenon may not be unique to our institution. In a study comparing the confidence and accuracy of medical diagnoses among physicians in training, Friedman et al. found that inexperienced medical students were least comfortable with their medical diagnosis, while residents were more likely to be overconfident, yet less accurate than faculty physicians.[14]

Compared with the interns and senior residents, the attending physicians generally exhibited better performance on the objective knowledge assessment, however, pain management competence even here was inadequate. Although our study results did not demonstrate a significant difference in knowledge between the three physician groups for each of the pain management skills in an acute care medical setting, our findings of pain knowledge deficits are consistent with similar published studies.[9,10,15-17] On the basis of a palliative care knowledge assessment completed by 81 physician trainees, Mortimer et al. discovered that 75 percent of respondents were unable to correctly perform an equianalgesic dosage conversion.[16] Von Gunten et al. also observed poor housestaff performance on a cancer pain knowledge assessment questionnaire prior to an intervention program designed to improve pain management skills.[17]

Several important limitations of this study should be considered. One is the issue of response bias. It is possible that those physicians who chose not to complete the questionnaire were less confident in their knowledge of pain management. The low response rate and small sample size also makes it difficult to identify true differences between groups and caution must be exercised when interpreting the results of nonsignificant findings. Finally, the authors recognize that the assessment of physicians' competency in pain management skills through the use of an objective survey tool may not correspond with actual clinical decision making. As the study questionnaires were anonymous, it was not possible to evaluate physician competence in the clinical practice setting.

Despite these limitations, educational initiatives are clearly needed to address overall low performance in pain management skills at our institution. These initiatives must target the discrepancy between physician confidence and relatively low level of competency, irrespective of the level of training. System wide resources that were developed as a means to ensure optimal pain management, such as our hospital-specific pain management guidelines, are currently underutilized and clearly insufficient. Rather, these resources should be developed within existing decision support methodologies (real time alerts for inappropriate dosage form selection, dose range checking, template pain management order sets, etc) to enhance physician work flow and improve pain management performance. Recent pain management initiatives at our institution have included the addition of a Department of Neurology consult service and a grand rounds symposium, which included several focused pain management sessions provided by each physician discipline. The adoption of case-based learning strategies and those that utilize emerging technology such as clinical simulators could also be employed to assess physician knowledge and competency as a measure of effective pain management practices.[18-20]

Mark A. Douglass, PharmD, Department of Pharmacy Practice, Northeastern University School of Pharmacy, Boston, Massachusetts.

Gail M. Sanchez, PharmD, BCPS, Department of Pharmacy Services, Boston Medical Center, Boston, Massachusetts.

Daniel P. Alford, MD, MPH, Section of General Internal Medicine, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts.

Gail Wilkes, RNC, MS, AOCN, Division of Nursing, Boston Medical Center, Boston, Massachusetts.

Jeffrey L. Greenwald, MD, Section of General Internal Medicine, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts.

ACKNOWLEDGMENTS

The abstract of this manuscript was published and displayed in the form of a poster presentation at the 2006 American College of Clinical Pharmacists Spring Practice and Research Forum, Monterey, California.

REFERENCES

1. National Institutes of Health: New Directions in Pain Research. September 1998: 98-102.

2. Chandler S, Payne R: Economics of unrelieved cancer pain. *Am J Hosp Palliat Care.* 1998; 15(4): 223-225.

3. Pressganey.com [homepage on the Internet]. Indiana: Press Ganey Associates, Inc.; c2009. Available at *www.pressganey.com/.* Accessed February 18, 2009.

4. Potter M, Schafer S, Gonzalez-Mendez E, et al.: Opioids for chronic nonmalignant pain: Attitudes and practices of primary care physicians in the UCSF/Stanford collaborative research network. *J Fam Pract.* 2001; 50(2): 145-151.

5. Gallagher R: Opioids in chronic pain management: Navigating the clinical and regulatory challenges. *J Fam Pract.* 2004; 53 (Suppl 10): S23-S32.

6. Loder E, Witkower A, McAlary P, et al.: Rehabilitation hospital staff knowledge and attitudes regarding pain. *Am J Phys Med Rehabil.* 2003; 82(1): 65-68.

7. Von Roenn J, Cleeland C, Gonin R, et al.: Physician attitudes and practice in cancer pain management: A survey from the Eastern cooperative oncology group. *Ann Int Med.* 1993; 119(2): 121-126.

8. Elliott T, Murray D, Elliott B, et al.: Physician knowledge and attitudes about cancer pain management: A survey from the Minnesota Cancer Pain Project. *J Pain Symptom Manage.* 1995; 10(7): 494-504.

9. Lebovits A, Florence I, Bathina R, et al.: Pain knowledge and attitudes of healthcare providers: Practice characteristic differences. *Clin J Pain.* 1997; 13(3): 237-243.

10. Ponte C, Johnson-Tribino J: Attitudes and knowledge about pain: An assessment of West Virginia family physicians. *Fam Med.* 2005; 37(7): 477-480.

11. Levin M, Berry J, Leither J: Management of pain in terminally ill patients: Physician reports of knowledge, attitudes, and behavior. *J Pain Symptom Manage.* 1998; 15: 27-40.

12. Wilkes G, Lasch K, Lee J, et al.: Evaluation of a cancer pain education module. *Oncol Nurs Forum.* 2003; 30(6): 1037-1043.

13. Lasch K, Wilkes G, Lee J, et al.: Is hands-on experience more effective than didactic workshops in postgraduate cancer pain education? *J Cancer Educ.* 2000; 15: 218-222.

14. Friedman CP, Gatti GG, Franz TM, et al.: Do physicians know when their diagnoses are correct? *J Gen Intern Med.* 2005; 20: 334-339.

15. Xue Y, Schulman-Green, D, Czaplinski C, et al.: Pain attitudes and knowledge among RN's, pharmacists, and physicians on an inpatient oncology service. *Clin J Onc Nurs.* 2007; 11(5): 687-695.

16. Mortimer JE, Bartlett NL: Assessment of knowledge about cancer pain management by physicians in training. *J Pain Symptom Manage.* 1997; 14(1): 21-28.

17. Von Gunten CF, Von Roenn JH, Weitzman S: Housestaff training in cancer pain education. *J Cancer Educ.* 1995; 9(4): 230-234.

18. Harasym PH, Tsai TC, Hemmati P: Current trends in developing medical students' critical thinking abilities. *Kaohsiung J Med Sci.* 2008; 24: 341-355.

19. Boulet JR: Summative assessment in medicine: The promise of simulation for high-stakes evaluation. *Acad Emerg Med.* 2008; 15: 1017-1024.

20. Srinivasan M, Hwang JC, West D, et al.: Assessment of clinical skills using simulator technologies. *Acad Psychiatry.* 2006; 30: 505-515.

Clopidogrel, Genetics, and Drug Responsiveness

Jane E. Freedman, M.D., and Elaine M. Hylek, M.D., M.P.H.

Despite great progress in the diagnosis and treatment of unstable coronary syndromes, it is estimated that 785,000 Americans will have new acute cardiac events and 470,000 will have recurrent events this year.[1] Central to the pathogenesis of acute coronary syndromes is the adhesion and activation of platelets leading to aggregation, thrombus formation, and vessel occlusion. Unfortunately, the absolute risk of recurrent vascular events among patients taking platelet inhibitors remains relatively high.[1]

The observation that platelet-dependent thrombosis occurs despite treatment with platelet inhibitors has led to a large number of studies assessing the cause of these treatment failures. Often termed "resistance," treatment failure in patients taking aspirin or clopidogrel has been ascribed to myriad causes, including nonadherence to drug regimens, inadequate doses of drugs, and coexisting medical conditions. Such a lack of response to therapy may affect 5 to 45% of patients.[2] It has also become apparent that heritable factors play a major role in determining endogenous platelet function[3] and that the platelet-activation response varies widely among patients. However, genetic variants in platelet receptors have not been consistently shown to influence platelet function, alter drug response, or be associated with cardiovascular disease.[4]

For clopidogrel, an inhibitor of platelet $P2Y_{12}$ receptor, there are data suggesting that genetics may affect drug responsiveness and efficacy. The responsible genetic variant appears to occur not in the expected $P2Y_{12}$ receptor but, rather, in an enzyme responsible for the metabolism of the drug. Clopidogrel is a prodrug that requires activation by specific hepatic cytochrome P-450 (CYP) enzymes. The genes encoding the CYP-dependent oxidative steps are polymorphic, and previous studies have shown that carriers of the specific alleles of *CYP2C19* and *CYP3A4* have a diminished response to the antiplatelet effects of clopidogrel.[5-8] A reduced response to clopidogrel has been specifically associated with the *CYP2C19*2* allele, which causes loss of function, in patients after coronary-stent placement[9] and after myocardial infarction without ST elevation.[10] It is consistent with these pharmacodynamic findings that prasugrel, another $P2Y_{12}$ inhibitor, appears to be unaffected by variability in CYP2C19 isoenzymes.[11] (At the time of this writing, prasugrel was not approved by the Food and Drug Administration.) It is also consistent with the importance of hepatic enzymes that the coadministration of omeprazole, which is metabolized by CYP2C19, has been shown to decrease the platelet-inhibitory effect of clopidogrel.[12]

Two articles in this issue of the *Journal* contribute to our understanding of the genetic variation in these enzymes in regulating the actions and efficacy of clopidogrel. In a study conducted by the French Registry of Acute ST-Elevation or Non–ST-Elevation Myocardial Infarction (FAST-MI) investigators (ClinicalTrials.gov number, NCT00673036), Simon et al.[13] report on a cohort of more than 2200 clopidogrel-treated patients who presented with acute myocardial infarction. The investigators, who looked at the relationship between genetic variants that are potentially relevant to platelet function and clinical outcome during a 1-year period, found that patients carrying any two *CYP2C19* loss-of-function alleles (*2, *3, *4, or *5) had a higher event rate. Carriers of the *ABCB1* variant that modulates clopidogrel absorption also had a modestly increased rate of events. However, they found no association with polymorphisms of $P2Y_{12}$ or glycoprotein IIb/IIIa or with coadministration of omeprazole.

In another study, Mega et al.[14] investigated the association between genetic variants in CYP genes, the plasma concentration of active metabolite, and platelet function in healthy subjects. In a subsequent analysis, they examined the association between CYP genetic variants and cardiovascular outcomes in nearly 1500 patients who presented with an acute coronary syndrome and who were treated with clopidogrel during the earlier Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition with Prasugrel–Thrombolysis in Myocardial Infarction (TRITON–TIMI) 38 (NCT00097591). They found that in healthy subjects, carriers of at least one *CYP2C19* loss-of-function allele had decreased levels of the active clopidogrel metabolite and less reduction in platelet aggregation, as compared with noncarriers. In clopidogrel-treated subjects from TRITON–TIMI 38, carriers of the

loss-of-function alleles had an increased risk of death from cardiovascular causes, myocardial infarction, or stroke, as compared with noncarriers.

In addition, subjects in TRITON–TIMI 38 who carried the *CYP2C19*2* allele had a risk of stent thrombosis that was three times that of noncarriers. In this study, the event curves diverged soon after treatment with clopidogrel, a finding that was consistent with the potential immediate loss of a platelet-inhibitory effect. Inconsistent with these observations is the fact that no trend was found for increased bleeding among noncarriers of the *CYP2C19* variant; however, the numbers of patients were small and the definition of hemorrhage was potentially too stringent to discern a difference.

Several of the findings in these studies are inconsistent with those of past reports. Previously, carriers of *CYP3A4* were also noted to have a reduced response to clopidogrel.[8] Although there were some differences in the populations studied, the clinical end points in the current studies[13,14] should supersede surrogates for thrombosis. Also, as described by Simon et al., the use of proton-pump inhibitors had no effect on the clinical response to clopidogrel. The causes of the discrepancy in this finding between the study by Simon et al. and the previous reports are unclear. However, until this question can be answered in a larger set of patients with clear clinical outcomes, the use of clinical outcomes, as opposed to platelet-function testing, provides some reassurance.

These two studies raise many pivotal questions. Could the loss of effect that was seen with the genetic variant be overcome by increasing the dose of clopidogrel? The mean dose in the French study was 300 mg per day; in TRITON–TIMI 38, a standard dose of 300 mg per day was given, with a discharge dose of 75 mg per day in both studies. Would patients with a loss-of-function CYP variant have improved platelet function and clinical outcomes (thrombosis and hemorrhage) with an alternative platelet inhibitor (such as prasugrel) that does not require similar hepatic transformation? Would genetic testing and adjustments in the dose or type of therapy enhance efficacy? The data currently available cannot answer these questions. Until a prospective study is completed demonstrating how best to treat patients, particularly those who have poor metabolism of clopidogrel, it is not clear that routine genetic testing will be clinically or fiscally advantageous.

In summary, these studies demonstrate that patients with loss-of-function genetic variants have altered pharmacokinetic and pharmacodynamic responses to clopidogrel and an increased cardiac risk that persists after adjustment for other known potential risk factors. What is striking about these two studies is their concordance despite distinct differences in populations of patients. The fundamental observations are similar: that loss-of-function *CYP2C19* alleles are associated with an increased risk of acute cardiovascular events, particularly among patients undergoing percutaneous coronary intervention. Since these genetic variants are common in the general population, this observation is not trivial. To optimally guide the selection of therapy, we must await further information concerning genetic testing, the role of extended genotypic classification, dose adjustment, and the effect of tailored therapeutic selection on thrombotic and hemorrhagic outcomes across a wide spectrum of patients in clinical practice.

Dr. Freedman reports receiving grant support from Boehringer Ingelheim; and Dr. Hylek, receiving grant support from AstraZeneca, serving as an advisor to Boehringer Ingelheim, Bristol-Myers Squibb, Sanofi-Aventis, and the Medicines Company, and participating in clinical symposia sponsored by Bayer and Bristol-Myers Squibb. No other potential conflict of interest relevant to this article was reported.

From the Department of Medicine, Boston University School of Medicine, Boston.

1. Lloyd-Jones D, Adams R, Carnethon M, et al. Heart disease and stroke statistics — 2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 2008 December 15 (Epub ahead of print).
2. Mason PJ, Jacobs AK, Freedman JE. Aspirin resistance and atherothrombotic disease. J Am Coll Cardiol 2005;46:986-93.
3. Frelinger AL III, Furman MI, Linden MD, et al. Residual arachidonic acid-induced platelet activation via an adenosine diphosphate-dependent but cyclooxygenase-1- and cyclooxygenase-2-independent pathway: a 700-patient study of aspirin resistance. Circulation 2006;113:2888-96.
4. O'Donnell CJ, Larson MG, Feng D, et al. Genetic and environmental contributions to platelet aggregation: the Framingham Heart Study. Circulation 2001;103:3051-6.
5. Giusti B, Gori AM, Marcucci R, et al. Cytochrome P450 2C19 loss-of-function polymorphism, but not CYP3A4 IVS10 + 12G/A and P2Y12 T744C polymorphisms, is associated with response variability to dual antiplatelet treatment in high-risk vascular patients. Pharmacogenet Genomics 2007;17:1057-64.
6. Kim KA, Park PW, Hong SJ, Park JY. The effect of CYP2C19 polymorphism on the pharmacokinetics and pharmacodynamics of clopidogrel: a possible mechanism for clopidogrel resistance. Clin Pharmacol Ther 2008;84:236-42.
7. Hulot JS, Bura A, Villard E, et al. Cytochrome P450 2C19 loss-of-function polymorphism is a major determinant of clopidogrel responsiveness in healthy subjects. Blood 2006;108:2244-7.
8. Angiolillo DJ, Fernandez-Ortiz A, Bernardo E, et al. Contribution of gene sequence variations of the hepatic cytochrome P450 3A4 enzyme to variability in individual responsiveness to clopidogrel. Arterioscler Thromb Vasc Biol 2006;26:1895-900.

9. Trenk D, Hochholzer W, Fromm MF, et al. Cytochrome P450 2C19 681G>A polymorphism and high on-clopidogrel platelet reactivity associated with adverse 1-year clinical outcome of elective percutaneous coronary intervention with drug-eluting or bare-metal stents. J Am Coll Cardiol 2008;51:1925-34.
10. Frere C, Cuisset T, Morange PE, et al. Effect of cytochrome p450 polymorphisms on platelet reactivity after treatment with clopidogrel in acute coronary syndrome. Am J Cardiol 2008; 101:1088-93.
11. Brandt JT, Close SL, Iturria SJ, et al. Common polymorphisms of CYP2C19 and CYP2C9 affect the pharmacokinetic and pharmacodynamic response to clopidogrel but not prasugrel. J Thromb Haemost 2007;5:2429-36.
12. Gilard M, Arnaud B, Cornily JC, et al. Influence of omeprazole on the antiplatelet action of clopidogrel associated with aspirin: the randomized, double-blind OCLA (Omeprazole CLopidogrel Aspirin) study. J Am Coll Cardiol 2008;51:256-60.
13. Simon T, Verstuyft C, Mary-Krause M, et al. Genetic determinants of response to clopidogrel and cardiovascular events. N Engl J Med 2009;360:363-75.
14. Mega JL, Close SL, Wiviott SD, et al. Cytochrome P-450 polymorphisms and response to clopidogrel. N Engl J Med 2009; 360:354-62.

Evaluating the Effects of Ambient Air Pollution on Life Expectancy

Daniel Krewski, Ph.D.

Air pollution is an important determinant of population health. In this issue of the *Journal*, Pope et al.[1] provide data that once again reinforce this fundamental concept. In an analysis that correlates reductions in fine particulate matter (i.e., particles less than 2.5 μm in aerodynamic diameter, or $PM_{2.5}$) in the air with life expectancies, the investigators found that a decrease in the concentration of $PM_{2.5}$ of 10 μg per cubic meter is associated with an increase in life expectancy of 0.77 year. Their analysis is based on correlating reductions in particulate air pollution over the past several decades with increases in life expectancy in 217 counties in 51 metropolitan areas in the United States. Although ecologic in nature (i.e., reflecting associations between air pollution and life expectancy at the county rather than the individual level), these results appear to be robust with respect to adjustment for changes in socioeconomic, demographic, and smoking patterns occurring over the same period.

The finding is comparable with previous predictions of reductions in life expectancy of 1.11 years in the Netherlands,[2] 1.37 years in Finland,[3] and 0.80 year in Canada[4] resulting from increases in ambient $PM_{2.5}$ concentrations of 10 μg per cubic meter. However, the strength of the study by Pope et al. resides in its ability to demonstrate an increase in life expectancy resulting from actual reductions in particulate air pollution. This finding provides direct confirmation of the population health benefits of mitigating air pollution and greatly strengthens the foundation of the argument for air-quality management.[5]

This work could be extended to take into account quality of life. For example, Coyle et al.[4] estimated that an increase of 10 μg per cubic meter in $PM_{2.5}$ concentrations would lead to a quality-adjusted reduction in life expectancy of 0.60 year, as compared with the unadjusted reduction of 0.80 year. The work by Pope et al. represents an important contribution to the large and growing body of evidence linking ambient air pollution with adverse health outcomes. At the global level, the World Health Organization[6] estimates that 1.4% of all deaths and 0.8% of disability-adjusted life-years are the result of particulate air pollution.

The short-term health effects of particulate and gaseous air pollutants have been well documented, largely through time-series studies relating short-term elevations in ambient levels of such pollutants to increases in morbidity and mortality from cardiorespiratory conditions. A recent combined analysis of time-series data from 124 of the largest cities in North America and Europe produced an estimated increase in the rate of death from any cause ranging from 0.2 to 0.6% for an increase in ambient PM_{10} concentrations of 10 μg per cubic meter,[7] depending on the assumed lag time between exposure to particulate matter and death and on the method used for seasonality control, the form of the temporal smoothing function, and degree of smoothing. Risk estimates for Europe and the United States were similar but were higher in Canada.

The long-term effects of exposure to "criteria" air pollutants (particulate matter, ozone, sulfates, sulfur dioxide, nitrous oxides, and carbon monoxide) have been documented in large-scale cohort studies, including the Harvard Six Cities Study[8] and the American Cancer Society Cancer

CORRESPONDENCE

Warfarin Pharmacogenetics

TO THE EDITOR: The study of a pharmacogenetic algorithm for estimating the appropriate initial dose of warfarin, reported by the International Warfarin Pharmacogenetics Consortium (Feb. 19 issue),[1] highlights the challenges in fashioning a generally applicable algorithm for warfarin dosing. The authors used an imprecise end point — namely, the dose predicted to achieve a stable international normalized ratio (INR). The investigators conclude that their model would be most helpful in the case of patients for whom the stable, therapeutic warfarin dose is less than 22 mg or more than 48 mg per week, but would be less useful for patients requiring intermediate doses. Although the genetic data improved the accuracy of available models for predicting warfarin dose, whether the addition of genetic testing can reduce the risk of bleeding or thrombosis, or both, is the key question, and the answer remains unknown.

In clinical practice, careful monitoring allows individualized dose adjustment during the initiation of warfarin therapy. Genetic testing cannot eliminate important hazards such as fragmented transitions of care, concomitant antiplatelet therapy, and socioeconomic barriers to frequent INR measurement. Moreover, the findings of the only high-quality, comparative trial of genotyping among patients starting warfarin therapy cast serious doubt on the hypothesis that genetics-based warfarin dosing will reduce adverse outcomes.[2] These facts, along with a recent analysis of the cost-effectiveness of genetics-based warfarin dosing,[3] leave us highly uncertain about whether patients stand to benefit from genetic testing in routine practice.

David A. Garcia, M.D.

University of New Mexico Health Sciences Center
Albuquerque, NM 87107

Elaine Hylek, M.D., M.P.H.

Boston University Medical Center
Boston, MA 02118

1. The International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. N Engl J Med 2009;360:753-64.
2. Anderson JL, Horne BD, Stevens SM, et al. Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. Circulation 2007;116:2563-70.
3. Eckman MH, Rosand J, Greenberg SM, Gage BF. Cost-effectiveness of using pharmacogenetic information in warfarin dosing for patients with nonvalvular atrial fibrillation. Ann Intern Med 2009;150:73-83.

TO THE EDITOR: The study by the International Warfarin Pharmacogenetics Consortium, as well as an article published previously,[1] indicates that genetics plays an important role in warfarin therapy. We are concerned about the higher dosing and accelerated response in elderly patients. The current recommendation is to use an even lower dose in elderly patients than that used in the younger adult population when treatment is initiated.[2] How was age treated as a covariate? In addition,

THIS WEEK'S LETTERS

2474 Warfarin Pharmacogenetics
2475 Treating Childhood Anxiety
2477 C3 Polymorphisms and Outcomes of Renal Allografts
2479 Dronedarone in Atrial Fibrillation
2481 A Pregnant Woman with Fever after a Trip to Africa
2482 An Unusual Outbreak of Hypoglycemia
2483 Insurance-Industry Investments in Tobacco

it would be useful to know whether levels of factor VII and of proteins C and S influence the extremes of warfarin dosing.

Asit B. Shil, M.D.
Maureen P. Strohm, M.D.
University of Southern California Keck School of Medicine
Los Angeles, CA 90015
shil@usc.edu

1. Schwarz UI, Ritchie MD, Bradford Y, et al. Genetic determinants of response to warfarin during Initial anticoagulation. N Engl J Med 2008;358:999-1008.
2. Ansell J, Hirsh J, Hylek E, et al. Pharmacology and management of the vitamin K antagonists. Chest 2008;133:160S-98S.

THE AUTHORS REPLY: We agree with Garcia and Hylek that it is uncertain whether genetic-based warfarin dosing will improve outcomes, and this dosing approach cannot solve the other management problems related to warfarin dosing (e.g., fragmented care). Randomized, controlled trials comparing genotype-guided care with a nongenetic approach are planned in the United States, Europe, and Korea and will be a critical step in showing that genotype-based dosing can improve anticoagulation control. Ultimately, very large trials may be needed to determine the effect of genotyping on clinical outcomes.

Nonetheless, it is not difficult to envision that genetics-guided warfarin dosing could improve important and costly outcomes, such as extra clinic and emergency room visits. Even a reduction in minor bleeding could improve a person's quality of life and decrease the need to discontinue a highly effective therapy. Our study suggests that the use of genetics might benefit nearly half the patients who are initiating warfarin therapy. Thus, there is the potential that genetic-guided warfarin dosing could prove to be cost-effective, particularly among patients at high risk for hemorrhage. In addition, as the cost of genotyping decreases, cost-effectiveness could be further enhanced.

In response to Shil and Strohm's point about age, our algorithms do estimate lower doses in the elderly. We also agree that there are other factors, both genetic and nongenetic, that may influence warfarin dosing requirements, particularly at the extremes.

Although we do not yet have all the answers regarding the value of genotype-guided warfarin dosing in clinical practice, our study provides an understanding of its potential benefit. Like most diagnostic tests, genetic testing will not benefit all persons. Clinicians must assess the current level of evidence and decide whether to implement genetics-guided warfarin dosing in practice now, await the results of the INR-focused clinical trials to adopt this approach, or adopt it only once differential clinical outcomes have been documented. Many diagnostic tests are widely adopted in practice before differential clinical outcomes have been documented; whether this will occur with warfarin dosing remains to be seen.

Teri E. Klein, Ph.D.
Stanford University
Stanford, CA 94143
Stephen E. Kimmel, M.D., M.S.C.E.
University of Pennsylvania
Philadelphia, PA 19104
Julie A. Johnson, Pharm.D.
University of Florida
Gainesville, FL 32610
teri.klein@stanford.edu

Behavioral Therapy, Sertraline, or Both in Childhood Anxiety

TO THE EDITOR: Walkup and colleagues (Dec. 25 issue)[1] conclude that the three active therapies they studied — a combination of cognitive behavioral therapy and sertraline, cognitive behavioral therapy alone, and sertraline alone — were effective treatment for anxiety in children, as compared with placebo. The authors further conclude that combination treatment had a superior response rate, as compared with active treatment alone. However, the study design invites questions. There was no treatment group in which cognitive behavioral therapy plus placebo was used. The absence of such a group prevented the investigators from determining whether the addition of sertraline to cognitive behavioral therapy resulted in more improvement than each treatment given separately because of an additive effect of two active treatments or because of the placebo effect of adding a pill to cognitive behavioral therapy.

Furthermore, the children who were given a pill without cognitive behavioral therapy did not

BMC Research Notes

BioMed Central

Short Report

Lifetime intimate partner violence exposure, attitudes and comfort among Canadian health professions students

Megan R Gerber*†[1,2] and André KW Tan†[3]

Address: [1]Women's Health, VA Boston Healthcare System, Boston, MA, USA, [2]Harvard Medical School, Boston, MA, USA and [3]Department of Otolaryngology, Queen's University, Kingston, Ontario, Canada

Email: Megan R Gerber* - meggerber@post.harvard.edu; André KW Tan - akwt@queensu.ca

* Corresponding author †Equal contributors

Published: 23 September 2009

BMC Research Notes 2009, **2**:191 doi:10.1186/1756-0500-2-191

Received: 6 April 2009
Accepted: 23 September 2009

This article is available from: http://www.biomedcentral.com/1756-0500/2/191

Abstract

Background: Intimate partner violence (IPV) is a widespread public health problem and training of health professions students has become common. Understanding students' prior knowledge, attitudes and personal exposure to IPV will aid educators in designing more effective curriculum. As interprofessional educational efforts proliferate, understanding differences across disciplines will be critical.

Findings: Students in the schools of Medicine, Nursing and Rehabilitation at a university in Ontario attend an annual daylong interprofessional IPV training. To measure perceived role and comfort with IPV and prior personal exposure, we administered a brief Likert scale survey to a convenience sample of students over three years. 552 students completed the survey; the overall response rate was 73%. The majority (82%) agreed that it was their role to intervene in cases of IPV; however Rehabilitation students expressed lower overall comfort levels than did their peers in other schools ($p < .0001$). Gender, age and prior training on the subject were not significant predictors of comfort. Seven percent reported lifetime IPV and one-fifth had witnessed IPV, but these exposures did not predict comfort in adjusted logistic regression models.

Conclusion: While the majority of professional students believe it is their role to address IPV in clinical practice, comfort level varied significantly by field of study. More than one fifth of the students reported some personal exposure to IPV. However this did not impact their level of comfort in addressing this issue. Educators need to take students' preexisting attitudes and personal exposure into account when planning curriculum initiatives in this area.

Background

Intimate partner violence (IPV) is a pattern of coercive behavior in which one person attempts to control another through threats or actual use of physical violence, sexual assault and verbal or psychological abuse. [1] Nearly one-third of Canadian women experience IPV in their lifetime and 21.2% report IPV in the preceding 5 years. [2,3] In Canadian family practice settings, the estimated prevalence is 14.6%. [4] IPV has well-established adverse health effects, [5-7] and results in frequent and regular contact between victims and healthcare providers. [4,8] It has thus become widely accepted that training of healthcare professionals is imperative. [9,10] Yet, the sensitive nature of IPV creates challenges for educators who train health professions students. [11-14] The clear limitations of the medical model to provide a straightforward remedy, or

"fix", for this problem may be frustrating to many learners. [13] Given the well-documented difficulty many healthcare providers have with inquiry for IPV, [15-17] it would not be surprising to find a dearth of effective role models available during clinical training.

Further adding to these challenges is the possibility of personal exposure to IPV among students. Students who have been victims may experience a range of responses to IPV curricular content including anxiety, vicarious retraumatization and feelings of helplessness. [11,14] Medical students with personal histories of violence express concern about their future efficacy in aiding patients who have had similar experiences. [18] Nonetheless, students who report histories of abuse favor IPV training. [19] In order to provide effective learner-centered curricula, educators need to understand the potential extent of IPV exposure among students.

Curricula to address IPV have proliferated over the last 15 years [9,20] and are most commonly reported in medical and nursing school settings. [21-24] Fewer citations are found for the field of physical therapy and rehabilitation. [25,26]

The prevalence of IPV among US medical students is between 6-12% for women [18,19] and 7% for men[27,28] In a US study of nursing students, 8% reported experiencing IPV[29] Among practicing physicians and nurses in Ontario, nearly 50% reported either personally experiencing or witnessing a close friend or relative experience abuse. [16] Thus, it is also probable that some proportion of Canadian students will have been exposed to IPV, [11,14,18] but to our knowledge, rates have not been reported in the literature.

Aims of the Study

The main objectives of the study are to explore how student comfort in addressing IPV is impacted by 1) gender, 2) program of study and 3) prior personal experience or training. A secondary aim was to measure students' understanding of the dynamics of abusive relationships and ascertain whether this differs across program of study. Students in the schools of Nursing, Medicine and Rehabilitation attending a one-day workshop on IPV completed a brief survey in order to provide some preliminary data to address these study questions.

Methods

A daylong interprofessional workshop on IPV is held annually at a large university in Ontario, Canada. Students from the Schools of Nursing, Medicine and Physical Rehabilitation attend the mandatory workshop. Students are warned of the potential for disturbing material and offered on-site resources. Counselors attend the workshop and are available to assist any student in immediate need of support. A voluntary brief, confidential Likert-scale survey was distributed to students at the morning break during the one-day workshop over the three study years (2003-2005). The medical and rehabilitation students were in their second year of graduate training, the nursing students were in the third year of an undergraduate program. Attendees answered basic demographic questions about age, country of origin and current school. They were also asked about any prior training pertaining to IPV. We included two questions about students' personal experience and history of witnessing of IPV:

1) "Have you ever been physically abused by an intimate partner?"

2) "Have you ever directly witnessed physical abuse in a relationship?"

We also queried students about their level of comfort with inquiry about IPV. Response categories included "strongly agree", "agree", "neutral", "disagree" and "strongly disagree". A dichotomous variable for comfort was created with the two agreement categories being used to model presence of comfort in addressing IPV. Summary and descriptive statistics were performed to examine basic demographic characteristics, attitudes toward and prevalence of IPV. The secondary aim of characterizing students' understanding of abusive relationship dynamics was addressed by measuring agreement with the statement, "I don't understand why victims remain in abusive relationships." Bivariate analyses examined whether rates of IPV varied by gender, country of origin and school. We examined potential predictors of student comfort with inquiry for IPV using logistic regression analysis. This model was adjusted for age, gender, country of birth; prior training for IPV, school, year the survey was taken and history of IPV or being a witness to IPV. The University Research Ethics Board approved the study. All analyses were conducted using SAS Version 9.1 (Cary, N.C.).

Results

Over a three-year period, a total of 552 students completed the survey; 37% of the students were medical students, 33% were rehabilitation students and the remaining 30% were nursing students (Table 1). The overall response rate was 73%. The majority of the health professions students attending the workshops over the three years were female (n = 415/552, 76%). Most of the students reported no prior IPV training (n = 338, 61%); for those who had training, the most common source was undergraduate education (n = 86/214, 40%). Medical students had the highest rate of previous training. (Table 1)

The majority of students (82%) in all schools expressed the belief that it was their role to intervene on behalf of abused patients (Table 1), but the rehabilitation students

Table 1: Characteristics of Students by Professional School N = 552

Variable	Medical N = 208 (%)	Rehabilitation N = 181 (%)	Nursing N = 163 (%)	P value
Mean Age (SD)§	24 (3.8)	25 (2.6)	21 (4.5)	< .0001
Gender (REFERENCE = female)	103 (50%)	158 (87%)	154 (94%)	< .0001
Born in Canada	175 (84%)	159 (88%)	137 (85%)	.51
Prior IPV training (REFERENCE = none)	139 (67%)	103 (57%)	96 (59%)	.11
Personal IPV history	18 (9%)	11 (6%)	9‡ (5.5%)	.44
Witnessed IPV	50 (24%)	37* (21%)	36 (22%)	.71
Summary of Attitudes and Comfort: percentage responding Agree Strongly or Agree				
"I don't understand why victims remain in abusive relationships."	36 (17%)	38 (21%)	19 (12%)	.02
"It is my role to intervene if a patient has been abused."	171 (82%)	145 (80%)	136 (83%)	.24
"I feel comfortable asking patients about IPV."	99 (48%)	56 (31%)	78 (48%)	< .0001

§Standard deviation is given in parentheses for age variable only.
‡ Missing 1 response on this question only (n = 162)
* Missing 1 response on this question only (n = 180)

expressed a lower self-report of comfort level than both nursing and medical students (Table 2). Similarly, the rehabilitation students were more likely to endorse a lack of understanding as to why someone would remain in an abusive relationship (Table 1). In the adjusted analyses, the only significant predictor of student-reported comfort was enrollment in either nursing or medical school. Gender, age, prior report of IPV training, year of workshop attendance and personal history of IPV were not predictors of comfort with IPV inquiry (Table 2).

Overall, a total of 38 students (7%) reported lifetime IPV; the majority of these were female (n = 30/38, 79%). However, this was not statistically significant ($\chi^2 = 3.77$, p = .15). Medical students had the highest rate of lifetime IPV (Table 1). Fewer foreign-born students (4%) reported IPV than did their Canadian counterparts (7%).

One fifth of all the students witnessed IPV at some point in their lives; the highest percentage was again found among medical students (Table 1). More female students reported witnessing IPV, however the difference was not statistically significant ($\chi^2 = 3.63$, p = 0.16).

Discussion

While the majority of students in our study agreed that it is their role to address IPV in clinical practice, knowledge and attitudes varied across schools. Age, prior training and even personal exposure to IPV did not change the relationship between field of study and comfort level with this issue. Rehabilitation students expressed lower comfort levels that may, in part, correspond to their report of less prior IPV training; however in logistic regression analysis, field of study remained a significant predictor of comfort even when prior training was controlled for. While rehabilitation students clearly viewed addressing IPV as part of their professional purview, their expressed comfort level and understanding of the dynamics of abusive relationships lagged behind those of nursing and medical students. This finding is unlikely due to level of study alone since rehabilitation students were second year postgraduate students comparable in age to the medical students, while the nursing students were younger undergraduates.

Table 2: Predictors of Student Reported Comfort with IPV Inquiry

Covariate	Unadjusted O.R. (95% C.I.)	Adjusted O.R. (95% C.I.)
Gender (REFERENCE = Female)	1.08 (0.73-1.59)	1.34 (0.84-2.12)
Age	0.99 (0.95-1.03)	1.01 (0.96-1.06)
Prior Training (REFERENCE = none)	0.76 (0.54-1.08)	0.74 (0.52-1.06)
Year of workshop attendance	1.17 (0.95-1.43)	1.09 (0.88-1.35)
School (REFERENCE = Rehab)	0.49 (0.33-0.75)	0.45 (0.28-0.70)
Country of Birth (REFERENCE = Canada)	1.073 (0.66-1.73)	1.10 (0.67-1.81)
Witness IPV	0.68 (0.45-1.01)	0.76 (0.49-1.17)
Lifetime IPV	0.50 (0.26-0.98)	0.64 (0.31-1.33)

Literature searches reveal a relative lack of publication in this field (compared to medicine and nursing) which may contribute to reduced awareness and familiarity among those entering this field. Personal exposure to IPV was a significant predictor of reduced comfort in unadjusted analyses, but this relationship did not remain significant with adjustment for potential confounders.

While prior work has shown that female medical students were more likely than their male counterparts to report prior IPV exposure, [18] this finding did not achieve statistical significance in our study. Interestingly, the school reporting the highest rates of IPV (Medicine) was also the group with the highest percentage of male students. Rates of exposure to lifetime IPV are known to increase with age, but in our study, age alone was unlikely responsible for medical students' higher reported IPV rates since the mean age of rehabilitation students was comparable.

It is notable that the rate of lifetime IPV measured in this study is significantly lower than that reported in Canadian population studies but comparable to estimates among U.S student cohorts. One possible explanation for this is the "healthy worker effect" theory, which posits that those with abuse histories may have lower educational attainment due to the adverse effects of the abuse, and thus be less likely to participate in professional training, lowering the rate of IPV in such populations. [28,30] Another potential contributing factor is underreporting of abuse history by students due to our administration of the survey in an open lecture hall with proximate seating of other students.

Foreign-born students' reported rates of IPV are similar to those of Canadian- born students. Since we did not query length of residency in Canada, we were unable to assess the level of acculturation of these students which may impact rates of IPV. Lower rates of IPV have been found among foreign born women in population-based studies in Canada [2] but the foreign born students enrolled in Canadian professional schools likely have higher levels of language proficiency and literacy than their counterparts in the general population.

Limitations

This study has a number of limitations. Because the survey was administered during the workshop, students may have had privacy concerns when completing it, possibly resulting in response bias. We could not query 12 month (current) IPV separately because students attending the training who had intimate relationships with fellow students could have been seated together in the lecture hall, limiting the safety of inquiry about current IPV. Selection bias may have occurred because questionnaire completion was voluntary. Another concern is our measurement of IPV. Due to the need for brevity, we used one question to ascertain prior exposure to physical IPV and one to query witnessing IPV. Neither question has been validated. The lack of questions about emotional abuse also likely underestimated the true prevalence of IPV in this population. The students from the different schools were all at different levels in their training, thus unmeasured effects of clinical experience could have impacted some of their expressed knowledge and attitudes about IPV. Moreover, interprofessional educational initiatives remain unusual, so the findings from this study may not readily generalize to other more traditional teaching settings.

Conclusion

Our study presents novel data regarding Canadian professional students and IPV which may aid educators developing curriculum in this field. While the majority of all students believed that it was their role to address IPV; further study of rehabilitation students, who will go on to work with vulnerable populations, is needed to explain why this group differs in expressed comfort and understanding of the dynamics of abusive relationships.

While we may have underestimated the true prevalence of IPV in this cohort, our study affirms that a proportion of Canadian health professions students are likely to have experienced IPV. IPV may also be a more salient issue for male students than previously described. Our findings require replication with validated, confidential measures. Study of this issue across Canadian institutions could also better inform educational initiatives in this challenging field. Future work should examine which teaching methods may be most effective for learners who have been victims or witnesses to IPV.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MG assisted with the design of the survey, carried out the statistical analyses and drafted the manuscript. AT developed and ran the workshop, assisted with design of the survey, oversaw administration of the survey and helped to draft the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

The authors wish to thank Nicholas C. Bala LLB, LLM, Diane M. Buchanan, RN, M.Sc.N., Ph.D., Wendy Pentland M.Ed, PhD, and David M.C. Walker, MD, FRCPC for their assistance and support of the workshop. The authors thank Ms. Mary Beth Gurr for her administrative work with both the workshop and survey. This work was presented in abstract form at the 2008 Canadian Conference on Medical Education May 3-7, 2008, Montréal, Canada.

References

1. **Centers for Disease Control** [http://www.cdc.gov/ViolencePrevention/intimatepartnerviolence/definitions.html]
2. Cohen MM, Maclean H: **Violence against Canadian Women.** *BMC Womens Health* 2004, **4(Suppl 1):**S22.
3. Romans S, Forte T, Cohen MM, Du Mont J, Hyman I: **Who is most at risk for intimate partner violence? A Canadian population-based study.** *J Interpers Violence* 2007, **22(12):**1495-514.
4. Ahmad F, Hogg-Johnson S, Stewart DE, Levinson W: **Violence involving intimate partners: prevalence in Canadian family practice.** *Can Fam Physician* 2007, **53(3):**461-8. 460
5. United States. Center for Disease Control [CDC]: **Adverse health conditions and health risk behaviors associated with intimate partner violence--United States, 2005.** *MMWR Morb Mortal Wkly Rep* 2008, **57(5):**113-7.
6. Coker AL, Davis KE, Arias I, Desai S, Sanderson M, Brandt HM, Smith PH: **Physical and mental health effects of intimate partner violence for men and women.** *Am J Prev Med* 2002, **23(4):**260-8.
7. Bonomi AE, Anderson ML, Rivara FP, Thompson RS: **Health outcomes in women with physical and sexual intimate partner violence exposure.** *J Womens Health (Larchmt)* 2007, **16(7):**987-97.
8. Ulrich YC, Cain KC, Sugg NK, Rivara FP, Rubanowice DM, Thompson RS: **Medical care utilization patterns in women with diagnosed domestic violence.** *Am J Prev Med* 2003, **24(1):**9-15.
9. Hamberger LK: **Preparing the next generation of physicians: medical school and residency-based intimate partner violence curriculum and evaluation.** *Trauma Violence Abuse* 2007, **8(2):**214-25.
10. Alpert EJ: **Making a place for teaching about family violence in medical school.** *Acad Med* 1995, **70(11):**974-8.
11. Lawler MK: **Teaching about family violence when the trainee is also a survivor.** *Fam Med* 1999, **31(2):**87-9.
12. Dickstein LJ: **Practical recommendations for supporting medical students and faculty in learning about family violence.** *Acad Med* 1997, **72(1 Suppl):**S105-9.
13. Lenahan P, Shapiro J: **Facilitating the emotional education of medical students: using literature and film in training about intimate partner violence.** *Fam Med* 2005, **37(8):**543-5.
14. Phelps BP: **Helping medical students help survivors of domestic violence.** *JAMA* 2000, **283(9):**1199.
15. Gerber MR, Leiter KS, Hermann RC, Bor DH: **How and why community hospital clinicians document a positive screen for intimate partner violence: a cross-sectional study.** *BMC Fam Pract* 2005, **6:**48.
16. Gutmanis I, Beynon C, Tutty L, Wathen CN, MacMillan HL: **Factors influencing identification of and response to intimate partner violence: a survey of physicians and nurses.** *BMC Public Health* 2007, **7:**12.
17. Jaffee KD, Epling JW, Grant W, Ghandour RM, Callendar E: **Physician-identified barriers to intimate partner violence screening.** *J Womens Health (Larchmt)* 2005, **14(8):**713-20.
18. Ambuel B, Butler D, Hamberger LK, Lawrence S, Guse CE E: **Female and male medical students' exposure to violence: Impact on well-being and perceived capacity to help battered women.** *Journal of Comparative Family Studies* 2003, **34:**113-135.
19. Cullinane PM, Alpert EJ, Freund KM: **First-year medical students' knowledge of, attitudes toward, and personal histories of family violence.** *Acad Med* 1997, **72(1):**48-50.
20. Alpert EJ, Tonkin AE, Seeherman AM, Holtz HA: **Family violence curricula in U.S. medical schools.** *Am J Prev Med* 1998, **14(4):**273-82.
21. Davila YR: **Teaching nursing students to assess and intervene for domestic violence.** *Int J Nurs Educ Scholarsh* 2005, **2:**. Article 4
22. Davidson LL, Grisso JA, Garcia-Moreno C, Garcia J, King VJ, Marchant S: **Training programs for healthcare professionals in domestic violence.** *J Womens Health Gend Based Med* 2001, **10(10):**953-69.
23. Woodtli MA, Breslin ET: **Violence-related content in the nursing curriculum: a follow-up national survey.** *J Nurs Educ* 2002, **41(8):**340-8.
24. Weiss LB, Kripke EN, Coonse HL, O'Brien MK: **Integrating a domestic violence education program into a medical school curriculum: challenges and strategies.** *Teach Learn Med* 2000, **12(3):**133-40.
25. McComas J: **Violence and abuse education: toward changes to physiotherapy curricula.** *Physiother Can* 1994, **46(4):**241-6.
26. Clark TJ, McKenna LS, Jewell MJ: **Physical therapists' recognition of battered women in clinical settings.** *Phys Ther* 1996, **76(1):**12-8. discussion 18-9
27. Ernst AA, Houry D, Nick TG, Weiss SJ: **Domestic violence awareness and prevalence in a first-year medical school class.** *Acad Emerg Med* 1998, **5(1):**64-8.
28. Frank E, Elon L, Saltzman LE, Houry D, McMahon P, Doyle J: **Clinical and personal intimate partner violence training experiences of U.S. medical students.** *J Womens Health (Larchmt)* 2006, **15(9):**1071-9.
29. Attala JM, Oetker D, McSweeney M: **Partner abuse against female nursing students.** *J Psychosoc Nurs Ment Health Serv* 1995, **33(1):**17-24.
30. Doyle JP, Frank E, Saltzman LE, McMahon PM, Fielding BD: **Domestic violence and sexual abuse in women physicians: associated medical, psychiatric, and professional difficulties.** *J Womens Health Gend Based Med* 1999, **8(7):**955-65.

Predictors of Retention in HIV Care Among a National Cohort of US Veterans

Thomas P. Giordano,[1,2] Christine Hartman,[1,2] Allen L. Gifford,[3] Lisa I. Backus,[4] and Robert O. Morgan[1,2*]

[1]Baylor College of Medicine, Houston, Texas, USA; [2]Houston Center for Quality of Care and Utilization Studies, Health Services Research and Development Service, Michael E. DeBakey Department of Veterans Affairs Medical Center, Houston, Texas, USA; [3]VA New England Healthcare and Boston University, Bedford, Massachusetts, USA; [4]Center for Quality Management in Public Health, Department of Veterans Affairs, Palo Alto, California, USA

Background: Poor retention in HIV care leads to poor survival. The predictors of poor retention in HIV care are not well understood, especially from US nationwide datasets. We determined the predictors of poor retention in HIV care among a group of US veterans and examined whether poor retention was confounded by other predictors of survival. **Methods:** We conducted a retrospective cohort study of 2,619 male US veterans who started antiretroviral therapy after January 1, 1998. Poor retention in HIV care was defined as having had at least 1 quarter-year without any primary care visit in the year after starting antiretroviral therapy. Survival was assessed through 2002. Logistic regression and Cox models were constructed. **Results:** Thirty-six percent of patients had poor retention in care. In multivariable analysis, younger age, Black race/ethnicity, CD4 cell count >350 x10^6/L, hepatitis C infection, and illicit drug use were predictive of poor retention in care. Having a chronic medical comorbidity and being identified as a man having sex with men (MSM) were associated with improved retention in care. In multivariable survival analyses, poor retention in care was not a confounder or moderator for other variables that predicted survival. **Conclusions:** Retention in HIV care is an independent predictor of survival. As routine HIV screening increases, more people with the characteristics predictive of poor retention in care will be identified. Interventions to improve retention in care are needed. **Key words:** *adherence, cohort study, HIV/AIDS, survival, Veterans Affairs*

HIV-infected patients who are poorly adherent to physician visits are less likely to receive HAART, have lower adherence to antiretroviral therapy (ART), are more likely to develop an infection with resistance to HAART, and are less likely to achieve HIV suppression.[1–5] Using national data from the US Department of Veterans Affairs (VA) health care system, we recently showed that poor retention in HIV care was predictive of less desirable changes in CD4 cell count and HIV RNA concentration ("viral load") while on ART and worse overall survival.[6] Others have reached similar results[7] and extended them to patients newly entering care for HIV infection.[8] Given these findings, interventions to improve retention in care are urgently needed.

Little is known about the predictors of poor retention in HIV care, especially from US nationwide patient datasets. We sought to determine the factors predictive of poor retention in care among a group of US veteran patients. It is important to determine predictors of poor retention so that interventions used in the context of clinical trials or quality improvement initiatives can be targeted, especially since such projects often have limited resources. We also sought to determine whether retention in care explains the effects of other predictors of poorer survival with HIV infection, including baseline HIV disease severity, race/ethnicity, substance use, psychiatric comorbidity, and socioeconomic instability. If, for example, it was known that retention in care mediates the effect of substance use on survival, interventions targeted at reducing substance use might be particularly successful at improving survival.

*Dr. Morgan is currently at University of Texas Health Science Center Houston School of Public Health.

Address for correspondence: Thomas P. Giordano, MD, MPH, Houston Center for Quality of Care and Utilization Studies, Michael E. DeBakey VA Medical Center (152), 2002 Holcombe Boulevard, Houston, Texas 77030 USA. Email: tpg@bcm.tmc.edu

HIV Clin Trials 2009;10(5):299–305

doi: 10.1310/hct1005-299

METHODS

The methods for this study are fully described elsewhere.[6] Briefly, we used the VA's Immunology Case Registry to conduct a retrospective cohort study involving persons newly identified as having HIV infection during 1997–1998 at any VA hospital or clinic in the United States. To be included in the study, patients had to have started ART after January 1, 1997, seen a clinician at least once after receiving their first ART prescription, have a baseline CD4 cell count result available, and have survived for at least 1 year. Patients were divided into four groups on the basis of the number of quarters during the year in which they had at least one HIV primary care visit. Survival was assessed through 2002 using VA databases and the National Death Index. Because data were available for only a small number of women, they were excluded. Because HIV RNA concentration was missing for 383 (14.6%) of the subjects and was not likely missing at random, we did not include that variable in the analyses.

We examined baseline characteristics using data up to and including the first HIV primary care visit after ART was started. To determine the presence of medical, substance use, and socioeconomic conditions, we examined *International Classification of Diseases, Ninth Revision,* codes (**Table 1**). We selected five of the most common chronic medical comorbidities among veterans for individual study (diabetes, hypertension, ischemic heart disease, cerebrovascular disease, and chronic obstructive pulmonary disease) to determine whether the need for chronic medical care for those conditions influences retention in HIV care. Because the prevalence of each condition was relatively uncommon at baseline (hypertension was most common, at 9.7%), we created a variable that represented the presence of any one of these conditions. There is no single ICD-9 code for "AIDS," so it was defined by the presence of the ICD-9 codes for many of the 1993 AIDS-defining conditions. We estimate that between 70% and 80% of AIDS-defining conditions would be captured by these ICD-9 codes.[9,10] Hepatitis C virus infection was defined as a positive antibody test result. HAART use was defined as the use of a protease inhibitor, a non-nucleoside reverse transcriptase inhibitor, or a combination of zidovudine, lamivudine, and abacavir within 30 days of the first ART prescription.

Because our previous work showed that an HIV primary care visit during the first year after starting ART in fewer than four quarters was associated with a statistically significant decrease in survival,[6] we dichotomized retention in care at four quarters or fewer than four quarters to assess the independent predictors of poor retention in care. Categorical data were compared with the chi-square test, and a multivariate logistic regression model was created. To assess the impact of retention in care on other known predictors of survival with HIV infection, we used the method of Baron and Kenny to determine whether retention in care was a moderator in that relationship.[11] First, in a Cox proportional hazard model that did not include the variables representing retention in care, we determined what those other predictors of survival were in our dataset. Next we determined whether these predictors were associated with retention in care using the logistic regression models described previously. Finally, we constructed a Cox proportional hazards model of survival that included these variables and the variables representing the number of quarters in care to assess whether the relationship between those predictors and the outcome was moderated by retention in care. Statistics were analyzed with SAS software (SAS Institute, Cary, North Carolina, USA). The study was approved by the Institutional Review Board for Baylor College of Medicine and Affiliated Institutions, as well as by the VA. Individual informed consent was not required.

RESULTS

The cohort included 2,619 patients. Sixty-four percent were seen in all four quarters, and 36% were seen in fewer than four quarters. As shown in **Table 1**, in univariate analysis, younger age, HIV risk factor other than men having sex with men (MSM), Black or Hispanic race/ethnicity, higher CD4 cell count, hepatitis C infection, illicit drug use, and alcohol abuse were predictors of poor retention in care, while the presence of any of the chronic medical comorbidities and a history of AIDS predicted good retention in care. In multivariable analysis, younger age, Black race/ethnicity, CD4 cell count >350 $x10^6$/L, hepatitis C infection, and illicit drug use were predictive of being seen in fewer than four quarters during the

Table 1. List of conditions assessed by ICD-9 codes and their corresponding codes

Condition	ICD-9 codes	Condition	ICD-9 codes
AIDS-defining conditions		**Chronic medical conditions**	
Candidiasis, pulmonary	112.4	Diabetes mellitus	250
Candidiasis, esophageal	112.84	Hypertension	401-405.99, 437.2
Coccidioidomycosis, disseminated or extrapulmonary	114.1-114.3, 114.9	Ischemic heart disease	410-414
Histoplasmosis, disseminated or extrapulmonary	115.01-115.04; 115.11-115.14;	Cerebrovascular disease	430-438
Kaposi sarcoma	115.91-115.94 176	Chronic obstructive pulmonary disease	491, 492, 493.2, 494, 496
Lymphoma, Burkitt's or immunoblastic (or equivalent terms)	200, 202.0, 202.1, 202.8	**Other conditions**	
Lymphoma, primary, of brain	200.5	Alcohol abuse (alcohol abuse, dependence, or intoxication; alcoholic psychoses, gastritis, fatty liver, or hepatitis	291, 303.0, 303.9, 305.0, 535.3, 571.0, 571.1
Mycobacterium avium complex, or other mycobacteria, disseminated or extrapulmonary	031.2	Illicit drug use (opioid, cocaine, or amphetamine or other stimulant dependence or abuse)	304.0, 304.2, 304.4, 304.7, 305.5, 305.6, 305.7
Mycobacterium tuberculosis, any site	010-018	Psychiatric disease (dementia, psychoses, or depression; neurotic personality, stress, anxiety, conduct, or developmental disorders)	290, 293-302, 306-319
Pneumocystis carinii (jiroveci) pneumonia	136.3		
Progressive multifocal leukoencephalopathy	046.3	Socioeconomic challenges or instability (housing, economic circumstances; family or psychosocial circumstances prompting seeking of medical care)	V60, V61, V62
Toxoplasmosis of the brain	130		

Note: Conditions not included in the "AIDS" variable include cervical cancer (no women in cohort); extrapulmonary cryptococcosis, cytomegalovirus disease other than liver, spleen, or nodes, cytomegalovirus retinitis with loss of vision, HIV-related encephalopathy, and wasting syndrome due to HIV (not specific enough ICD-9 codes); and chronic intestinal cryptosporidiosis (>1 month), chronic ulcers (>1 month) or bronchitis, pneumonitis, or esophagitis due to herpes simplex virus, chronic intestinal (>1 month) isosporiasis, recurrent pneumonia, and recurrent salmonella septicemia (unable to define chronicity and/or recurrence by ICD-9 code).

first year. Having a medical comorbidity and being identified as MSM were associated with improved retention in care.

To determine whether retention in care moderates the effect of other characteristics associated with poor survival with HIV infection, we constructed multivariable models. Using the method of Baron and Kenny,[11] we ran three sets of models (**Table 2**). First, in a model that did not include the variables representing the number of quarters in care, we found that increasing age, decreasing CD4 cell count, and chronic medical comorbidity were predictive of poorer survival (columns 7 and 8 of **Table 2**). Next we showed that these variables, among others, were associated with retention in care (columns 5 and 6 of **Table 2**). Finally, we constructed a model that included these variables and the variables representing the

number of quarters in care (columns 9 and 10 of **Table 2**). Increasing age, decreasing baseline CD4 cell count, any chronic medical comorbidity, and poorer retention in care were strongly predictive of increased hazard of death ($p \leq .01$), while illicit drug use was weakly predictive of decreased hazard of death ($p = .03$). In comparing the survival models with and without the retention in care variables, none of the adjusted hazard ratios or p values changed substantially when adjusted for the number of quarters in care, indicating that retention in care is independently predictive of survival and is not confounded by other variables nor is it a mediator in the pathway by which they affect survival.

DISCUSSION

We studied retention in care in the VA health care system, a system that has few barriers to care for eligible veterans. Thirty-six percent of veterans had poor retention in care at 1 year, and we found that the predictors of poor retention in care are younger age and less advanced HIV disease, Black race, and substance use. MSM and men with chronic medical conditions had better retention in care. Retention in care appears to be an independent predictor of survival that is not a confounder and does not mediate the effects of other predictors of survival.

The finding that Black patients had more difficulty remaining in care, even after adjusting for socioeconomic status, disease severity, age, and other factors, is discouraging. The VA health care system provides access to low-cost care, though some copayments and other real and opportunity costs are incurred by the patient. We could not account for these costs in the analysis, and it is possible that these costs were a barrier for more of the Black patients than the other patients. Unmeasured noneconomic barriers may also contribute to this finding, including for example different levels of mistrust or stigma.[12–14] More work is needed to fully explain this observation, because differences in retention in care may contribute to race-based disparities in outcomes.[15,16]

The other predictors of poor retention in care were substance use or its markers, such as hepatitis C infection. Others have found substance use associated with poor retention in care.[17] These factors also often cause difficulty with adherence to ART. Adherence to ART and retention in care are, in fact, related behaviors: both are chronic, complex behaviors requiring accurate information, internal and external motivation and support, and an array of behavioral skills, including regular interaction with and navigation of the health care system. These conceptual similarities and the overlapping predictors suggest that interventions to improve adherence might benefit from focus on retention in care and vice versa. In contrast, medical comorbidities and more advanced HIV disease were predictive of better retention. This finding offers some assurance that the sickest patients are getting the care they need. Psychiatric disease was not associated with either retention in care or survival. Treatment for psychiatric disease may be at least partially responsible for these negative findings.

Adjusting for retention in care had little effect on the hazard ratios for the other variables in the model, indicating that it is not a confounder or mediator of their effects. In other words, substance use or more advanced HIV disease at baseline does not lead to worse survival through poorer retention in HIV care. Retention in care is associated with adherence to ART,[6] which clearly impacts survival; retention in care is easily and accurately ascertained in routine care and should be considered when identifying patients at high risk for poor outcomes. Further work is needed to disentangle the effects of retention in care and adherence to ART, but clearly they are both important forms of adherence to HIV care.

This study supplements and improves upon our earlier work in a number of ways. First, we herein report the results of a multivariable model of the predictors of poor retention in care. Second, the earlier report used a comorbidity index[18] rather than specific comorbidities; the latter are more useful clinically. To our knowledge, no other studies have assessed the impact of medical comorbidities on retention in care. Third, the present study demonstrates that retention in care is an independent predictor of survival, not a mediator or confounder for other predictors of survival with HIV infection. The study has limitations. We could not account for incarceration or transfer out of VA care, though the latter is unlikely for patients accessing ART through the VA.[19] These data are from early in the HAART era, but it is not likely that patterns and associations we observed would be substantially

Table 2. Univariate and multivariate analyses of characteristics predictive of poor retention in care and death among 2,619 veterans

	Predictors of retention in care					Predictors of death			
						Model without retention in care variables		Model with retention in care variables	
Characteristic	Visits in all 4 quarters ($n=1685$)	Visits in <4 quarters ($n=934$)	Univariate p value	Multivariate odds ratio (95% CI)	Multivariate p value	Adjusted hazard ratio (95% CI)	p value	Adjusted hazard ratio (95% CI)	p value
Age, years			<.001						
21–29	4.3	7.7		2.87 (1.95, 4.22)	<.001	Referent		Referent	
30–39	23.1	28.3		1.75 (1.38, 2.23)	<.001	2.46 (0.99, 6.13)	.05	2.60 (1.04, 6.47)	.04
40–49	40.8	44.0		1.41 (1.13, 1.75)	<.01	3.58 (1.46, 8.78)	<.01	3.87 (1.58, 9.49)	<.01
50+	31.8	20.0		Referent		5.13 (2.09, 12.6)	<.001	5.73 (2.33, 14.1)	<.001
Race			<.001						
White	36.4	29.2		Referent		Referent		Referent	
Black	50.7	60.3		1.34 (1.11, 1.62)	<.01	1.19 (0.95, 1.49)	.13	1.17 (0.93, 1.46)	.18
Hispanic	8.6	7.5		1.01 (0.73, 1.40)	.96	1.35 (0.94, 1.96)	.11	1.35 (0.93, 1.95)	.11
Other/ unknown	4.4	3.0		0.82 (0.51, 1.30)	.39	0.79 (0.42, 1.51)	.48	0.83 (0.43, 1.58)	.57
HIV risk factor			<.01						
Sex with men	25.0	19.2		0.73 (0.59, 0.90)	<.01	0.83 (0.64, 1.10)	.19	0.86 (0.65, 1.12)	.26
IDU	18.1	19.7		0.82 (0.65, 1.04)	.09	1.22 (0.94, 1.57)	.13	1.24 (0.96, 1.60)	.09
Other/ unknown	56.9	61.1		Referent		Referent		Referent	
CD4+ cell count, x10^6/L			<.001						
≤200	48.1	40.3		Referent		2.31 (1.77, 3.00)	<.001	2.35 (1.82, 3.05)	<.001
201 to 350	22.3	24.6		1.19 (0.96, 1.47)	.11	1.32 (0.96, 1.81)	.09	1.32 (0.97, 1.82)	.08
>350	29.6	35.1		1.25 (1.02, 1.52)	.03	Referent		Referent	
History of AIDS	15.3	11.3	<.01	0.81 (0.63, 1.05)	.11	0.99 (0.76, 1.27)	.91	1.04 (0.81, 1.35)	.75
HAART prescribed	81.8	79.0	.09	0.94 (0.76, 1.15)	.53	0.80 (0.64, 1.01)	.06	0.82 (0.66, 1.04)	.10

(Continued)

Table 2. *Continued*

Chronic medical condition[a]	27.1	20.9	<.001	0.81 (0.66, 0.99)	.04	1.40 (1.13, 1.73)	<.01	1.42 (1.15, 1.76)	.001
Hepatitis C infection	20.2	26.7	<.001	1.32 (1.06, 1.64)	0.01	1.24 (0.98, 1.58)	.08	1.24 (0.97, 1.57)	.09
Alcohol abuse	21.3	30.3	<.0001	1.28 (0.99, 1.66)	0.06	1.17 (0.88, 1.56)	.28	1.13 (0.85, 1.50)	.40
Illicit drug use	17.1	27.3	<.0001	1.42 (1.08, 1.87)	0.01	0.74 (0.54, 1.01)	.05	0.71 (0.52, 0.97)	.03
Psychiatric disease	29.4	31.9	.19	0.91 (0.74, 1.11)	0.34	1.14 (0.90, 1.43)	.28	1.14 (0.91, 1.44)	.26
Socioeconomic instability	27.4	29.9	.18	0.90 (0.73, 1.10)	0.30	1.08 (0.86, 1.36)	.50	1.09 (0.87, 1.37)	.46
Retention in care									
Visit in 4 of 4 quarters	N/A	N/A		N/A		N/A		Referent	
Visits in 3 of 4 quarters	N/A	N/A		N/A		N/A		1.38 (1.07, 1.77)	.01
Visits in 2 of 4 quarters	N/A	N/A		N/A		N/A		1.65 (1.23, 2.23)	<.001
Visits in 1 of 4 quarters	N/A	N/A		N/A		N/A		1.92 (1.35, 2.74)	<.001

Note: HAART = highly active antiretroviral therapy; IDU = injection drug use. [a]Chronic medical condition includes diabetes, hypertension, ischemic heart disease, cerebrovascular disease, and chronic obstructive pulmonary disease. AIDS defining conditions are defined in Table 1.

different with more contemporary HAART. As reviewed earlier, potentially important influences on retention in care could not be assessed in this retrospective cohort study. We could not capture all AIDS-defining conditions, as noted in **Table 1**. This limitation may account for the low prevalence of a history of AIDS and its poor prognostic value in predicting survival in our data. The analyses excluded all patients who died within 1 year of their first visit after starting ART, so the low prevalence and poor prognostic ability are not completely unexpected.

Retention in HIV care is an important, independent predictor of survival with HIV infection. Persons at highest risk for poor retention in care are young, otherwise healthy, Black patients with less advanced HIV disease who use illicit drugs. As HIV screening becomes routine,[20] more people with these characteristics will be identified and retaining them in care may be difficult. Interventions to improve retention are needed.

ACKNOWLEDGMENTS

This research was supported by the National Institute of Mental Health, National Institutes of Health (K23MH067505), and, in part, the Houston VA HSR&D Center of Excellence (HFP90-020). The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. None of the authors have any associations that might pose a conflict of interest.

REFERENCES

1. Kleeberger CA, Buechner J, Palella F, et al. Changes in adherence to highly active antiretroviral therapy medications in the Multicenter AIDS Cohort Study. *AIDS*. 2004;18(4):683–688.
2. Wagner GJ, Kanouse DE, Koegel P, Sullivan G. Correlates of HIV antiretroviral adherence in persons with serious mental illness. *AIDS Care*. 2004;16(4):501–506.
3. Giordano TP, White AC Jr, Sajja P, et al. Factors associated with the use of highly active antiretroviral therapy in patients newly entering care in an urban clinic. *J Acquir Immune Defic Syndr*. 2003;32(4):399–405.
4. Lucas GM, Chaisson RE, Moore RD. Highly active antiretroviral therapy in a large urban clinic: risk factors for virologic failure and adverse drug reactions. *Ann Intern Med*. 1999;131(2):81–87.
5. Sethi AK, Celentano DD, Gange SJ, Moore RD, Gallant JE. Association between adherence to antiretroviral therapy and human immunodeficiency virus drug resistance. *Clin Infect Dis*. 2003;37(8):1112–1118.
6. Giordano TP, Gifford AL, White AC Jr, et al. Retention in care: a challenge to survival with HIV infection. *Clin Infect Dis*. 2007;44(11):1493–1499.
7. Park WB, Choe PG, Kim SH, et al. One-year adherence to clinic visits after highly active antiretroviral therapy: a predictor of clinical progress in HIV patients. *J Intern Med*. 2007;261(3):268–275.
8. 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):248–256.
9. Jones JL, Hanson D, Dworkin M, et al. Surveillance for AIDS-defining opportunistic illnesses, 1992-1997. *CDC Surveillance Summaries, MMWR*. 1999;48(No. SS-2).
10. Centers for Disease Control and Prevention. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR*. 1992;41(RR-17).
11. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 1986;51(6):1173–1182.
12. Whetten K, Leserman J, Whetten R, et al. Exploring lack of trust in care providers and the government as a barrier to health service use. *Am J Public Health*. 2006;96(4):716–721.
13. Clark A, Mayben JK, Hartman C, Kallen MA, Giordano TP. Conspiracy beliefs about HIV infection are common but not associated with delayed diagnosis or adherence to care. *AIDS Patient Care STDS*. 2008;22(9):753–759.
14. Ross MW, Essien EJ, Torres I. Conspiracy beliefs about the origin of HIV/AIDS in four racial/ethnic groups. *J Acquir Immune Defic Syndr*. 2006;41(3):342–344.
15. Mugavero MJ, Lin H-Y, Allison JJ, et al. Racial disparities in HIV virologic failure: Do missed visits matter? *J Acquir Immune Defic Syndr*. 2009;50(1):100–108.
16. Cheever LW. Engaging HIV-infected patients in care: their lives depend on it. *Clin Infect Dis*. 2007;44(11):1500–1502.
17. Naar-King S, Bradford J, Coleman S, Green-Jones M, Cabral H, Tobias C. Retention in care of persons newly diagnosed with HIV: outcomes of the Outreach Initiative. *AIDS Patient Care STDS*. 2007;21(Suppl 1):S40–48.
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(6):613–619.
19. Menke TJ, Giordano TP, Rabeneck L. Utilization of health care resources by HIV-infected white, African-American, and Hispanic men in the era before highly active antiretroviral therapy. *J Natl Med Assoc*. 2003;95(9):853–861.
20. Branson BM, Handsfield HH, Lampe MA, et al. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. *MMWR*. 2006;55(RR-14).

Evaluation of the Sustainability of an Intervention to Increase HIV Testing

Matthew Bidwell Goetz, MD[1,5], Tuyen Hoang, PhD[2], S. Randal Henry, DPH[2], Herschel Knapp, PhD, MSSW[2], Henry D. Anaya, PhD[2,5], Allen L. Gifford, MD[3,4], and Steven M. Asch, MD, MPH[2,5]and the QUERI-HIV/Hepatitis Program

[1]Infectious Diseases Section (111-F), VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA; [2]General Medicine (111G), VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA; [3]VA Bedford Center for Health Quality, Outcomes & Economic Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, USA; [4]Boston University Schools of Public Health and Medicine, Boston, MA, USA; [5]David Geffen School of Medicine at UCLA, Los Angeles, CA, USA.

BACKGROUND: Sustainability—the routinization and institutionalization of processes that improve the quality of healthcare—is difficult to achieve and not often studied.

OBJECTIVE: To evaluate the sustainability of increased rates of HIV testing after implementation of a multicomponent intervention in two Veterans Health Administration healthcare systems.

DESIGN: Quasi-experimental implementation study in which the effect of transferring responsibility to conduct the provider education component of the intervention from research to operational staff was assessed.

PATIENTS: Persons receiving healthcare between 2005 and 2006 (intervention year) and 2006 and 2007 (sustainability year).

MEASUREMENTS: Monthly HIV testing rate, stratified by frequency of clinic visits.

RESULTS: The monthly adjusted testing rate increased from 2% at baseline to 6% at the end intervention year and then declined reaching 4% at the end of the sustainability year. However, the stratified, visit-specific testing rate for persons newly exposed to the intervention (i.e., having their first through third visits during the study period) increased throughout the intervention and sustainability years. Increases in the proportion of visits by patients who remained untested despite multiple, prior exposures to the intervention accounted for the aggregate attenuation of testing during the sustainability year. Overall, the percentage of patients who received an HIV test in the sustainability year was 11.6%, in the intervention year 11.1%, and in the pre-intervention year 5.0%

CONCLUSIONS: Provider education combined with informatics and organizational support had a sustainable effect on HIV testing rates. The effect was most pronounced during patients' early contacts with the healthcare system.

Received May 21, 2009
Revised August 21, 2009
Accepted September 02, 2009
Published online October 2, 2009

KEY WORDS: HIV testing; provider education; sustainability; VA hospitals.
J Gen Intern Med 24(12):1275–80
DOI: 10.1007/s11606-009-1120-8

INTRODUCTION

There is a growing literature on the types of interventions required to improve healthcare quality[1]. To reap long-term benefits, the gains brought about by such programs must be sustained beyond the initial interventional period. However, achieving sustainability (i.e., the routinization and institutionalization of improved processes), is difficult and may be dependent on characteristics of the intervention that are not examined during the trial that demonstrates effectiveness. Sustainability is not often studied, and when it is, the results are often disappointing[2–4].

Herein we report on the sustainability of a successful intervention to increase HIV testing. The clinical benefits of identifying and treating asymptomatic human immunodeficiency virus (HIV) infected individuals are firmly established and more cost-effective than many other general population preventive services[5–14]. However, 21% of the 1.1 million HIV-infected persons in the United States remain undiagnosed[15]. Similarly, only 30% to 50% of Veterans Administration (VA) patients with known, documented risk factors for HIV infection have been tested[16,17]. Therefore, we previously implemented a multi-modal intervention based upon computerized decision support, provider education and feedback, and organizational changes that significantly increased HIV testing rates in at-risk individuals who receive care at VA medical facilities[18]. Over a one-year period, implementation of this program increased the cumulative rate of ever being tested for HIV from 20.1% to 53.7% ($p<0.001$). In contrast, there was no change in three control facilities.

Once the interventional year was over, we turned project responsibility over to preexisting primary care clinical leadership. This leadership chose to dramatically reduce the labor-intensive provider education campaign and merged what little that remained into routine clinical management (e.g. weekly staff meetings). They did, however, continue the largely "fixed" changes in the systems infrastructure for HIV testing, which

required substantially less support to maintain (i.e. the computerized decision support, feedback reports, and maintenance of organizational changes). We now report on the intervention's sustainability in the second, sustainability year of this project.

METHODS

As previously described[18], the intervention program was put in place for one year in two of the five geographically separate VA regional healthcare systems (HCS) in southern Nevada and California. HCS A and B were comprised of 12 and five subfacilities, respectively, in which primary care were provided by mixtures of academic and non-academic staff physicians, postgraduate medical trainees and mid-level providers. This study was approved by the appropriate institutional review boards.

In brief, the components of the intervention were:

1.) *A continuously updated, electronic clinical reminder* that identifies patients at increased risk for HIV infection and encourages providers to offer HIV testing to such individuals. This reminder is triggered by HIV risk factors available in the VA electronic medical record. These include evidence of Hepatitis B or C infection, illicit drug use, sexually transmitted diseases, homelessness, and Hepatitis C risk factors[18]. Once triggered, the reminder was resolved by ordering an HIV test, recording the result of an HIV test performed elsewhere, or indicating that the patient was not competent to consent to testing or refused HIV testing. Once resolved, the reminder was no longer triggered.

2.) *An audit-feedback system:* Providers were given quarterly reports of clinic-level HIV testing performance[19].

3.) *The reduction of organizational barriers*: Under federal laws specific to the VA, written, informed consent and pre-test HIV counseling have been required for all HIV tests[20]. To expedite this process we encouraged nurse-based rather than physician-based pre-test counseling, use of streamlined HIV counseling, and both telephone notification and brief post-test counseling after negative HIV test results[18,21].

4.) *A provider education (activation) program:* This included academic detailing, social marketing, and educational materials[22,23]. The academic detailing component involved regular informal discussions by project staff to encourage providers to prioritize the performance of HIV testing[24,25]. Social marketing involved having physician and nursing clinical opinion leaders encourage HIV testing by primary care healthcare providers[26]. Finally, we developed and distributed educational hand-outs, pocket cards and posters to promote HIV testing and increase provider comfort and abilities to provide pre- and post-test HIV counseling.

All aspects of the program were implemented in the first month of the intervention year at HCS A and HCS B and maintained during the subsequent 11 months. In support of the provider education program, members of the study team made frequent visits to the clinics to informally promote HIV testing in one-on-one ad hoc meetings with primary care providers. In addition, senior members of the study team regularly attended clinic and facility-wide meetings of primary care physicians, nurses and clinic leadership to promote HIV testing.

The study team did not participate in provider education activities during the second (sustainability) year of the study and instead fully transferred responsibility for this activity to clinic leadership. Qualitative evaluation indicated that provider education activities were much reduced and merged into routine clinical management activities such as staff meetings. Leadership did maintain other aspects of the intervention, including quarterly feedback reports of the rate of HIV testing, and the electronic clinical reminder. Organizational changes that had eased the documentation requirements for HIV testing and broadened the number of people authorized to initiate testing and counseling persisted. Distribution of educational activities, pocket cards and handouts continued at a reduced rate

Our primary analytical goal was to assess the trajectory of the monthly rate of HIV testing during the intervention and sustainability years. In addition, we assessed changes in the proportion of patients who agreed to be tested.

Data sources. We obtained administrative and clinical data, including patient demographics, laboratory tests, diagnostic codes and health factors of the inpatient and outpatient encounters from August 2004 to July 2007 from a pre-existing regional VA database[18]. The medical records were linked across the data files by encrypted identifiers.

Study population. We evaluated outcomes during clinical visits of patients who were identified as being at-risk for HIV infection but had not been offered HIV testing (i.e., the HIV Testing Clinical Reminder had been previously been resolved). Visits by eligible patients were removed from the database subsequent to the month during which the reminder was resolved.

Statistical methods. To assess the adjusted rates of HIV testing and refusal, we performed logistic regression analyses in which the unit of analysis was the patient who was seen at the VHA facilities in each month, had HIV risk factors, but the HIV Testing Clinical Reminder had not previously been resolved. The dependent variables were performance of HIV testing and documentation of patient refusal to be tested. The independent variables included patient demographic and clinical factors such as age, race and ethnicity, marital status, lack of housing, co-payment status, being at-risk for hepatitis C, hepatitis C infection, hepatitis B infection, illicit substance use and sexually transmitted diseases[18]. The two VHA healthcare systems comprised of 17 facilities where the patients were seen. To adjust for any systemic effects on patient likelihood of accepting or refusing HIV testing, we included facility-level annual patient loads and baseline HIV testing rates in the pre-intervention period as independent variables. Finally, we adjusted the covariance of the regression model for patient clustering within facilities using the Generalized Estimating Equation method. The data analysis was generated using SAS v9.1 *proc genmod* (SAS version 9.1. SAS Institute, Cary, NC, USA).

RESULTS

Table 1 compares the demographic features and factors of patients with known risk for HIV infection who received care in the intervention and sustainability years. In the sustainability year, at-risk patients were somewhat younger and less often married. This largely represents an influx of veterans from recent military campaigns into VA care[27–30]. Otherwise there were no meaningful differences in demographic and clinical characteristics between the two years. The number of patients in the sustainability year was lower than in the intervention year as all patients in whom the HIV Testing Clinical Reminder was resolved in the intervention year were excluded from the analyses of the sustainability year.

We previously reported that our multi-modal intervention more than doubled the rate of HIV testing rates among at-risk individuals[18]. The percentage of at-risk patients who received an HIV test was 11.1% in the intervention year versus 5.0% in the year prior to the intervention ($p<0.001$). In the sustainability year, 11.6% of at-risk patients were tested. To better assess whether this result represented actual sustainability of the intervention, we assessed the trajectory of the monthly HIV testing rates[31,32]. This rate increased from 2% at baseline (prior to implementation of the program) to 6% in month 12 (Fig. 1). Although the monthly testing rate declined in the sustainability year, the rate in month 24 remained more than twice the baseline rate (4% versus 2%). These results were consistent across all patient subgroups (data not shown).

As only patients in whom the HIV Testing Clinical Reminder remained unresolved were eligible for testing in the sustainability year, the previous analyses are susceptible to bias from differences in system-, provider- or patient-level characteristics for patients in whom the reminder was or was not resolved in the intervention year. To reduce this bias, we analyzed HIV testing rates by the order of visits since the start of the intervention period (i.e. first visit, second visit, etc.). This analysis was prompted by discussions with providers which indicated that a more comprehensive approach to detecting undiagnosed disease is taken in new patients. As shown on Figure 2a, the HIV testing rate was consistently greatest on a patient's first visit during the study period (i.e., on the first possible exposure to the intervention). For such patients, the testing rate increased from 2% at baseline (pre-intervention) to 6% on month 1; the rate continued to increase throughout the 24-month observation period. For each subsequent visit, the magnitude of the increase in the HIV testing rate was less than for patients having their first visit, but remained greater than during the prestudy period for patients

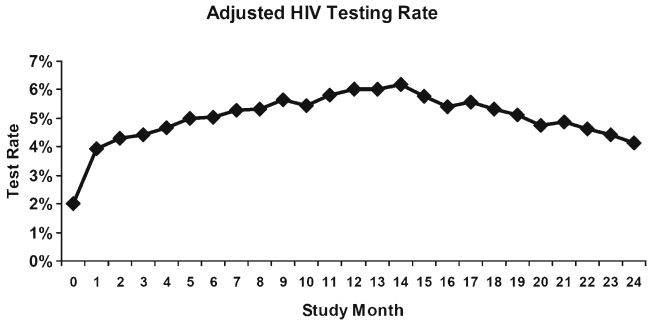


Figure 1. Adjusted HIV testing rates among all patients with identified risk factors for HIV infection. The active intervention period started in study month one and lasted through study month 12. The sustainability period started in study month 13.

Table 1. Patient Demographic and Clinical Characteristics

Characteristics	1st year	2nd year
N	29885	18486
Age (%)		
• 18–30	4.4	7.3
• 31–50	16.5	21.5
• 51–64	42.9	45.1
• 65+	36.2	26.1
Race/ethnicity (%)		
• Caucasian	17.8	16.1
• African American	8.8	9.6
• Hispanic	3.6	3.4
• Asian, Native American	8.0	6.5
• Missing	61.8	64.4
Marital status (%)		
• Single	23.1	27.1
• Married	35.7	31.1
• Widow/divorced/separated	41.3	41.7
Low income (%)	69.8	72.6
Risk factors (%)		
• Hepatitis C infection	18.7	23.4
• Hepatitis B infections	12.5	13.4
• Prior sexually transmitted disease	4.1	5.2
• History of substance abuse	14.7	19.3
• History of homelessness	18.1	22.7
• Presence of risk factors for HCV infection	66.9	56.2

Patients were included if they had identified risk factors for HIV infection, were not known to be HIV-infected and had no documentation of previously having had an HIV test, refusing an HIV test or being incompetent to consent to HIV testing.

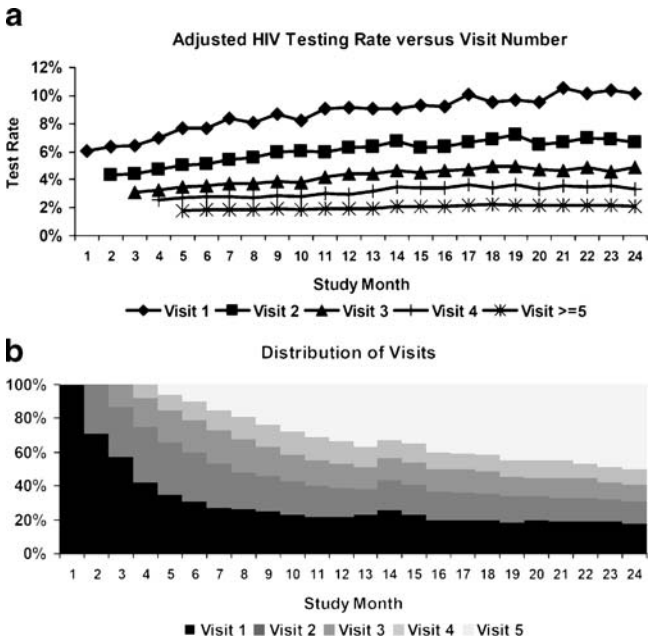


Figure 2. (a). Adjusted HIV testing rates among patients as stratified by outpatient study visit number. The starting period for the strata are offset at monthly intervals as very few patients had more than one visit per month. (b). Proportion of outpatient visits grouped by visit number.

having their second to fourth visits. Time series analyses demonstrated that the probability of being tested increased over time for patients having their second or third visits. Minimal increases were seen on the fourth visit and the testing rate on the fifth and later visits did not increase. Over time the proportion of patients being seen on their first to third visits decreased while the proportion being seen on visit number four and greater increased (Fig. 2b). This change in patient distribution explained the attenuation of the rate of HIV testing in the overall population. Further analyses did not identify any demographic, clinical or facility characteristics that differed between persons who were or were not tested for HIV by their fourth visit (data not shown).

As discussed in METHODS, the HIV Testing Clinical Reminder can be resolved by performing an HIV test or by documenting that the patient refused to be tested. While allowing for patient choice with respect to HIV testing, minimization of the refusal rate is an important goal; once "refused" is selected, the HIV Testing Clinical Reminder did not prompt providers to re-offer HIV testing during future visits. However, we hypothesized that some "refusals" might actually reflect provider discomfort offering an HIV test[33,34], and therefore that the refusal rate might decrease as providers gained more HIV testing experience.

We found that there was a substantial, continuous decrease in the HIV test refusal rate (Fig. 3). The net result was that among persons in whom the HIV Testing Clinical Reminder was resolved, the likelihood that reminder resolution resulted in HIV testing increased from 17% of all reminder responses in the first month of the intervention to 60% in the final month.

DISCUSSION

We previously demonstrated that implementation of an integrated package of quality improvement interventions that utilizes decision support, a provider education (activation) campaign, feedback reports and organizational changes more than doubled HIV testing rates for at-risk individuals[18]. These results were robust with dramatic increases in the likelihood of being tested for HIV being observed across patient-level, provider-level and subfacility-level factors. Furthermore, the fraction of HIV test results that were positive remained constant (0.45%) and well within the range at which HIV testing costs less than $50,000 per quality-adjusted life year when societal benefits of testing are considered[6].

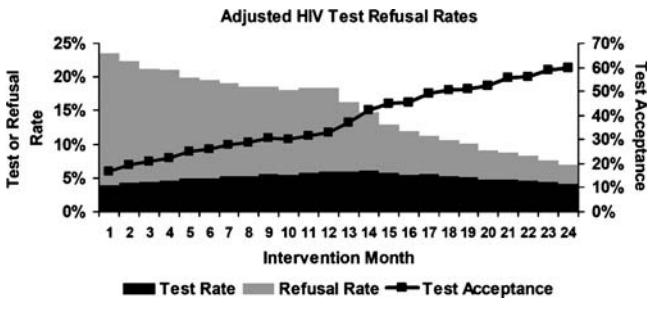


Figure 3. The vertical bars depict the adjusted rates at which patients with HIV risk factors underwent tests or were stated to refuse testing. The lines indicate the proportion of patients who were offered HIV testing and then underwent testing.

We now report on the sustainability of this program during the twelve-month period after overall responsibility for the interventional program was transferred to preexisting clinical management, who chose to greatly deintensify the provider education campaign and other labor and time-intensive aspects of the intervention[18,35]. Remarkably, we found that the rate of HIV testing continued to increase for patients making their first, second or third visits during the sustainability period. These results indicate that despite the de-emphasis of the provider education campaign, when the frequency of medical contact is considered, the program's impact on HIV testing rates was fully sustainable. The observation that overall testing rates declined was related to the changing make-up of the study population as patients with their first through third visits accounted for 100% of the study population in month 1, 54% of the population in month 12 and 41% of the population in month 24.

We also found that the rate at which patients refused HIV testing decreased over time. Correspondingly, the likelihood of having the HIV Testing Clinical Reminder being resolved by HIV testing increased. These results suggest that providers became more proficient at offering and discussing HIV tests and may have integrated HIV testing into their normal practice. Others have observed that normalization of HIV testing is associated with increased patient acceptance of testing[36,37].

The importance of reporting the sustainability of health care interventions and of choosing appropriate measurement metrics is receiving increasing attention[32]. Our results indicate that assessments of the sustainability of the outcome of an intervention are critically dependent on the mode of analysis. We found that when applied to homogeneous patient population (as defined by prior use of VA healthcare), increased HIV testing rates were sustained after de-emphasis of the provider education campaign and continued to increase among patients newly exposed to the intervention (Fig. 2a). This suggests that our intervention has become part of the institutional culture of our facility, does not overburden providers and fits the implementing culture and variations of the patient population[32].

Stratified analysis by the number of visits during each year reveals that our intervention was least sustained among established patients who had not previously been offered testing. We conclude that interventions that aim to maximize sustainability should consider a "tail" of provider education or other components focused on patients who do not receive recommended services on the first exposure. Also, further work needs to be done to determine the determinants of repeated non-performance. We believe that such failures are likely due to systemic barriers or a lack of provider agreement/knowledge. Notably, although theoretical[38–40] and empirical observations[22,23,41–43] demonstrate that the use of provider education (or activation) campaigns are necessary to transform group norms and maximize quality improvement, there is far less literature regarding the importance of maintaining these activities to sustain whatever gains are achieved during their use[32].

The strengths of our sustainability analysis include, as recommended, use of a time-series analysis of monthly rates of HIV testing which allowed us to better assess the trajectory of HIV testing rates[32,44]. Furthermore, we examined the effectiveness of the intervention in an unselected population of at-risk veterans receiving care in a routine, real-world clinical setting.

Limitations include the fact that the sustainability analysis was done immediately after the withdrawal of study personnel from active maintenance of the intervention. It is therefore difficult to distinguish between lingering improvements from the implementation and true persistence of effects from institutionalization[45]. Moreover, this study was undertaken within the quality improvement infrastructure in the VA, which includes an electronic medical record, clinical reminder software and familiarity with performance measurements. Although such tools are increasingly common, this intervention might not be generalizable to other healthcare systems. Another limitation is that while sustainability can be defined as continued use of the core elements of the interventions, and persistence of improved performance[32], we did not formally evaluate the continued use of the core elements of the interventions or their individual contributions to the successful sustenance of the intervention. However, surveys of the two HCSs involved in this project indicate that the organizational changes that favor HIV testing and the HIV Testing Clinical Software package have been maintained. Another limitation is that there was still room for improvement and it is unknown whether the rates of HIV testing would have increased further had the provider activation campaign been continued. Furthermore, while guidelines now recommend that all patients be offered HIV testing and that yearly testing be offered to persons who continue to engage in high risk activities[14,46–48], this intervention was targeted to ensure one-time testing in patients with known risk factors. This strategy was purposely undertaken to prioritize testing for patients at the highest known risk for HIV infection and in deference to concerns that a program to promote HIV testing in all patients would be impractical in the VA as long as written informed consent was required for testing. Finally, the achieved rate of HIV testing remained less than desired. It will be important to determine the effect of removal of the written informed consent requirement for VA HIV testing in August 2009 on the rates of HIV testing[49].

In conclusion, we found that when assessed in homogeneous patient populations, the impact of implementation of the coordinated use of a computerized clinical reminder, feedback reports, provider education and organizational change is sustainable after cessation of external support of the provider education component. Maintenance of the gains after withdrawal of support by the research team suggests that the organizational and behavioral changes that led to the enhanced performance of HIV testing were successfully institutionalized. These findings have substantial implications for the assessment and sustenance of quality improvements programs for clinical preventive services and beyond.

Acknowledgements: *Support: This project was supported by a research grant to Drs. Goetz and Asch by the Health Services Research & Development Service (SDP 06–001)*

Conflicts of Interest: *Matthew Bidwell Goetz: consultancy with Monogram Biosciences; grants in the last 3 years (Gilead Pharmaceuticals), Henry D. Anaya: stock ownership in Trinity Biotechnology, which develops biomarker devices, one of which is a test for the HIV virus, and educational support in the form of unrestricted grants from both Trinity Biotechnology and OraSure Technologies. Allen Gifford: royalties for authorship of Living Well With HIV And AIDS, Ball Publishing Company. Steven Asch: unrestricted travel grant from Trinity Pharmaceuticals. The other authors have no conflicts of interest.*

Corresponding Author: *Matthew Bidwell Goetz, MD; Infectious Diseases Section (111-F), VA Greater Los Angeles Healthcare System, 11301 Wilshire Blvd, Los Angeles, CA 90073, USA (e-mail: matthew.goetz@va.gov; mgoetz@ucla.edu).*

REFERENCES

1. **Committee on Quality of Health Care in America IoM.** Crossing the Quality Chasm. A New Health System for the 21st Century. National Academy Press. Washington, DC; 2001:1–337.
2. **Goodman RM, McLeroy KR, Steckler AB, Hoyle RH.** Development of level of institutionalization scales for health promotion programs. Health Educ Q. 1993;20:161–78.
3. **Johnson K, Hays C, Center H, Daley C.** Building capacity and sustainable prevention innovations: a sustainability planning model. Eval Program Plann. 2004;27:135–49.
4. **Knox KL, Litts DA, Talcott GW, Feig JC, Caine ED.** Risk of suicide and related adverse outcomes after exposure to a suicide prevention programme in the US Air Force: cohort study. BMJ. 2003;327:1376.
5. **DHHS Panel on Antiretroviral Guidelines for Adults and Adolescents.** Guidelines for the Use of Antiretroviral Agents in HIV-infected adults and adolescents. Available at http://aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf. Accessed September 2009.
6. **Sanders GD, Bayoumi AM, Sundaram V, et al.** Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. N Engl J Med. 2005;352:570–85.
7. **Paltiel AD, Weinstein MC, Kimmel AD, et al.** Expanded screening for HIV in the United States–an analysis of cost-effectiveness. N Engl J Med. 2005;352:586–95.
8. **Yeni PG, Hammer SM, Hirsch MS, et al.** Treatment for adult HIV infection: 2004 recommendations of the International AIDS Society-USA Panel. JAMA. 2004;292:251–65.
9. **Goetz MB, Morreale A, Berman S, et al.** Effect of highly active antiretroviral therapy (HAART) on outcomes in Veterans Affairs Medical Centers (VAMC). Clin Infect Dis. 1997;25:396.
10. **Palella FJ, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA.** Declining morbidity and mortality among patients with advance human immunodeficiency virus infection. N Engl J Med. 1998;338:853–60.
11. **Quinn TC, Wawer MJ, Sewankambo N, et al.** Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med. 2000;342:921–9.
12. **Marks G, Crepaz N, Senterfitt JW, Janssen RS.** Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States: implications for HIV prevention programs. J Acquir Immune Defic Syndr. 2005;39:446–53.
13. **Marks G, Crepaz N, Janssen RS.** Estimating sexual transmission of HIV from persons aware and unaware that they are infected with the virus in the USA. AIDS. 2006;20:1447–50.
14. **Paltiel AD, Walensky RP, Schackman BR, et al.** Expanded HIV screening in the United States: effect on clinical outcomes, HIV transmission, and costs. Ann Intern Med. 2006;145:797–806.
15. **Campsmith M, Rhodes P, Hall I.** Estimated Prevalence of Undiagnosed HIV Infection: US, End of 2006. 16th Conference on Retroviruses and Opportunistic Infections. 2–11–2009; Montreal, Canada. Abstract 1036.
16. **Perlin JB.** Need for routine Human Immunodeficiency Virus (HIV) risk assessment and testing. Under Secretary for Health's Information Letter. IL 10–2005–017. Available at http://vaww.vhaco.va.gov/pubarchives/docs/10200517.pdf. Accessed September 2, 2009.
17. **Owens DK, Sundaram V, Lazzeroni LC, et al.** HIV testing of at risk patients in a large integrated health care system. J Gen Intern Med. 2007;22:315–20.
18. **Goetz MB, Hoang T, Bowman C, et al.** A systemwide intervention to improve HIV testing in the veterans health administration. JGIM. 2008;23:1200–7.
19. **Lomas J, Enkin M, Anderson GM, Hannah WJ, Vayda E, Singer J.** Opinion leaders vs. audit and feedback to implement practice guidelines: delivery after previous cesarean section. JAMA. 1991;265:2202–7.
20. **VA Policy on Confidential HIV Testing and Counseling.** Available at http://www.hiv.va.gov/vahiv?page=prtop02-ov-02#S2X. Accessed September 2009.

21. **Anaya HD, Hoang T, Golden JF, et al.** Improving HIV screening and receipt of results by nurse-initiated streamlined counseling and rapid testing. J Gen Intern Med. 2008;23:800–7.
22. **Thomson O'Brien MA, Oxman AD, Haynes RB, Davis DA, Freemantle N, Harvey EL.** Local opinion leaders: effects on professional practice and health care outcomes. Cochrane Database Syst Rev. 2005: CD000125. doi:10.1002/14651858.CD000125.
23. **Jamtvedt G, Young J, Kristoffersen D, O'brien M, Oxman A.** Audit and feedback: effects on professional practice and health care outcomes. Cochrane Database Syst Rev. 2006;19:CD000259.
24. **Collins B, Hawks J, Davis R.** From theory to practice: Identifying authentic opinion leaders to improve care. Managed Care. 2000;9:56–62.
25. **Solomon D, Van Houten L, Glynn R, et al.** Academic detailing to improve use of broad spectrum antibiotics at an academic medical center. Arch Intern Med. 2001;161:1897–902.
26. **Lefebvre RC, Rochlin L.** Social marketing. In: Lewis FM, Rimer BK, eds. Health Behavior and Health Education: Theory, Research, and Practice. 2nd ed. Jossey-Bass; 1988:
27. **Heron MP, Smith BL.** Deaths: leading causes for 2003. National Vital Statistics Report. 2007;55:1–93.
28. **Bray RM, Hourani LL, Olmsted KLR, Witt M, Brown JM, Pemberton MR, Marsden ME, Faulkner DL, Marriott B, Scheffler S, Vandermaas-Peeler R, Weimer B, Calvin S, Bradshaw M, Close K, Hayden D.** 2005 Department of Defense Survey of Health Related Behaviors Among Military Personnel. Available at http://www.ha.osd.mil/special_reports/2005_Health_Behaviors_Survey_1-07.pdf. Accessed September 2009.
29. **Holmes KK, Levine R, Weaver M.** Effectiveness of condoms in preventing sexually transmitted infections. Bull World Health Organ. 2004;82:454–61.
30. **Bryant KJ.** Expanding research on the role of alcohol consumption and related risks in the prevention and treatment of HIV/AIDS. Subst Use Misuse. 2006;41:1465–507.
31. **Shojania KG, Grimshaw JM.** Evidence-based quality improvement: the state of the science. Health Aff (Millwood). 2005;24:138–50.
32. **Bowman CC, Sobo EJ, Asch SM, Gifford AL.** Measuring persistence of implementation: QUERI Series. Implement Sci. 2008;3:21.
33. **Epstein RM, Morse DS, Frankel RM, Frarey L, Anderson K, Beckman HB.** Awkward moments in patient-physician communication about HIV risk. Ann Intern Med. 1998;128:435–42.
34. **Epstein RM, Levenkron JC, Frarey L, Thompson J, Anderson K, Franks P.** Improving physicians' HIV risk-assessment skills using announced and unannounced standardized patients. J Gen Intern Med. 2001;16:176–80.
35. **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: QUERI Series. Implement Sci. 2008;3:16.
36. **Bayer R, Fairchild AL.** Changing the paradigm for HIV testing–the end of exceptionalism. N Engl J Med. 2006;355:647–9.
37. **Sobo EJ, Bowman C, Halloran J, Asch S, Goetz MB, Gifford AL.** "A Routine Thing": clinician strategies for implementing HIV testing for at-risk patients in a busy healthcare center. Anthropol Med. 2008;15:213–5.
38. **Berwick DM.** Developing and testing changes in delivery of care. Ann Intern Med. 1998;128:651–66.
39. **Bodenheimer T, Wagner EH, Grumbach K.** Improving primary care for patients with chronic illness: the chronic care model, Part 2. JAMA. 2002;288:1909–14.
40. **Bodenheimer T, Wagner EH, Grumbach K.** Improving primary care for patients with chronic illness. JAMA. 2002;288:1775–9.
41. **Stone EG, Morton SC, Hulscher ME, et al.** Interventions that increase use of adult immunization and cancer screening services: a meta-analysis. Ann Intern Med. 2002;136:641–51.
42. **Patterson ES, Doebbeling BN, Fung CH, Militello L, Anders S, Asch SM.** Identifying barriers to the effective use of clinical reminders: bootstrapping multiple methods. J Biomed Inform. 2005;38:189–99.
43. **Patterson ES, Nguyen AD, Halloran JP, Asch SM.** Human factors barriers to the effective use of ten HIV clinical reminders. J Am Med Inform Assoc. 2004;11:50–9.
44. **Shojania KG, Grimshaw JM.** Evidence-based quality improvement: the state of the science. Health Aff (Millwood). 2005;24:138–50.
45. **Titler MG.** Methods in translation science. Worldviews Evid Based Nurs. 2004;1:38–48.
46. **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;55:1–17.
47. **Bartlett JG, Branson BM, Fenton K, Hauschild BC, Miller V, Mayer KH.** Opt-out testing for human immunodeficiency virus in the United States: progress and challenges. JAMA. 2008;300:945–51.
48. **Qaseem A, Snow V, Shekelle P, Hopkins R Jr, Owens DK.** Screening for HIV in health care settings: a guidance statement from the American College of Physicians and HIV Medicine Association. Ann Intern Med. 2009;150:125–31.
49. **Anonymous.** Elimination of Requirement for Prior Signature Consent and Pre- and Post-Test Counseling for HIV Testing. Federal Register. 2009;74:34500–3.

Racial and Ethnic Differences in End-of-Life Costs

Why Do Minorities Cost More Than Whites?

Amresh Hanchate, PhD; Andrea C. Kronman, MD, MSc; Yinong Young-Xu, ScD, MS; Arlene S. Ash, PhD; Ezekiel Emanuel, MD, PhD

Background: Racial and ethnic minorities generally receive fewer medical interventions than whites, but racial and ethnic patterns in Medicare expenditures and interventions may be quite different at life's end.

Methods: Based on a random, stratified sample of Medicare decedents (N = 158 780) in 2001, we used regression to relate differences in age, sex, cause of death, total morbidity burden, geography, life-sustaining interventions (eg, ventilators), and hospice to racial and ethnic differences in Medicare expenditures in the last 6 months of life.

Results: In the final 6 months of life, costs for whites average $20 166; blacks, $26 704 (32% more); and Hispanics, $31 702 (57% more). Similar differences exist within sexes, age groups, all causes of death, all sites of death, and within similar geographic areas. Differences in age, sex, cause of death, total morbidity burden, geography, socioeconomic status, and hospice use account for 53% and 63% of the higher costs for blacks and Hispanics, respectively. While whites use hospice most frequently (whites, 26%; blacks, 20%; and Hispanics, 23%), racial and ethnic differences in end-of-life expenditures are affected only minimally. However, fully 85% of the observed higher costs for nonwhites are accounted for after additionally modeling their greater end-of-life use of the intensive care unit and various intensive procedures (such as, gastrostomies, used by 10.5% of blacks, 9.1% of Hispanics, and 4.1% of whites).

Conclusions: At life's end, black and Hispanic decedents have substantially higher costs than whites. More than half of these cost differences are related to geographic, sociodemographic, and morbidity differences. Strikingly greater use of life-sustaining interventions accounts for most of the rest.

Arch Intern Med. 2009;169(5):493-501

Author Affiliations: Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts (Drs Hanchate, Kronman, and Ash); Lown Cardiovascular Research Foundation, Brookline, Massachusetts (Dr Young-Xu); and Department of Bioethics, National Institutes of Health, Bethesda, Maryland (Dr Emanuel).

RACIAL AND ETHNIC DISPARITIES pervade US health care.[1-9] Many studies show blacks and Hispanics receiving fewer medical services and spending less than whites. For example, minorities receive fewer cardiac procedures, prescriptions for life-saving medications, and narcotic medications for pain relief. Despite efforts by policy makers to address these disparities, they persist.[5,10] At the end of life, however, this pattern may be reversed.[11] Several studies have found higher Medicare costs and service use for nonwhites at life's end.[2,12-16] These studies examined differences in sociodemographic and geographic factors as contributors to these disparities. Shugarman et al[14] reported that in the 2 years before the last year of life, spending by blacks was significantly lower. However, in the last year, this deficit "flipped"; estimated final-year spending was 19% higher for blacks than for whites ($P = .10$). They did not study Hispanics.

In analyses restricted to Medicare Part A (hospital bills) and hospital referral regions (HRRs) with substantial numbers of blacks, several studies by researchers at Dartmouth Medical School, Hanover, New Hampsire, have attributed most of black-white cost differences to geography: basically, more blacks live in regions with high hospital use.[2,15] However, even after geographic differences were controlled for, costs in the last 6 months of life remained 29% higher for blacks than for whites.[2]

To better understand the racial differences in end-of-life care and to extend comparisons to Hispanics and other minorities, we constructed a national random sample of nearly 160 000 Medicare decedents, oversampled for nonwhites. We tallied all Medicare costs (Parts A and B) in the final 6 months of life and quantified the effects on racial and ethnic cost differences of a range of factors, including age, sex, preexisting comorbidities (from the penultimate 6 months), cause of death, geography, and socioeconomic indicators, as well as markers for conservative or aggressive use of end-of-life interventions (eg, hospice, intensive care unit [ICU], or ventilator use).

METHODS

DATA SOURCE

Among the 1.76 million Medicare beneficiaries aged 66 or older who died in 2001, we selected 241 655, including random samples of 85 000 each for blacks and whites and for all Hispanic (approximately 30 000) and other minority (approximately 42 000) decedents. To ensure complete health and health care records, we required enrollment in "traditional" fee-for-service Medicare, for which services were individually billed throughout 2000, with both Medicare Parts A and B entitlement (for inpatient and ambulatory care) continuously for 12 months preceding death, as well as a positive match in the National Death Index. We excluded beneficiaries in the end-stage renal disease program and those residing in Puerto Rico or other nonmainland territories, because care for these groups is administered quite differently than for others in Medicare. After these exclusions, there were 158 780 decedents in the analytic sample. The proportions excluded among Hispanics (55%) and others (47%) were much higher than among whites (24%) and blacks (32%), principally owing to the exclusion of residents of Puerto Rico, to the high rates of non–National Death Index match for Hispanic decedents, and to the higher rates of Part B coverage among whites.

OUTCOMES

Our primary outcome is total Medicare-covered health care expenditures at the end of life: specifically, in the 180 days (6 months) preceding death. For each beneficiary, we added Medicare-allowed payments for all covered services, including hospital and skilled nursing facility care, hospice and home health services, physician services, and durable medical equipment purchases. We also examined intermediate outcomes, such as total costs by type of service, and any use of hospice or selected life-sustaining procedures, such as ventilators.[11]

OTHER VARIABLES

Covariates included age, sex, race, total morbidity burden in the penultimate 6 months of life, geographic location, and socioeconomic status, as well as direct concurrent contributors to, or proximate causes of, final 6-month costs, such as cause of death, hospice use, and receipt of specific, intensive life-sustaining procedures. Age, sex, and race were obtained from Medicare's denominator file, using its racial/ethnic categories of white, black, and Hispanic, grouping everyone else into "other," which included Asian (37%), North American Natives (8%), other (32%), and unknown (33%). Socioeconomic status was proxied by (1) median income of the patient's zip code of residence, and (2) whether the patient received Medicaid assistance (buy-in) to pay Medicare Part B premiums. Morbidity burden was summarized using (1) the Charlson comorbidity score and (2) a Diagnostic Cost Group prospective relative risk score (DxCG version 6.1 for SAS Windows). Each score was based on *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*), diagnoses recorded during the 6 months preceding the last 6 months of life.[17,18] The Charlson score assigns points to 19 disease conditions; scores in these data range from 0 to 37. The DxCG software organizes all *ICD-9-CM* codes into 184 "condition categories" and summarizes their expected impact on future expenditures via a relative risk score (RRS); an RRS of 1.00 refers to average expected next-year expenditures among all Medicare beneficiaries (not just decedents) observed for 1 year. Because the choice of morbidity measure did not affect estimates of racial/ethnic differences, we report analyses using only the more predictive RRS.

To see the effect of aggressive end-of-life care on differences in expenditures, we used surgical procedure codes and other markers in the Medicare inpatient file during the last 6 months to identify decedents with any (nonpsychiatric) ICU admission, and each of 10 intensive life-sustaining interventions: cardiac catheterization, implantation of a cardiac assistance device, pulmonary artery wedge monitoring, cardiopulmonary resuscitation or cardiac conversion, gastrostomy, blood transfusions, dialysis, and use of mechanical ventilators, intravenous antibiotics, and cancer chemotherapies.[13] We identified these procedures using the clinical classifications software (CCS) of the Agency for Healthcare Research and Quality, excluding carotid sinus stimulation (code 99.64); finally, we identified chemotherapy use from diagnoses and procedures, as described elsewhere.[19]

We examined differences in expected costs by where decedents had lived, in 3 ways. First, we classified the county of residence descriptively, using Beale rural-urban continuum codes to distinguish "metropolitan counties by size and nonmetropolitan counties by degree of urbanization and proximity to metropolitan areas."[20] We also mapped the zip code of residence into Dartmouth Atlas–based HRRs and hospital service areas (HSAs). The 306 HRRs are aggregations of more than 3000 HSAs that distinguish geographic regions whose residents access the same hospital(s) and physicians.[21] We rely principally on models that use HSA (as so-called fixed effects indicators), because each HSA defines a small geographic cluster of persons who typically rely on the same hospital systems. In a sensitivity analysis, we quantify the modest differences associated with coding place of residence by county, HRR, or HSA.

ANALYSIS

We used STATA version 9.2 for all analyses, which are weighted to adjust for oversampling of nonwhites.[22] We summarized outcome and covariate variables for all Medicare decedents and compared them by race and ethnic group (using χ^2 and analysis of variance to test for differences). We used regression to estimate differences by race/ethnicity in total end-of-life expenditures after accounting for differences in covariates. Models successively added covariate sets: age and sex only (model B); underlying cause of death determined from death certificates and morbidity (using RRSs) in the 6 months before the final 6 months (model C); geography (using HSAs, model D); socioeconomic status (model E); hospice use (model F); and use of the ICU and 10 intensive procedures during the end-of-life period itself (model G). The order in which explanatory variables are added affects the size of the contribution attributed to each. We first adjusted for pure patient characteristics (age, sex, and medical problems) and then for geographic and socioeconomic indicators. We included geography early in the sequence to focus on differences by race among (otherwise similar) persons who face the same health care delivery systems.[23] Finally, we examined differences in the use of specific services only after all these factors that might otherwise confound associations between race and procedure use had been accounted for. For example, to the extent that nonwhites use more aggressive interventions because they have the same utilization patterns as their white neighbors, we wanted that to be attributed to geography. By entering utilization variables last, the estimated effects focus exclusively on cost differences that arise because nonwhites have different utilization patterns than their neighbors. We also explored the effect of sequencing on the apparent importance of different sets of variables.

Table 1. Characteristics of 2001 Medicare Decedents Within Racial/Ethnic Groups[a]

Characteristic	Whites (n=64 819)	Blacks (n=58 182)	Hispanics (n=13 634)	Other Minorities (n=22 145)	All (N=158 780)
Women, age, y, %					
66-74	9	13	12	10	10
75-84	21	21	23	17	21
≥85	26	22	18	29	26
Men, age, y, %					
66-74	11	16	16	12	12
75-84	19	18	23	17	19
≥85	13	10	9	15	12
Age, mean (SD), y	81.8 (8.0)	80.2 (8.4)	79.7 (7.3)	82.3 (8.5)	81.7 (8.0)
DCG prospective risk score >3, %[b]	14	18	21	16	14
DCG score, mean (SD)	1.7 (1.3)	1.7 (1.3)	1.8 (1.5)	1.8 (1.6)	1.8 (1.5)
Cause of death, %					
Heart disease	36	37	36	34	36
Cancer	22	23	20	21	22
Stroke and brain diseases	13	11	11	13	13
Chronic obstructive pulmonary disease	6	4	4	5	6
Pneumonia and influenza	3	4	6	3	3
Diabetes	3	3	4	4	3
Chronic liver disease and cirrhosis	0.5	0.3	1.7	0.6	0.5
Injuries, homicide, or suicide	2	2	2	3	2
Other	15	17	15	16	15
Site of death, %					
Hospital	38	46	52	45	39
Nursing home	31	21	17	25	31
Residence	21	19	20	19	21
Other	10	14	11	11	10
Urbanicity, county of residence, %[c]					
Metropolitan >1 million	40	54	57	52	42
Metropolitan 250 000 to 1 million	21	17	23	21	21
Town <250 000	12	9	9	7	12
Nonmetropolitan, including rural	26	19	11	18	26
Median income quartile, residence zip code, %					
Lowest	10	38	33	15	12
Second lowest	26	27	26	21	26
Third lowest	30	21	23	26	30
Highest	34	14	18	39	33
Medicaid buy-in, %					
No	80	51	32	52	77
Yes	20	49	68	48	23
Hospice use, %					
No	74	80	77	80	74
Yes	26	20	23	20	26

[a] The All column lists the rates for all Medicare decedents (in 2001), obtained by adjusting for the stratified sampling. The similarity of each measure across racial and ethnic cohorts was rejected ($P<.05$).

[b] A measure of total morbidity burden based on diagnoses in the billing data during the 6 months that precede the last 6 months of life and quantified using the Diagnostic Cost Group (DCG) prospective relative risk score.

[c] As indicated by the Beale urbanization score applied to county of residence in 2001.

Because of large sample sizes, racial/ethnic differences are almost always statistically significant at the $P<.05$ level. Therefore, we only explicitly remark on it when race and ethnic differences are not significant.

RESULTS

Among the 158 780 decedents, blacks and Hispanics were younger than whites and more often lived in metropolitan areas with a population of more than 1 million (**Table 1**). The cause of death was similar across racial groups. However, the site of death differed, with blacks and Hispanics dying more often than whites in hospitals and less often in nursing homes.

DO EXPENDITURES DIFFER BY RACE AND ETHNICITY?

Black and Hispanic decedents have significantly higher end-of-life Medicare Parts A and B expenditures than whites (**Table 2**). White decedents average $20 166 in the last 6 months of life; blacks, $26 704 (32% more); and Hispanics, $31 702 (57% more). These racial disparities persist within strata defined by age, sex, morbidity-burden level, and cause and site of death. End-of-life expenditures for blacks are significantly higher than for whites in almost every state (**Figure 1**) and in cities as well as rural areas (Table 2). Interestingly, several southeastern states have both the lowest overall spending and

Table 2. Medicare Expenditures in the Last 6 Months of Life by Decedent Characteristics and Racial/Ethnic Group[a]

Characteristic	Mean Expenses, ×$1000				
	Whites	Blacks	Hispanics	Other Minorities	All
All	20.2	26.7	31.7	25.5	20.7
Women, age, y					
66-74	26.7	32.9	38.0	32.8	27.3
75-84	21.8	28.5	34.9	28.6	22.3
≥85	14.3	22.5	25.0	17.5	14.8
Men, age, y					
66-74	24.4	27.5	33.2	31.6	24.9
75-84	22.6	26.3	31.1	30.6	23.0
≥85	17.2	23.5	27.4	22.0	17.6
Diagnostic Cost Group score category					
<1.0	17.2	20.5	26.1	21.0	17.5
1.0-3.0	19.6	26.8	30.9	24.1	20.1
>3.0	29.7	39.9	42.7	39.6	30.7
Cause of death					
Heart disease	17.9	23.2	29.0	22.7	18.3
Cancer	23.8	27.5	32.3	29.3	24.2
Stroke and brain diseases	14.7	24.1	26.0	20.5	15.3
Chronic obstructive pulmonary disease	23.5	32.7	38.7	30.2	23.9
Pneumonia and influenza	20.2	27.9	32.4	27.7	21.3
Diabetes	22.3	35.9	38.9	33.1	23.3
Chronic liver disease and cirrhosis	23.9	25.5	28.2	28.0	24.2
Injuries, homicide, or suicide	18.5	24.5	22.3	22.0	18.9
Other	23.1	32.0	39.0	27.1	23.8
Site of death					
Hospital	28.0	34.8	40.4	35.7	28.7
Nursing home	15.8	24.6	27.8	18.7	16.2
Residence	14.9	15.7	19.1	15.4	15.0
Other	14.8	17.8	19.9	16.5	15.0
Urbanicity, county of residence					
Metropolitan >1 million	23.8	31.7	36.7	31.7	24.7
Metropolitan 250 000 to 1 million	18.4	21.9	28.4	20.0	18.7
Town <250 000	17.9	21.1	22.6	18.0	18.0
Nonmetropolitan, including rural	17.0	19.8	20.4	17.1	17.1
Income quartile, residence zip code median income					
Lowest	20.4	27.9	34.1	27.0	22.0
Second lowest	18.9	25.2	28.7	23.1	19.4
Third lowest	19.2	25.7	32.1	24.9	19.6
Highest	21.9	27.9	31.1	26.5	22.2
Medicaid buy-in					
No	20.7	27.0	27.7	22.1	21.0
Yes	17.9	26.4	33.6	29.2	19.7
Hospice use					
No	19.8	26.9	32.9	26.0	20.5
Yes	21.1	25.9	27.7	23.3	21.3

[a]The All column lists the rates for all Medicare decedents (in 2001), obtained by adjusting for stratified sampling. The similarity of each measure across racial and ethnic cohorts was rejected ($P<.05$).

the smallest black-white differences in cost; cost disparities were largest in urban areas, where more minorities live and end-of-life costs are also high for whites. Specifically, the average end-of-life expenditures in the largest metropolitan areas are $24 700 for all races; in areas with population numbering 250 000 to 1 million, they are $18 700; and in rural areas, they are $17 100. Finally, these racial and ethnic cost differences occur in each of the last 6 months of life (**Figure 2**).

HOW MUCH OF THE RACIAL AND ETHNIC COST DISPARITY IS ACCOUNTED FOR BY DIFFERENCES IN DEMOGRAPHICS, MORBIDITY, GEOGRAPHY, AND SOCIOECONOMIC FACTORS?

In **Table 3**, we quantify the importance of various factors on these raw cost differences by sequentially controlling first for demographic, health, and geographic variables. Model A shows the raw differences in costs for the

last 6 months of life among the racial and ethnic groups. Blacks' costs are $6538 higher than whites' costs; Hispanics' costs are $11 536 higher, and other minorities' costs are $5307 higher.

Model B controls for age and sex. These 2 factors reduce the cost differences between blacks and Hispanics and whites by about 10%. Model C also controls for morbidity burden and cause of death and reduces these differences by another 7% to 9%, with nearly all of the reduction attributable to morbidity burden. Using the Charlson index instead of the DCG morbidity measure produced very similar reductions.

Additionally accounting for place of residence by including HSAs in model D brings unexplained extra costs down to only $2924 for blacks (together eliminating 55% of the raw difference for blacks), $3705 (eliminating 68% of the raw difference) for Hispanics, and $2103 (eliminating 60%) for decedents of other races. Using either HRR or county as the geographic unit instead of HSA in model D yields similar findings. Adding socioeconomic indicators (zip code, median income, and Medicaid buy-in) yields very modest further reductions in cost differences by race and ethnicity (model E). In total, between 55% and 68% of the raw differences in end-of-life expenditures between whites and the other 3 groups are accounted for by differences in demographics, morbidity, geography, and socioeconomic indicators, with geography contributing the largest part.

HOW MUCH OF THE REMAINING COST DISPARITY IS ACCOUNTED FOR BY SPECIFIC END-OF-LIFE INTERVENTIONS?

Most of the remaining racial and ethnic differences in end-of-life costs can be attributed to differences in the use of hospital-based, life-sustaining interventions (**Table 4**). Blacks and Hispanics are significantly more likely to be admitted to the ICU (32.5% for blacks, 39.6% for Hispanics, and 27.0% for whites). Minorities also receive significantly more intensive procedures, such as resuscitation and cardiac conversion (4.4% of blacks, 4.0% of Hispanics, and 2.7% of whites), mechanical ventilation (18.0% blacks, 21.0% Hispanics, and 11.6% whites), and gastrostomy for artificial nutrition (10.5% blacks, 9.1% Hispanics, and 4.1% whites). In contrast, whites are slightly more likely to receive inpatient cancer chemotherapy (7.9%) than either blacks (7.6%) or Hispanics (7.4%) and are more likely than blacks, but not Hispanics, to receive cardiac catheterization, cardiac balloon assistance devices, and pulmonary artery pressure measurements.[5,9]

Hospice is used more frequently by whites (26%) than blacks (20%) or Hispanics (23%) and is associated with an average reduction in end-of-life expenditures of $784 per beneficiary overall (model F, Table 3). However, differential hospice use has essentially no effect on racial and ethnic differences in end-of-life costs (model F).

Finally, controlling for the use of all 10 life-sustaining interventions, such as ICU admissions, ventilators, and gastrostomies (model G), eliminates more than half of the remaining differences between whites and each of the other groups (Table 3). Of the original $6538 excess cost for black over white decedents, only $997 (15%) remains after this final adjustment, while for Hispanics, only $1902 (16%) of the original $11 536 in excess costs over whites remains "unexplained." Most of the life-sustaining interventions are associated with strikingly higher total costs. For example, among otherwise similar persons, those who use the ICU cost $12 000 more than those who do not; the use of gastrostomy adds $22 800, and mechanical ventilation adds $15 200.

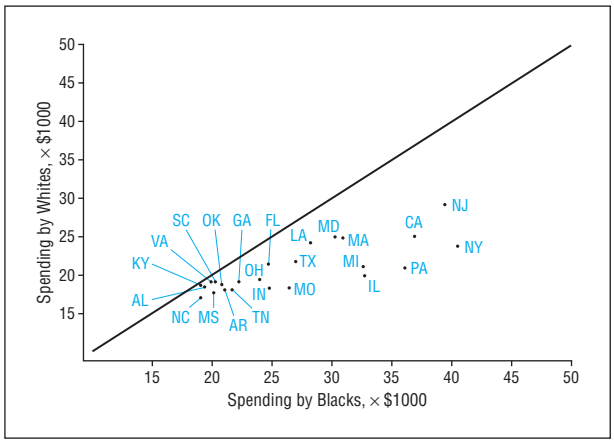


Figure 1. Comparison of black and white average Medicare expenditures 6 months before death, by state. Only the 24 states with at least 400 sample sizes for each racial cohort are plotted.

COMMENT

This study of nearly 160 000 Medicare decedents and their total Medicare (Parts A and B) costs in the last 6 months of life shows substantial differences by race and ethnicity. While there are differences in the magnitude of their cost differences with whites, all 3 groups of nonwhite decedents are similar to whites in their causes of death, but they incur substantially higher Medicare expenditures. In each group, before differences in specific services used are considered, the largest part (between 38% and 58% across the 3 nonwhite groups) of the total excess cost over whites is explained by geography. After all other factors are controlled for, the use of aggressive end-of-life interventions, such as ICU care, ventilators, and gastrostomies, accounts for between 21% and 33% of the difference in total end-of-life costs. Differences in hospice use contribute little to racial and ethnic differences in total end-of-life costs, both because estimated savings from hospice are small (less than $800) and because differences in use by race are modest (20% for blacks, 23% for Hispanics, and 26% for whites).

Using a large and nationally representative sample that linked Medicare data with cause-of-death data from the National Death Index, this study confirms and extends findings regarding black-white differences in costs at the end of life. Raw costs of health care in the last 6 months of life are 32% higher for black Medicare beneficiaries than those for white decedents. Our findings contrast with those of numerous non–end-of-life studies in which minorities received fewer services and especially fewer technologically intensive interventions.[1-10] Medicare expen-

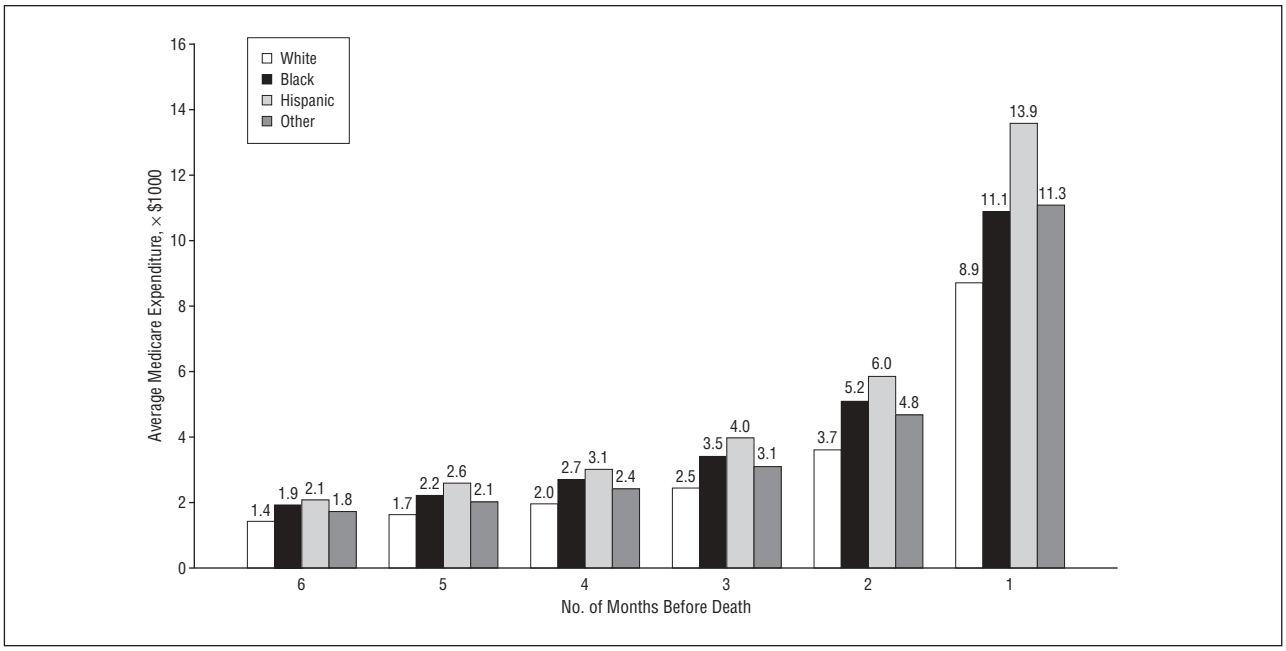


Figure 2. Thirty-day Medicare expenditures for 2001 decedents, by months before death.

ditures for Hispanics were even higher: fully 57% higher than for whites.

Because racial and ethnic differences in cause of death are minimal, they do not contribute to racial differences in end-of-life costs. However, geography is very important, whether measured as HRR, HSA, or county of residence. Blacks and Hispanics are far more likely than whites to live in large urban areas, where medical care in general, and end-of-life care in particular, is more expensive than in smaller cities and rural areas. In part, Medicare costs for black and Hispanic decedents are higher because more of them live and die in higher-cost locations. However, even within the same geographic locations, black and Hispanic decedents have notably higher end-of-life Medicare costs than their white neighbors. In contrast to previous studies, our method of using indicators for each geographic unit instead of area-level measures, such as county hospital beds and physician supply, adjusts not only for measured geographical factors related to variations in practice patterns but also for unknown, and therefore unmeasured, factors.[2,23]

Despite the cumulative importance of age, sex, cause of death, geography, morbidity burden, and socioeconomic status on decedent costs, 45% of the excess costs for blacks and 32% of the excess costs for Hispanics are not explained by these factors. Most of this residual difference is accounted for by more end-of-life ICU admissions and life-sustaining interventions for nonwhites. Black Medicare decedents are significantly more likely to receive resuscitation, mechanical ventilation, and gastrostomy for artificial feedings than are white decedents, even when they reside in the same HSA or county.[11] Hispanics are even more likely than blacks to receive ICU care, mechanical ventilation, dialysis, and cardiac catheterization.

While black, Hispanic, and other minority decedents receive more intensive life-sustaining interventions at the end of life, blacks receive less cancer chemotherapy, cardiac catheterization, and other aggressive cardiac interventions, such as balloon assistance devices, than whites. The lower level of cardiac and oncologic interventions for blacks may be because such interventions require access to subspecialists—oncologists and cardiologists—with whom blacks may have fewer prior relationships. It is not clear why the same is not true for Hispanics. Indeed, why blacks, Hispanics, and other minorities receive so much more of many intensive, life-sustaining interventions now emerges as an important area for further research.

Differences in the use of aggressive end-of-life interventions may reflect patient preferences.[24-26] Some studies have found minorities to be (1) more reluctant than whites to have do-not-resuscitate orders, (2) more likely to prefer life-sustaining treatments at the end-of-life, and (3) less likely to use hospice.[27-35] Such differences in preferences and use of high-technology interventions at the end of life are in contrast with other life stages, in which whites get more intensive interventions.[10] Even if such racial preferences are real, they are not a "first cause"; they raise important policy issues.

Are health care resources for nonwhites misallocated over a lifetime, with racial and ethnic minorities receiving fewer life-extending and life-enhancing interventions than whites throughout their lives[1-10] but more at the end, when there is less opportunity to improve the quantity and quality of life? Perhaps the use of aggressive, hospital-based interventions at the end of life is a well-considered preference. However, even if such interventions are a choice, the decision to use them may stem less from settled views than from distrust of the medical care system or from economic constraints.[10] Nonwhites who receive timely, effective care throughout their lives may find it easier to reject cardiac resuscitation, mechanical ventilation, and artificial nutrition at the end.[36,37] We also know that blacks receive lower-quality primary

Table 3. Models Examining Racial/Ethnic Differences in Total Medicare Expenditures Per Capita in the Last 6 Months of Life[a]

Variable	Model A, Race Only	Model B, A + Age and Sex	Model C, B + Cause of Death + Illness Burden	Model D, C + HSA Effect	Model E, D + SES	Model F, E + Hospice	Model G, F + Medical Interventions
Reference, constant	20 166	14 462	4901	6178	6243	6455	2488
Race							
White	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
Black	6538	5840	5395	2924	3068	3031	997
Hispanic	11 536	10 504	9524	3705	4282	4278	1902
Other	5307	5613	5203	2103	2477	2454	717
Women, age, y							
66-74		12 258	11 427	11 777	11 640	11 607	4156
75-84		7438	6922	6973	6839	6813	2075
≥85		1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
Men, age, y							
66-74		9690	9358	9865	9543	9478	2716
75-84		7959	7115	7538	7177	7131	1691
≥85		2783	1687	2022	1717	1695	–231*
Cause of death							
Stroke and brain diseases			1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
Heart disease			1915	1051	941	856	110*
Cancer, all			5333	4790	4555	4797	4178
Chronic obstructive pulmonary disease			5586	5345	5294	5278	4135
Pneumonia and influenza			1048	1203	1182	1110	1145
Diabetes			7891	6937	6943	6814	3494
Chronic liver disease and cirrhosis			4021	3541	3443	3531	2683
Unintentional injuries, homicide, suicide, etc			3371	3489	3310	3176	1585
Other			7503	7292	7269	7213	4264
DCG risk score			3731	3327	3393	3410	2907
Geography, HSA							
Income quartile, residence zip code median income							
Lowest					1851	1831	772
Second lowest					997	982	251*
Third lowest					171*	162*	–287*
Highest					1 [Reference]	1 [Reference]	1 [Reference]
Medicaid buy-in					–1963	–1984	–616
Enrolled for hospice in last 6 mo						–784	3470
End-of-life interventions							
ICU							12 094
Resuscitation, cardiac conversion							–5501
Ventilation							15 208
Gastrostomy							22 827
Vascular transfusion							11 023
Dialysis							13 072
Chemotherapy							7039
Cardiac catheterization							20 377
Cadiac assistance device							7470
PAP, wedge							2927
Antibiotic							12 360

Abbreviations: DCG, Diagnostic Cost Group; HSA, hospital service area; ICU, intensive care unit; PAP, pulmonary artery pressure; SES, socioeconomic status.
[a] All coefficients are significant at $P < .05$ except values with an asterisk.

care[38,39] and fewer preventive services than whites.[38-40] Perhaps not having a usual source of care and an established relationship with a physician does not allow for an expression of preferences for less intensive treatments at the end of life. Indeed, we found in this Medicare population that more primary care visits just before the last 6 months of life are associated with lower costs and less hospital use at the end of life.[41]

This study has limitations. First, the Medicare claims files contain few clinical descriptors. The findings may not generalize to Medicare managed care, to patients without Medicare coverage, or to patients who do not self-identify as black or Hispanic. Previous work has indicated that sensitivity of Medicare data in identifying minorities is good for blacks but less so for Hispanics.[42] But specificity (the proportion of identified minorities correctly classified) is high for both. The Hispanic effect that we found on end-of-life costs pertains directly only to mainland Hispanics identified as such in Medicare's database. Percentage reductions in racial and ethnic expenditure differences attributed to individual covariates are only rough guides to their relative importance, especially because how a variable affects a model depends on what other factors have previously been accounted

Table 4. Use of Life-Sustaining Interventions in the Last 6 Months of Life by Racial/Ethnic Group[a]

	% of Column With Any Such Use at the End of Life				
	Whites	Blacks	Hispanics	Other Minorities	All[b]
Intensive care unit	27.0	32.5	39.6	30.6	27.5
Resuscitation and cardiac conversion	2.7	4.4	4.0	3.6	2.8
Ventilation	11.6	18.0	21.0	16.6	12.1
Gastrostomy	4.1	10.5	9.1	6.6	4.5
Vascular transfusion	10.4	16.6	18.0	13.7	10.9
Dialysis	1.2	2.7	3.4	2.9	1.3
Chemotherapy	7.9	7.6	7.4	6.8	7.9
Cardiac catheterization	3.4	2.6	4.6	3.0	3.4
Cardiac assistance device	0.6	0.3	0.8	0.6*	0.5
Pulmonary artery pressure and wedge	1.3	1.1	1.7	1.5*	1.3
Antibiotic	0.9	1.2	1.6	1.9	1.0
Hospice enrollment, last 6 mo	25.9	20.1	22.9	20.3	25.5

[a] All white-nonwhite differences in rates of use, except values with an asterisk, are statistically significant ($P < .05$).
[b] The All column lists the rates for all Medicare decedents (in 2001), obtained by adjusting for stratified sampling.

for. Nonetheless, the primacy of "geography" and "aggressive interventions" in accounting for differences in end-of-life costs is a robust finding. Medicare expenditures do not capture all health care expenditures, especially pharmaceuticals and long-term nursing home costs, which may displace some hospital care. Medicare data also do not capture the resources expended by private or government supplemental insurers or financial or in-kind support from families.[43] Importantly, we had no direct data on patient preferences for the various interventions at the end of life. While the models control for most other variables known to influence medical service use—age, sex, race, geography, cause of death, and morbidity burden—patient preference, overt or covert racism in how the same providers treat patients, and differences in which providers and systems of care are accessed might all contribute to these differences in health care system use at the end of life.

In conclusion, blacks and Hispanics die of similar causes, but the costs involved in their last 6 months of life are substantially higher than those of whites. Although 40% to 60% of these excess differences are associated with geography, ie, living in high–medical-expenditure areas, substantial differences remain, even after adjustment for many patient characteristics in addition to geographic variables. Strikingly higher rates of use of intensive end-of-life treatments such as ICU and ventilators account for most of these residual differences. Therefore, at life's end, minorities often receive more expensive but not necessarily life-enhancing care. It is unclear how much of this was actively sought, or the extent to which racial and ethnic differences are principally driven by how choices are presented or how they are "heard." These would be fruitful questions for future research.

Accepted for Publication: July 21, 2008.
Correspondence: Ezekiel Emanuel, MD, PhD, Department of Clinical Bioethics, Bldg 10, Room 1C118, The Clinical Center, National Institutes of Health, Bethesda, MD 20892-1156 (eemanuel@nih.gov).

Author Contributions: Dr Hanchate had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Hanchate, Young-Xu, Ash, and Emanuel. *Acquisition of data:* Hanchate, Young-Xu, and Ash. *Analysis and interpretation of data:* Hanchate, Kronman, Young-Xu, Ash, and Emanuel. *Drafting of the manuscript:* Hanchate, Kronman, and Ash. *Critical revision of the manuscript for important intellectual content:* Hanchate, Kronman, Young-Xu, Ash, and Emanuel. *Statistical analysis:* Hanchate, Young-Xu, and Ash. *Obtained funding:* Emanuel. *Administrative, technical, and material support:* Hanchate, Kronman, Young-Xu, and Ash. *Study supervision:* Hanchate and Ash. *Arranged for use of Centers for Medicare & Medicaid Services data:* Ash and Emanuel.
Financial Disclosure: Dr Ash is cofounder of, and consulting senior scientist at, DxCG Inc, a company whose software was used to measure morbidity burden in this study.
Funding/Support: This study was supported by contracts with the Department of Bioethics of the Clinical Center of the National Institutes of Health.
Role of the Sponsor: The funding agency 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, review, or approval of the manuscript.
Disclaimer: The ideas and opinions expressed are the authors' views. They do not represent any official position or policy of the National Institutes of Health, the Public Health Service, or the Department of Health and Human Services.

REFERENCES

1. Agency for Healthcare Research and Quality. *National Healthcare Disparities Report 2005.* Rockville, MD: Agency for Healthcare Research and Quality; 2006.
2. Baicker K, Chandra A, Skinner JS, Wennberg JE. Who you are and where you live: how race and geography affect the treatment of Medicare beneficiaries. *Health Aff (Millwood).* 2004(suppl Web exclusives):VAR33-VAR44.

3. Barnato AE, Lucas FL, Staiger D, Wennberg DE, Chandra A. Hospital-level racial disparities in acute myocardial infarction treatment and outcomes. *Med Care*. 2005;43(4):308-319.
4. Degenholtz HB, Thomas SB, Miller MJ. Race and the intensive care unit: disparities and preferences for end-of-life care. *Critical Care Med*. 2003;31(5)(suppl) :S373-S378.
5. Jha AK, Fisher ES, Li Z, Orav EJ, Epstein AM. Racial trends in the use of major procedures among the elderly. *N Engl J Med*. 2005;353(7):683-691.
6. Laditka JN, Laditka SB, Mastanduno MP. Hospital utilization for ambulatory care sensitive conditions: health outcome disparities associated with race and ethnicity. *Soc Sci Med*. 2003;57(8):1429-1441.
7. Rathore SS, Krumholz HM. Differences, disparities, and biases: clarifying racial variations in health care use. *Ann Intern Med*. 2004;141(8):635-638.
8. Trivedi AN, Zaslavsky AM, Schneider EC, Ayanian JZ. Trends in the quality of care and racial disparities in Medicare managed care. *N Engl J Med*. 2005; 353(7):692-700.
9. Vaccarino V, Rathore SS, Wenger NK, et al. Sex and racial differences in the management of acute myocardial infarction, 1994 through 2002. *N Engl J Med*. 2005; 353(7):671-682.
10. Epstein AM, Ayanian JZ. Racial disparities in medical care. *N Engl J Med*. 2001; 344(19):1471-1473.
11. Barnato AE, Chang CC, Saynina O, Garber AM. Influence of race on inpatient treatment intensity at the end of life. *J Gen Intern Med*. 2007;22(3):338-345.
12. Hogan C, Lunney J, Gabel J, Lynn J. Medicare beneficiaries' costs of care in the last year of life. *Health Aff (Milwood)*. 2001;20(4):188-195.
13. Levinsky NG, Yu W, Ash A, et al. Influence of age on Medicare expenditures and medical care in the last year of life. *JAMA*. 2001;286(11):1349-1355.
14. Shugarman LR, Campbell DE, Bird CE, Gabel J, Louis TA, Lynn J. Differences in Medicare expenditures during the last 3 years of life. *J Gen Intern Med*. 2004; 19(2):127-135.
15. Dartmouth Institute for Health Policy and Clinical Practice. *Tracking the Care of Patients With Severe Chronic Illness: The Dartmouth Atlas of Health Care 2008*. Lebanon, NH: Dartmouth Institute for Health Policy & Clinical Practice; 2008.
16. Wennberg JE, Fisher ES, Stukel TA, Skinner JS, Sharp SM, Bronner KK. Use of hospitals, physician visits, and hospice care during last six months of life among cohorts loyal to highly respected hospitals in the United States. *BMJ*. 2004; 328(7440):607.
17. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.
18. *DxCG Risk Adjustment Software: User's Guide Release 6.1*. Boston, MA: DxCG; 2004.
19. Emanuel EJ, Young-Xu Y, Levinsky NG, Gazelle G, Saynina O, Ash AS. Chemotherapy use among Medicare beneficiaries at the end of life. *Ann Intern Med*. 2003; 138(8):639-643.
20. US Department of Agriculture. *Measuring Rurality: Rural-Urban Continuum Codes*. Washington, DC: US Dept of Agriculture; 2006.
21. Dartmouth Medical School. *The Dartmouth Atlas of Health Care in the United States 1998*. Hanover, NH: Center for the Evaluative Clinical Sciences, Dartmouth Medical School; 1999.
22. *Stata Statistical Software: Release 9*. College Station, TX: StataCorp LP; 2005.
23. Johnston J, DiNardo J. *Econometric Methods*. 4th ed. New York, NY: McGraw-Hill Co; 1997.
24. Fried TR, Bradley EH, Towle VR, Allore H. Understanding the treatment preferences of seriously ill patients. *N Engl J Med*. 2002;346(14):1061-1066.
25. Phipps E, True G, Harris D, et al. Approaching the end of life: attitudes, preferences, and behaviors of African-American and white patients and their family caregivers. *J Clin Oncol*. 2003;21(3):549-554.
26. Pritchard RS, Fisher ES, Teno JM, et al. Influence of patient preferences and local health system characteristics on the place of death. *J Am Geriatr Soc*. 1998; 46(10):1242-1250.
27. Campbell DE, Lynn J, Louis TA, Shugarman LR. Medicare program expenditures associated with hospice use. *Ann Intern Med*. 2004;140(4):269-277.
28. Chen H, Haley W, Robinson B, Schonwetter R. Decisions for hospice care in patients with advanced cancer. *J Am Geriatr Soc*. 2003;51(6):789-797.
29. Crawley L, Payne R, Bolden J, et al. Palliative and end-of-life care in the African American community. *JAMA*. 2000;284(19):2518-2521.
30. Degenholtz HB, Arnold RA, Meisel A, Lave JR. Persistence of racial disparities in advance care plan documents among nursing home residents. *J Am Geriatr Soc*. 2002;50(2):378-381.
31. DePalma J. How race and culture influence advance directive decisions. *Am J Nurs*. 1998;98(3):31.
32. Earle CC, Neville BA, Landrum MB, Ayanian JZ, Block SD, Weeks JC. Trends in the aggressiveness of cancer care near the end of life. *J Clin Oncol*. 2004;22 (2):315-321.
33. Flory J, Yinong YX, Gurol I, Levinsky N, Ash A, Emanuel E. Place of death: U.S. trends since 1980. *Health Aff (Milwood)*. 2004;23(3):194-200.
34. Kiely DK, Mitchell SL, Marlow A, Murphy KM, Morris JN. Racial and state differences in the designation of advance directives in nursing home residents. *J Am Geriatr Soc*. 2001;49(10):1346-1352.
35. McCarthy EP, Burns RB, Ngo-Metzger Q, Davis RB, Phillips RS. Hospice use among Medicare managed care and fee-for-service patients dying with cancer. *JAMA*. 2003;289(17):2238-2245.
36. Silveira MJ, DiPiero A, Gerrity MS, Feudtner C. Patients' knowledge of options at the end of life: ignorance in the face of death. *JAMA*. 2000;284(19):2483-2488.
37. 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.
38. Franks P, Fiscella K, Meldrum S. Racial disparities in the content of primary care office visits. *J Gen Intern Med*. 2005;20(7):599-603.
39. Schneider EC, Zaslavsky AM, Epstein AM. Racial disparities in the quality of care for enrollees in medicare managed care. *JAMA*. 2002;287(10):1288-1294.
40. Fiscella K, Holt K. Impact of primary care patient visits on racial and ethnic disparities in preventive care in the United States. *J Am Board Fam Med*. 2007; 20(6):587-597.
41. Kronman AC, Ash AS, Freund KM, Hanchate A, Emanuel EJ. Can primary care visits reduce hospital utilization among Medicare beneficiaries at the end of life? *J Gen Intern Med*. 2008;23(9):1330-1335.
42. McBean AM. *Medicare Race and Ethnicity Data*. Minneapolis, MN: National Academy of Social Insurance; 2004.
43. Kirby JB, Taliaferro G, Zuvekas SH. Explaining racial and ethnic disparities in health care. *Med Care*. 2006;44(5)(suppl):I64-I72.

Published Ahead of Print on May 14, 2009, as 10.2105/AJPH.2008.150730
The latest version is at http://www.ajph.org/cgi/doi/10.2105/AJPH.2008.150730

Perceived Racial Discrimination in Health Care: A Comparison of Veterans Affairs and Other Patients

Leslie R.M. Hausmann, PhD, Kwonho Jeong, BA, James E. Bost, PhD, Nancy R. Kressin, PhD, and Said A. Ibrahim, MD, MPH

Objectives. We compared rates of perceived racial discrimination in health care settings for veteran and nonveteran patients and for veterans who used the Veterans Affairs health care system and those who did not.

Methods. Data were drawn from the 2004 Behavioral Risk Factor Surveillance System. We used logistic regression to examine whether perceived racial discrimination in health care was associated with veteran status or use of Veterans Affairs health care, after adjusting for patient characteristics.

Results. In this sample of 35902 people, rates of perceived discrimination were equal for veterans and nonveterans (3.4% and 3.5%, respectively; crude odds ratio [OR]=1.00; 95% confidence interval [CI]=0.77, 1.28; adjusted OR=0.92; 95% CI=0.66, 1.28). Among veterans (n=3420), perceived discrimination was more prevalent among patients who used Veterans Affairs facilities than among those who did not (5.4% vs 2.7%; OR=2.08; 95% CI=1.04, 4.18). However, this difference was not significant after adjustment for patient characteristics (OR=1.30; 95% CI=0.54, 3.13).

Conclusions. Perceived racial discrimination in health care was equally prevalent among veterans and nonveterans and among veterans who used the Veterans Affairs health care system and those who did not. (*Am J Public Health.* 2009;99:XXX–XXX. doi:10.2105/AJPH.2008.150730)

Amidst national efforts to understand and eliminate pervasive racial and ethnic health disparities,[1,2] research has documented the deleterious effects of perceived racial or ethnic discrimination on the health of patients. There is strong and consistent evidence that patients who perceive racial or ethnic discrimination are at greater risk for poor health, as defined by a host of health outcomes (e.g., mortality, depression, self-assessed health status) and health-related behaviors (e.g., smoking, substance abuse).[1–7] Moreover, racial or ethnic discrimination that occurs specifically within health care settings is associated with poorer health status, lower patient satisfaction with care, and, in some cases, less health care utilization.[3,8–15]

To support the development of targeted interventions to reduce instances of discrimination and minimize the negative consequences of perceived discrimination for those most at risk, it is necessary to identify vulnerable patient populations or health care settings in which patients are more likely to perceive discrimination. A higher prevalence of perceived racial discrimination in health care settings has already been found in some patient populations, such as racial/ethnic minorities and those who have limited economic resources.[3,12,16–19]

We examined rates of perceived discrimination for another patient population: military veterans. Veterans are a minority population (about 10% of US adults[20]) with unique health care needs. Not only do those serving in the military face substantial physical and psychological challenges that put their health at risk, but those who have served also become part of a culture of veterans that can shape the way they interact with the health care system.[21–23] One way to gauge whether the health care system adequately adapts to the needs of this patient population is to compare reports of perceived discrimination in health care among veterans and nonveterans.

We also compared the prevalence of perceived racial discrimination in health care among veterans who received health care from the Veterans Affairs (VA) health care system with veterans who received care outside of this system. Unique features of the VA health care system make it an interesting setting in which to examine rates of perceived discrimination. The VA patient population includes a disproportionate number of patients who belong to racial and socioeconomic groups that are at increased risk of experiencing discrimination.[24,25] This might suggest that perceived racial discrimination would be more prevalent among VA patients than among others. However, the VA has undertaken substantial efforts in recent years to improve its delivery of health care, including addressing potential racial/ethnic disparities in care.[26–28] These changes have yielded extraordinary improvements across a variety of quality indicators,[29] making the VA a model health care system both nationally and internationally.[30–32] These developments might suggest that patients in the VA health care system would be less likely than would other patients to perceive racial discrimination. We explored potential differences in perceived racial discrimination among VA and other patient populations with data from a national survey.

METHODS

We used data from the Behavioral Risk Factor Surveillance System (BRFSS), a national telephone survey conducted annually to monitor health conditions and risk behaviors of US adults.[33] It uses state-level sampling plans and data weights to obtain a sample that represents the population of households with telephones within each state. Complete BRFSS data files are publicly available on the BRFSS Web site.[33] We analyzed 2004 BRFSS data from the District of Columbia and states (Arkansas, Colorado, Delaware, Mississippi, Rhode Island, South Carolina, and Wisconsin) that administered the optional Reactions to Race module, which included a question about whether respondents perceived racial discrimination while seeking health care in the past 12 months.

Measures

The outcome of interest was perceived racial discrimination in health care, which

TABLE 1—Sample Characteristics of Veterans and Nonveterans: Behavioral Risk Factor Surveillance System, 2004

	Veterans, %	Nonveterans, %	*P*
Race/ethnicity			<.001
White	84.5	79.3	
Hispanic	3.1	6.4	
African American	12.4	14.3	
Racial salience[a]			.45
≤1/mo	82.3	81.8	
1/wk	7.1	6.8	
≥1/d	10.5	11.4	
Gender			<.001
Women	6.9	59.4	
Men	93.1	40.6	
Age, y			<.001
18–24	4.5	15.4	
25–34	9.3	19.4	
35–44	13.9	21.2	
45–54	17.2	19.2	
55–64	22.6	11.3	
≥65	32.5	13.5	
Income, $			<.001
<15 000	6.2	11.3	
15 000–24 999	16.1	18.2	
25 000–34 999	15.5	15.0	
35 000–49 999	20.6	17.1	
>50 000	41.6	38.5	
Education			<.001
<High school	6.2	11.6	
High school graduate	31.2	32.6	
Some college	29.1	25.6	
College degree	33.5	30.2	
Health care coverage			<.001
Yes	90.8	84.1	
No	9.2	15.9	
Cost of medical care prohibitive in past 12 mo			<.001
No	91.3	86.0	
Yes	8.7	14.0	
Health status			<.001
Excellent, very good, or good	82.1	85.0	
Fair or poor	17.9	15.0	
State			<.001
Arkansas	13.0	12.0	
Colorado	20.3	20.3	
Delaware	4.1	3.7	
District of Columbia	2.0	2.8	
Mississippi	11.4	12.9	

Continued

was assessed with the following item: "Within the past 12 months when seeking health care, do you feel your experiences were worse than, the same as, or better than for people of other races?" Possible responses were worse than other races; the same as other races; better than other races; worse than some races, better than others; and only encountered people of the same race. We excluded the latter 2 responses from our analyses because relatively few people chose these responses (0.3% each) and they did not unambiguously indicate the presence or absence of discrimination. "Worse than other races" responses were coded as having experienced discrimination. Responses of "the same as other races" and "better than other races" were coded as not having experienced discrimination.

The primary predictors were veteran status and whether veteran respondents used VA medical facilities in the past 12 months. Veteran status was assessed with the yes or no item, "Have you ever served on active duty in the United States Armed Forces, either in the regular military or in a National Guard or military reserve unit?" Respondents who answered yes were asked whether they were currently on active duty or in a National Guard or reserve unit, retired from military service, medically discharged from military service, or discharged from military service. To assess VA health care utilization, veterans were asked, "In the last 12 months have you received some or all of your health care from VA facilities?" Possible answers were yes, all; yes, some; and no VA health care received. Analyses of VA health care utilization included the subsample of veteran patients who reported receiving all or none of their medical care from VA facilities; those who received some care from the VA (n=422) were excluded because we could not determine whether health care discrimination reported by these patients occurred in VA or non-VA settings. Respondents who were on active duty or in a National Guard or reserve unit were also excluded from these analyses because they were not eligible to receive health care at VA medical facilities.

The following patient variables served as covariates in multivariable models: self-reported racial/ethnic group, racial salience (how often respondents thought about their race), gender, age, annual household income, highest educational attainment, health care coverage, affordability of medical care, health status,

TABLE 1—*Continued*

Rhode Island	4.5	5.1
District of Columbia	22.4	18.2
Wisconsin	22.3	25.0

Note. Percentages were based on weighted data. For veterans, unweighted n = 5 233; weighted n = 2 363 540. For nonveterans, unweighted n = 30 669; weighted n = 13 502 210.
[a]Defined as frequency of thoughts about one's own race.

and state of residence. Race/ethnicity was categorized as White, African American, Hispanic, other (Asian, Native Hawaiian, Pacific Islander, American Indian, or Alaska Native), or multiple (more than 1 racial/ethnic group). Other and multiple racial/ethnic groups were excluded from the analyses because of the relatively small size of these groups (3.2% and 1.4% of respondents, respectively) and because their heterogeneity made drawing conclusions difficult.

Racial salience was included because it has been shown to be positively associated with perceptions of racial discrimination.[3] It was assessed by the item, "How often do you think about your race?" Possible responses were never, once a year, once a month, once a week, once a day, once an hour, and constantly. Responses were collapsed into 3 categories: once a month or less, once a week, and once a day or more. Health care coverage was assessed with the yes or no item, "Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare?" Affordability of medical care was assessed with the yes or no item, "Was there a time in the past 12 months when you needed to see a doctor but could not because of the cost?" Self-reported health status was assessed with the item, "Would you say that in general your health is excellent, very good, good, fair, or poor?" Responses were dichotomized into excellent, very good, or good versus fair or poor.

Statistical Analyses

We summarized respondent characteristics for veterans and nonveterans and compared them using the χ^2 test. We used logistic regression models to estimate the crude association between veteran status and the prevalence of perceived racial discrimination, and to estimate the association after adjusting for respondent characteristics.

We conducted similar analyses on the subsample of veteran respondents, with VA health care utilization as the primary predictor. Specifically, we compared respondent characteristics for those who used VA medical facilities with those who did not. We then used logistic regression models to estimate the crude and adjusted associations between VA utilization and perceived racial discrimination.

To explore whether the associations between veteran status or VA utilization and perceived racial discrimination varied across racial/ethnic groups, we ran additional models to test for interactions between race/ethnicity and veteran status, as well as between race/ethnicity and VA utilization. These interactions were not significant and are not reported here.

TABLE 2—Sample Characteristics of Veterans Who Used the Veterans Affairs (VA) Health Care System and Veterans Who Did Not: Behavioral Risk Factor Surveillance System, 2004

	VA Users, %	VA Nonusers, %	*P*
Race/ethnicity			<.001
White	73.2	87.8	
Hispanic	3.9	2.2	
African American	22.9	10.0	
Racial salience[a]			.12
≤1/mo	79.6	83.5	
1/wk	6.3	7.0	
≥1/d	14.1	9.5	
Gender			.85
Women	6.2	6.4	
Men	93.8	93.6	
Age, y			.02
18–24	2.6	1.6	
25–34	4.7	7.8	
35–44	10.3	13.8	
45–54	23.4	18.4	
55–64	30.9	24.8	
≥65	28.2	33.6	
Income, $			<.001
<15 000	20.4	4.9	
15 000–24 999	28.5	14.2	
25 000–34 999	20.9	14.2	
35 000–49 999	14.4	20.6	
>50 000	15.8	46.2	
Education			<.001
<High school	11.6	5.7	
High school graduate	36.1	30.4	
Some college	33.3	27.9	
College degree	19.0	36.1	
Health care coverage			<.001
Yes	81.9	91.5	
No	18.1	8.5	

Continued

TABLE 2—*Continued*

Cost of medical care prohibitive in past 12 mo			.04
No	88.0	91.4	
Yes	12.0	8.6	
Health status			<.001
Excellent, very good, or good	62.2	83.9	
Fair or poor	37.8	16.1	
State			<.001
Arkansas	18.2	11.8	
Colorado	14.0	21.5	
Delaware	3.5	4.4	
District of Columbia	3.4	1.7	
Mississippi	15.9	10.4	
Rhode Island	4.2	4.5	
South Carolina	26.8	21.3	
Wisconsin	14.0	24.5	

Note. Percentages were based on weighted data. For VA users, unweighted n = 362; weighted n = 140 672. For VA nonusers, unweighted n = 3 058; weighted n = 1 406 880.
[a]Defined as frequency of thoughts about one's own race.

State was included as a control variable in all adjusted models. In all analyses, we incorporated the BRFSS weighting and design variables into the models with Stata/IC version 10.0 for Windows (StataCorp LP, College Station, TX).

RESULTS

The study sample included 35 902 respondents, who represented 15 865 750 people when data were weighted to reflect state populations. Veterans and nonveterans differed significantly ($P<.001$) on almost all background characteristics (Table 1). For example, veterans were more likely than were nonveterans to be White (84.5% versus 79.3%) and male (93.1% versus 40.6%) and to have health care coverage (90.8% versus 84.1%). Veterans were also significantly older: 32.5% of veterans and 13.5% of nonveterans were older than 65 years.

Analyses comparing veteran VA health care users and nonusers included a subsample of 3420 respondents, who represented 1 547 552 people when data weights were applied (Table 2). VA users and nonusers differed significantly ($P<.04$) on the majority of background characteristics. VA users were more likely than were nonusers to be African American (22.9% versus 10.0%), Hispanic (3.9% versus 2.2%), and middle aged (23.4% versus 18.4% were aged 45–54 years, and 30.9% versus 24.8% were aged 55–64 years). VA users also differed significantly from nonusers on several variables indicative of socioeconomic status. VA users had lower incomes than did nonusers (20.4% versus 4.9% had annual incomes below $15 000), were less educated (11.6% versus 5.7% did not finish high school), were less likely to have health care coverage (81.9% versus 91.5%), and were more likely to report having to forgo medical care because of cost (12.0% versus 8.6%). VA users were also more likely than were nonusers to report fair or poor health status (37.8% versus 16.1%).

Perceived racial discrimination in health care was reported by 3.4% of veterans and 3.5% of nonveterans. This difference was not statistically significant in an unadjusted analysis (odds ratio [OR] = 1.00; 95% confidence interval [CI] = 0.77, 1.28) or in a multivariable model that adjusted for respondent characteristics (OR = 0.92; 95% CI = 0.66, 1.28; Table 3). In the adjusted model, higher odds of perceived discrimination were significantly associated with African American race, greater racial salience, male gender, younger age, an annual income of less than $15 000, having a high school diploma, having to forgo medical care because of cost, and fair or poor health status (Table 3).

In the veterans subsample, reports of perceived racial discrimination in health care were significantly more common among VA users than nonusers (5.4% versus 2.7%; $P<.03$). In a model that did not take into account additional respondent characteristics, veterans who received health care from the VA were 2.08 (95% CI = 1.04, 4.18) times as likely to report perceived racial discrimination as veterans who received health care from non-VA facilities (Table 4). After we controlled for respondent characteristics, however, the likelihood of reporting perceived discrimination was not significantly different for VA users and nonusers (OR = 1.30; 95% CI = 0.54, 3.13). In the adjusted analysis, higher odds of perceived discrimination were significantly associated with African American race, greater racial salience (thinking about race once a day or more), age (25–34 years), an annual income of less than $15 000, and fair or poor health status (Table 4).

DISCUSSION

We used data from a national survey to examine whether rates of perceived racial discrimination in health care varied across different patient populations and health care settings. Specifically, we compared the prevalence of perceived discrimination among veterans and nonveterans and among veterans who received care in VA and non-VA health care systems. We found that rates of perceived racial discrimination in health care were low overall and did not differ for veterans and nonveterans. To our knowledge, this is the first study to compare rates of perceived racial discrimination in health care for veterans and nonveterans. Our findings suggest that, despite the unique experiences and health care needs of veterans, those who have served in the military are not more likely to perceive racial discrimination in health care settings.

We also found that, in the subsample of veterans who were eligible to receive health care from VA medical facilities, veterans who received all of their care in the VA system were twice as likely to report perceptions of racial discrimination in the health care setting than were veterans who received their care outside of the VA system. This difference, however, was eliminated after we controlled for differences in patient characteristics, such as race, indicators of socioeconomic status, access to health care, and health status.

Our findings are consistent with 2 previous studies that found no differences in rates of

TABLE 3—Crude and Adjusted Odds Ratios (ORs) of Perceived Racial Discrimination in Health Care Among Veterans and Nonveterans: Behavioral Risk Factor Surveillance System, 2004

	Reported Discrimination, %	Crude OR[a] (95% CI)	Adjusted OR[b] (95% CI)
Veteran status			
Veteran	3.4	1.00 (0.77, 1.28)	0.92 (0.66, 1.28)
Nonveteran (Ref)	3.5	1.00	1.00
Race/ethnicity			
White (Ref)	2.0	1.00	1.00
Hispanic	5.2	2.73 (1.88, 3.98)	1.11 (0.66, 1.85)
African American	10.9	6.02 (5.03, 7.21)	3.27 (2.50, 4.28)
Racial salience[c]			
≤ Once a mo (Ref)	2.4	1.00	1.00
Once a wk	4.2	1.83 (1.36, 2.47)	1.48 (1.03, 2.12)
≥ Once a d	10.6	4.93 (4.06, 5.97)	2.55 (1.96, 3.32)
Gender			
Women (Ref)	3.1	1.00	1.00
Men	3.8	1.22 (1.03, 1.44)	1.34 (1.09, 1.66)
Age, y			
18–24	3.7	2.30 (1.59, 3.31)	1.69 (1.04 2.75)
25–34	4.1	2.55 (1.86, 3.50)	2.06 (1.35, 3.14)
35–44	3.4	2.08 (1.54, 2.82)	1.81 (1.22, 2.70)
45–54	4.0	2.48 (1.84, 3.32)	2.03 (1.37, 3.01)
55–64	3.5	2.17 (1.59, 2.96)	1.97 (1.33, 2.93)
≥ 65 (Ref)	1.6	1.00	1.00
Income, $			
< 15 000	8.9	5.63 (4.30, 7.36)	1.80 (1.21, 2.69)
15 000–24 999	5.7	3.45 (2.64, 4.50)	1.40 (0.99, 1.98)
25 000–34 999	2.8	1.68 (1.23, 2.29)	1.01 (0.69, 1.46)
35 000–49 999	3.0	1.75 (1.28, 2.38)	1.16 (0.82, 1.63)
> 50 000 (Ref)	1.7	1.00	1.00
Education			
< High school	5.6	2.81 (2.13, 3.71)	1.12 (0.74, 1.70)
High school graduate	4.6	2.29 (1.83, 2.87)	1.53 (1.15, 2.03)
Some college	3.0	1.46 (1.13, 1.89)	0.99 (0.73, 1.35)
College degree (Ref)	2.1	1.00	1.00
Health care coverage			
Yes (Ref)	2.7	1.00	1.00
No	7.9	3.08 (2.56, 3.72)	1.18 (0.92, 1.51)
Cost of medical care prohibitive in past 12 mo			
No	2.1	6.32 (5.32, 7.52)	3.64 (2.86, 4.63)
Yes (Ref)	12.2	1.00	1.00
Health status			
Excellent, very good, or good (Ref)	2.8	1.00	1.00
Fair or poor	7.6	2.90 (2.45, 3.44)	1.76 (1.39, 2.24)

Note. CI = confidence interval.
[a]Unadjusted ORs reflect the bivariate associations between perceived discrimination and each variable (weighted n = 13 374 133).
[b]Adjusted ORs reflect the association between perceived discrimination and each variable, after adjustment for all the other variables. State was included as an additional covariate in the adjusted model (weighted n = 11 036 142).
[c]Defined as frequency of thoughts about one's own race.

perceived racial discrimination in health care between VA users and nonusers.[13,34] The issue of perceived racial discrimination among veterans received national attention when a report was released in 2007 indicating that more than 50% of African American veterans could recall a situation in which they experienced discrimination wherever they received health care services, in either VA or non-VA facilities.[34] Although the overall rate of perceived discrimination documented in that study was much higher than in our study, that report found no differences in perceived discrimination between veterans who received care at VA facilities and those who did not. The earlier report was greatly limited by its reliance on a small convenience sample of 141 African American veterans within a single geographic location.

Another study examined the prevalence of racial discrimination in health care among patients drawn from university, community, and VA clinics.[13] Although examining the prevalence of discrimination across different health care settings was not the focus of the study, the authors reported that the prevalence of perceived racial discrimination was not significantly different among those recruited from VA clinics than among those from university or community clinics. That study was limited by its inclusion of patients from only 3 VA facilities and 2 non-VA health care systems; it also did not account for the possibility that patients received care from more than 1 health care system.

Our study, which used data from a national survey and included respondents from several states, provided more robust evidence that the prevalence of racial discrimination in health care settings does not differ between veterans and nonveterans or between patients who receive care at VA facilities and those who do not, once differences in characteristics of VA users and nonusers are taken into account. Although we found significantly higher odds of perceived discrimination among VA users than among nonusers in unadjusted analyses, this difference was likely attributable to the higher prevalence of patient characteristics that put patients at higher risk of discrimination (e.g., minority race and lower socioeconomic status) among VA users rather than to systemic differences between VA and non-VA health care settings that increased the likelihood of discriminatory experiences. The

TABLE 4—Crude and Adjusted Odds Ratios (ORs) of Perceived Racial Discrimination in Health Care Among Veterans Affairs (VA) Health Care System Users and Nonusers: Behavioral Risk Factor Surveillance System, 2004

	Reported Discrimination, %	Crude OR[a] (95% CI)	Adjusted OR[b] (95% CI)
VA health care utilization			
VA user	5.4	2.08 (1.04, 4.18)	1.30 (0.54, 3.13)
VA nonuser (Ref)	2.7	1.00	1.00
Race/ethnicity			
White (Ref)	1.6	1.00	1.00
Hispanic	3.0	1.93 (0.48, 7.73)	1.43 (0.25, 8.23)
African American	13.0	9.23 (5.26, 16.20)	4.33 (1.70, 11.0)
Racial salience[c]			
≤ Once a mo (Ref)	2.0	1.00	1.00
Once a wk	4.1	2.11 (0.80, 5.56)	1.43 (0.30, 6.67)
≥ Once a d	11.5	6.46 (3.67, 11.39)	2.57 (1.02, 6.49)
Gender			
Women (Ref)	3.1	1.00	1.00
Men	3.1	0.98 (0.42, 2.29)	2.10 (0.56, 7.90)
Age, y			
18–24	7.3	4.70 (0.59, 37.45)	. . .
25–34	6.8	4.31 (1.69, 10.98)	4.21 (1.12, 15.9)
35–44	3.1	1.88 (0.85, 4.18)	1.81 (0.66, 4.96)
45–54	3.3	2.05 (1.02, 4.12)	1.90 (0.72, 4.99)
55–64	3.2	1.96 (1.03, 3.73)	1.93 (0.77, 4.85)
≥ 65 (Ref)	1.7	1.00	1.00
Income, $			
< 15 000	9.0	6.40 (2.77, 14.78)	3.55 (1.26, 10.0)
15 000–24 999	4.3	2.93 (1.32, 6.47)	1.33 (0.47, 3.76)
25 000–34 999	3.4	2.28 (1.00, 5.20)	1.81 (0.70, 4.68)
35 000–49 999	2.6	1.71 (0.72, 4.09)	1.21 (0.45, 3.25)
> 50 000 (Ref)	1.5	1.00	1.00
Education			
< High school	2.9	1.56 (0.55, 4.44)	0.83 (0.17, 3.96)
High school graduate	4.2	2.33 (1.22, 4.44)	1.83 (0.78, 4.29)
Some college	3.3	1.82 (0.91, 3.65)	1.26 (0.49, 3.24)
College degree (Ref)	1.9	1.00	1.00
Health care coverage			
Yes (Ref)	2.6	1.00	1.00
No	8.5	3.55 (1.79, 7.01)	0.85 (0.37, 1.93)
Cost of medical care prohibitive in past 12 mo			
No (Ref)	2.4	1.00	1.00
Yes	9.9	4.44 (2.47, 7.99)	1.82 (0.77, 4.33)
Health status			
Excellent, very good, or good (Ref)	2.4	1.00	1.00
Fair or poor	6.2	2.64 (1.61, 4.34)	1.66 (0.81, 3.38)

Note. CI = confidence interval. Ellipsis indicates too few respondents with complete data to estimate in adjusted model.
[a]Unadjusted ORs reflect the bivariate associations between perceived discrimination and each variable (weighted n = 1 727 967).
[b]Adjusted ORs reflect the association between perceived discrimination and each variable, after adjustment for all the other variables. State was included as an additional covariate in the adjusted model (weighted n = 1 293 909).
[c]Defined as frequency of thoughts about one's own race.

VA should keep the special nature of its patient population in mind when seeking to promote equity and fairness in the health care it provides for veterans.

Our study had several limitations. Because only a subset of states administered the Reactions to Race module that we analyzed, our sample was not representative of the entire US population. Furthermore, although the study sample was representative of the states from which respondents were drawn, it was not representative of veterans within those states. The findings may therefore not be generalizable to the entire population of US veterans, nor was it possible to assess potential nationwide geographic differences in perceived discrimination with the available BRFSS data.

Another limitation was that the BRFSS survey only assessed perceived racial discrimination in health care. It is possible that other types of perceived discrimination are more common in VA than in non-VA settings. For example, 1 study found that gender discrimination in health care was more commonly reported among women recruited from VA facilities than among those recruited from non-VA facilities.[13] It is also possible that veterans who receive care in non-VA settings may be more likely to perceive discrimination related to their veteran status than are those who receive care in the VA, but we could not examine this possibility within this data set.

Finally, our focus on discrimination in VA and non-VA settings represented a crude attempt to examine whether discrimination varied across different types of health care settings; information that would allow a more in-depth exploration of features of specific settings or facilities in which discrimination was perceived most often was not available. Future studies should examine whether rates of perceived discrimination are associated with additional features of health care settings, such as location, size, or the proportion of racial/ethnic minority or female providers.

Our study used the best available data to explore whether rates of perceived racial discrimination in health care varied among veterans and nonveterans and among veterans treated in VA and non-VA settings. Our findings of significant differences in unadjusted rates of perceived discrimination between veterans treated in VA settings and those treated in other settings suggest that the VA serves a special patient population that

may be more vulnerable to experiences of discrimination. However, rates of perceived racial discrimination among veterans treated in VA and non-VA settings were similar once differences in characteristics of the patients served in these settings were taken into account. This suggests that the VA is doing as well as other health care providers in the way patients perceive they are treated while obtaining health care. ■

About the Authors

Leslie R.M. Hausmann and Said A. Ibrahim are with the Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, Pittsburgh, PA. Said A. Ibrahim is also with the Division of General Internal Medicine, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh. Kwonho Jeong and James E. Bost are with the Center for Research on Health Care Data Center, University of Pittsburgh, Pittsburgh. Nancy R. Kressin is with the Center for Health Quality, Outcomes and Economic Research, Boston VA Healthcare System, Boston, MA, and the Section of General Internal Medicine, Boston University School of Medicine, Boston.

Requests for reprints should be sent to: Leslie R.M. Hausmann, VA Pittsburgh Healthcare System, 7180 Highland Dr (151C-H), Pittsburgh, PA 15206-1206 (e-mail: leslie.hausmann@gmail.com).

This article was accepted October 16, 2008.

Contributors

L.R.M. Hausmann originated the study, played a key role in implementing the study and writing the article, and participated in data analysis and interpretation. K. Jeong completed the analyses and assisted with interpretation of the findings. J.E. Bost assisted with data analysis and interpretation. N.R. Kressin contributed to interpreting the findings. S.A. Ibrahim contributed to designing the study and interpreting the findings. All authors helped to conceptualize ideas and review drafts of the article.

Acknowledgments

L.R.M. Hausmann was supported by a VA Health Services Research and Development Career Development Award (RCD 06-287). N.R. Kressin was supported by a Research Career Scientist award from the Department of Veterans Affairs, Health Services Research and Development (RCS 02-066-1). S.A. Ibrahim was supported by a VA Health Services Research Career Development Award, the Harold Amos Robert Wood Johnson Scholar Award, and an award from the National Institutes of Musculoskeletal and Skin Disorders (1K24AR055259-01). This study was also supported by the National Center for Research Resources (grant UL1 RR024153) and the National Institutes of Health Roadmap for Medical Research.

Information on the National Center for Research Resources is available at http://www.ncrr.nih.gov/. Information on Re-engineering the Clinical Research Enterprise can be obtained from http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp.

Note. This article is solely the responsibility of the authors and does not necessarily represent the official view of National Center for Research Resources, National Institutes of Health, or the VA.

Human Participant Protection

This study was approved by the VA Pittsburgh Healthcare System institutional review board.

References

1. Smedley BD, Stith AY, Nelson AR, eds. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academies Press; 2003.

2. Healthy People 2001 midcourse review. Available at: http://www.healthypeople.gov/Data/midcourse/default.htm. Accessed April 18, 2008.

3. 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.

4. Paradies Y. A systematic review of empirical research on self-reported racism and health. *Int J Epidemiol*. 2006;35(4):888–901.

5. Williams DR, Neighbors HW, Jackson JS. Racial/ethnic discrimination and health: findings from community studies. *Am J Public Health*. 2003;93(2):200–208.

6. 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.

7. 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.

8. 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.

9. 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.

10. Trivedi AN, Ayanian JZ. Perceived discrimination and use of preventive health services. *J Gen Intern Med*. 2006;21(6):553–558.

11. 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):721–730.

12. 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.

13. 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.

14. Ryan AM, Gee GC, Griffith D. The effects of perceived discrimination on diabetes management. *J Health Care Poor Underserved*. 2008;19(1):149–163.

15. 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.

16. Stepanikova I, Cook KS. Effects of poverty and lack of insurance on perceptions of racial and ethnic bias in health care. *Health Serv Res*. 2008;43(3):915–930.

17. 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.

18. 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.

19. LaVeist TA, Rolley NC, Diala C. Prevalence and patterns of discrimination among US health care consumers. *Int J Health Serv*. 2003;33(2):331–344.

20. US Census Bureau American Community Survey. Available at: http://www.census.gov/acs/www/index.html. Accessed October 7, 2008.

21. Hobbs K. Reflections on the culture of veterans. *AAOHN J*. 2008;56(8):337–341.

22. Morgan RO, Teal CR, Reddy SG, Ford ME, Ashton CM. Measurement in Veterans Affairs Health Services Research: veterans as a special population. *Health Serv Res*. 2005;40(5 pt 2):1573–1583.

23. Williams JW Jr. Serving the health needs of our military and veterans. *N C Med J*. 2008;69(1):23–26.

24. Agha Z, Lofgren RP, VanRuiswyk JV, Layde PM. Are patients at Veterans Affairs medical centers sicker? A comparative analysis of health status and medical resource use. *Arch Intern Med*. 2000;160(21):3252–3257.

25. Wilson NJ, Kizer KW. The VA health care system: an unrecognized national safety net. *Health Aff (Millwood)*. 1997;16(4):200–204.

26. Hayward RA, Hofer TP, Kerr EA, Krein SL. Quality improvement initiatives: issues in moving from diabetes guidelines to policy. *Diabetes Care*. 2004;27(suppl 2): B54–B60.

27. Reiber GE, Au D, McDonell M, Fihn SD. Diabetes Quality Improvement in Department of Veterans Affairs Ambulatory Care Clinics: a group-randomized clinical trial. *Diabetes Care*. 2004;27(suppl 2):B61–B68.

28. Reiber GE, Boyko EJ. Diabetes research in the Department of Veterans Affairs. *Diabetes Care*. 2004; 27(suppl 2):B95–B98.

29. Jha AK, Perlin JB, Kizer KW, Dudley RA. Effect of the transformation of the Veterans Affairs Health Care System on the quality of care. *N Engl J Med*. 2003; 348(22):2218–2227.

30. Fooks C, Decter M. The transformation experience of the Veterans Health Administration and its relevance to Canada. *Healthc Pap*. 2005;5(4):60–64.

31. Weatherill S. The VHA's commitment to accountability: a "Third Way" for Medicare? *Healthc Pap*. 2005; 5(4):38–42.

32. Ramirez B, Nazaretian M. Quality, value, accountability and information as transforming strategies for patient-centred care: a commentary from an international perspective. *Healthc Pap*. 2005;5(4):52–55.

33. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System survey data. Available at: http://www.cdc.gov/brfss/index.htm. Accessed June 12, 2008.

34. Rickles NM, Dominguez S, Fennell H, Roberts LE, Warren JD, Amaro H. *Health Care Experiences and Health Outcomes of African-American Veterans*. Boston, MA: Institute on Urban Health Research; 2007.

Substance Abuse, 30:118–126, 2009

ISSN: 0889-7077 print / 1547-0164 online
DOI: 10.1080/08897070902802034

Routledge
Taylor & Francis Group

Alcohol-Related Problems Among Younger Drinkers Who Misuse Prescription Drugs: Results from the National Epidemiologic Survey of Alcohol and Related Conditions (NESARC)

J. Hermos, MD
M. Winter, MPH
T. Heeren, PhD
R. Hingson, PhD

ABSTRACT. The authors determined whether lifetime prescription drug misuse (PDM) associated with increased risks for alcohol-related problems among 18- to 34-year-old, NESARC respondents. Among 8222 "ever-drinkers," 15.4% reported ever "*misusing sedatives, tranquilizers, painkillers or stimulants... as prescriptions or from indirect sources.*" Outcomes were within two alcohol-related problem domains, "risk-taking behaviors," including driving while drinking, fights, injuries, and arrests, and "interpersonal troubles," including problems with jobs, family, or friends. Among all drinkers and among alcohol-dependent and cannabis-using subsamples, those reporting PDM were significantly more likely to report alcohol-related "risk-taking behaviors" or "interpersonal troubles" than were those without PDM. In adjusted analysis, young age drinking onsets, and heavy and dependent drinking independently increased these risks. Results of this cross-sectional analysis support the need for longitudinal data to more clearly define the association between drinking problems and PDM, and which can support prevention, treatment, and harm-reduction efforts for younger, multisubstance users.

KEYWORDS. Alcohol-related problems, prescription drug misuse, young drinkers

Misuse of prescribed psychoactive drugs with abuse potential ("prescription drug misuse" [PDM]), a problem of increasing national concern, may range from intentional or unintentional overuse of personal prescriptions to overt recreational use of diverted drugs

J. Hermos is affiliated with Youth Alcohol Prevention Center and Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts, USA.
M. Winter and T. Heeren are affiliated with Youth Alcohol Prevention Center and Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA.
R. Hingson is affiliated with Youth Alcohol Prevention Center and Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts, USA; and Division of Epidemiology and Prevention, National Institute of Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland, USA.
Address correspondence to: J. Hermos, MD, VA Boston Healthcare System (151MAV), 150 S. Huntington Avenue, Boston, MA 02130 USA (E-mail: jhermos@bu.edu).
This work was supported by Core Funds from the National Institute on Alcohol Abuse and Alcoholism for the Youth Alcohol Prevention Center, Boston University School of Public Health, and Core Funds from the Statistical Coordinating Center, Boston University School of Public Health.

(1–3). The classifications of prescribed psychoactive drugs most commonly currently misused are "painkillers" (opioid analgesics), minor tranquilizers (principally benzodiazepines), sedative-hypnotics, and central nervous system (CNS) stimulants.

Recent studies utilizing data from the 2001–2002 National Epidemiological Survey on Alcohol and Related Conditions (NESARC) indicate a close association of PDM with alcohol use disorders. Stinson et al. (4) and Huang et al. (5) reported strong concordance between alcohol use and any drug use disorders, including those relating to misuse of opioids, sedatives, tranquilizers, and amphetamines, among all NESARC respondents but especially so among 18- to 29-year-olds. McCabe et al. (6) reported that about 25% of 18- to 24-year-old NESARC respondents meeting criteria for alcohol dependence reported concurrent PDM. Also using the NESARC database, we reported, among 18- to 34-year-old NESARC respondents, that the risks of PDM increased steadily with younger ages of drinking onsets, and that dependent drinking and marijuana use may be key mediators of that association (7). These studies, however, have not demonstrated to what extent young drinkers reporting PDM might also be at risk for frank alcohol-related problems that may be independent of criteria for alcohol-use disorders, and that may incur highly prevalent and severe personal, societal, and economic costs on teenagers and young adult drinkers (3,5–6, 8–12).

We have conducted cross-sectional analysis of data from the 2001–2002 NESARC to determine whether young drinkers who reported PDM were more likely to report alcohol-related problems than were those not reporting PDM. In our analysis, alcohol-problems were categorized into two domains, alcohol-related "interpersonal troubles," which included items specifying drinking problems that adversely affected family, friends, school, or work, and alcohol-related "risk-taking behaviors," which included items specifying behaviors that might jeopardize personal safety. Our findings indicate that misuse of prescription drugs may be an important factor associated with alcohol problems among younger drinkers, findings that might stimulate longitudinal studies to better establish cause and effect associations.

METHODS

Study Population

The NESARC is designed as a longitudinal survey on topics related to alcohol and drug use, abuse, and dependence and their associated disabilities. Its first wave of interviews was fielded in 2001–2002. The NESARC includes a probability sample of the United States population; 43,093 Americans participated in the first wave of the survey. The target population of the NESARC is the noninstitutionalized household population, 18 years and older residing in the United States. The sampling frame for randomly selected housing units were the "primary sampling units" determined from the supplemental housing survey of the 2000–2001 National Census. Samples of residents in group quarters such as boarding houses, group homes, college quarters; nontransient motels and hotels are also surveyed. African Americans, Hispanics, and young adults were oversampled. Face-to-face interviews were conducted in home or group quarters by survey-experienced and trained interviewers using laptop computer–assisted software. Subjects were not paid for their participation. The overall response rate of the 2001–2002 survey was 81%. The sample was weighted to adjust for probabilities of nonresponse, household size, and oversampling of young adults and data were adjusted to be representative of the U.S. population on a variety of sociodemographic variables (13–15).

Analysis for the current study was limited to 18- to 34-year-old NESARC respondents. This cohort had the highest prevalence of self-reported prescription drug misuse (5–7). Further, recall bias would be less likely from younger than from older respondents and results would be more relevant to current teenagers and young adults. Respondents were categorized as "ever-drinkers" and "life-time abstainers" on the basis of "yes" or "no" responses to the question *"In your entire life, have you had a least 1 drink of any kind of alcohol, not counting tastes or sips?"*

"Ever-drinkers" included both 18- to 34-year-old current-drinkers (*"... at least one drink of any kind of alcohol ... in the prior 12 months"*) and ex-drinkers at the time of the survey. To be consistent with previous age of drinking onset analysis conducted with the National Longitudinal Alcohol Epidemiology Survey (NLAES), analyses focused only on "ever-drinker" respondents who drank 12 or more drinks in at least one year of their life (16).

Measures

Predictor Variables

Prescription drug misuse (PDM) was determined from responses to the to the following survey question:

> *Now I'd like to ask you about your experiences with medicines and other kinds of drugs that you may have used ON YOUR OWN—that is, either WITHOUT a doctor's prescription, in GREATER amounts, MORE OFTEN, or LONGER than prescribed; or for a reason other than a doctor said you should use them. People use these medicines and drugs ON THEIR OWN to feel more alert, to relax or quiet their nerves, to feel better, to enjoy themselves, or to get high or just see how they would work.*

The four targeted psychoactive prescription drugs classes used in the analysis, as they were listed in the survey, were *"Sedatives, for example sleeping pills, barbiturates, Seconal, Qualudes or Chloral Hydrate"; "Tranquilizers or anti-anxiety drugs, for example Valium, Librium, muscle relaxants, Xanax"; "Painkillers, for example Codeine, Darvon, Percodan, Dilaudid or Demoral;* and *"Stimulants, for example Preludin, Benzadrine, Methadrine, uppers or speed."* With any positive response for a drug class, respondents were asked to name the specific agents they used within that class so as to validate the initial response. Responses were analyzed as a dichotomous, "yes/no" variable. We determined the reported onsets of PDM and drinking [reported elsewhere (7)], but did not determine the duration or periods of concurrence for these two behaviors.

Primary Outcome Variables

The primary outcome variables were alcohol-related problems within two derived domains, labeled "risk-taking behaviors" and "interpersonal troubles." Items within each domain were chosen as they conveyed alcohol problems similar in nature. Standardized Cronbach coefficient alphas were .716 for the five items included in "risk-taking behaviors" and .606 for the two items included in "interpersonal problems." "Yes" or "no" responses were recorded for the survey questions in each of the following domain.

> *Alcohol-Related "Risk-Taking Behaviors."* In your entire life did you EVER
>
> *... more than once drive a car or other vehicle WHILE you were drinking?*
> *... more than once drive a car, motorcycle, truck, boat or other vehicle after having too much to drink?*
> *... get into situations while drinking or after drinking that increased your chances of getting hurt... like swimming, using machinery, or walking in a dangerous area or around heavy traffic?*
> *... get into physical fights while drinking or right after drinking?*
> *... get arrested, held at a police station, or have any other legal problems because of your drinking?*
>
> *Alcohol-Related "Interpersonal Troubles."* In your entire life did you EVER
>
> *... have job or school troubles because of your drinking or being sick from drinking? (some examples provided)*
> *... continue to drink even though you knew it was causing you trouble with your family or friends?*

Covariates. Demographic and family history covariates included were categorized as follows:

Sex—male (referent), female;
Age at interview—18–20, 21–34 (referent);
Race—white, not Hispanic or Latino (referent); black, not Hispanic or Latino; Hispanic or Latino; and other;
Family history of alcoholism or drug abuse—positive if respondents reported that first-degree relatives, mother, father, sister, brother, son, or daughter had alcohol or drug abuse problems; negative (referents).

Drinking and recreational drug use covariates were the following:

Drinking age-of-onset—response to "*How old were you when you first started drinking, not counting small tastes and sips*": ≤15 years, 16–17 years, 18–20 years, 21+ years (referent); *Lifetime alcohol dependence*—based on the DSM IV dependence diagnosis variables in NESARC that have been previously validated (13);
Heavy drinking episodes—drank 5 or more drinks at least once/week during period of heaviest drinking;
Cannabis use, ever—lifetime use ("ever use") of cannabis and was determined from positive responses to a direct question, "*Have you ever used any of these medications or drugs: cannabis, THC or grass?*"

These substance-use covariates were included to determine potential confounders and/or mediators of associations between PDM and alcohol-problems.

Data Analysis

The NESARC followed a multistage, probability-sampling plan, in which minorities and younger people were oversampled. Our analyses were conducted using the SUDAAN statistical package to account for the sampling design and survey weights of the NESARC survey (17). Lifetime history of PDM is described through percentages and corresponding standard errors of the mean. Multiple logistic regression models were used to examine the association between reported PDM and reported alcohol-related problems for all "ever-drinkers" and for those who met lifetime criteria for alcohol dependence and those who reported lifetime (ever) use of cannabis. Associations were described through odds ratios. Associations between demographic and behavioral factors and PMD were also examined through these logistic regression models. Associations were considered significant at the $P < .05$ level and expressed as point estimates along with their 95% confidence intervals.

RESULTS

Respondents ages 18 to 34 years old totaled 12,958 (31.5% of total NESARC cohort) men and women, of whom 8306 (64.1%) were classified as "ever-drinkers" and 4652 (35.9%) as "lifetime abstainers." Within the cohort of "ever-drinkers," the prevalence for misusing each of the classes of drugs was as follows: sedatives, 6.3%; tranquilizers, 6.1%; opioid analgesics, 9.4%; stimulants, 6.6%; one or more of the four classes, 15.4%. Among the "lifetime abstainers," only 3.0% reported misuse of one or more of the classes of prescribed drugs; no further analyses were conducted on "lifetime abstainers."

Table 1 presents the weighted proportions of the 8257 respondents who reported PDM by personal demographics and substance use and family histories of alcohol or drug problems. Respondents with the highest prevalence for reporting PDM were those who reported lifetime cannabis use (34.9%), who met NESARC criteria for alcohol dependence (34.0%), who reported episodes of heavy drinking (26.0%), and who reported a history among first-degree family members of either alcoholism (24.0%) or drug abuse (30.4%) and who reported a drinking age-of-onset of 15 years or less (33.1%). Men and women had roughly similar proportions reporting PDM (16.3% and 14.2%, respectively), but white respondents had substantially higher rates of PDM than did black and Hispanic/Latino respondents (17.5%, 6.8%, and 10.9%, respectively).

Respondents who reported PDM were significantly more likely to report specific, alcohol-related problems than were those who did not report PDM (Table 2). Overall, 72.3% of

TABLE 1. Prevalence of Reported PDM Among NESARC "Ever-Drinkers" Ages 18 to 34 by Demographic and Substance Use Variables

Variable (*N*)	"Ever-drinkers" reporting PDM Weighted % (SE)	Variable (*N*)	"Ever-drinkers" reporting PDM Weighted % (SE)
Age at Survey		Family Hx: Alcoholism	
18–20 years (1035)	21.6 (1.60)	Yes (2800)	24.0 (1.18)
21–34 years (7222)	14.4 (0.72)	No (5012)	10.7 (0.65)
Sex		Family Hx: Drug Abuse	
Male (4148)	16.3 (0.93)	Yes (1562)	30.4 (1.64)
Female (4109)	14.2 (0.76)	No (6126)	11.6 (0.59)
Race/Ethnicity		Drinking Age-of-onset	
White (4605)	17.5 (0.85)	≤ 15 years old (1492)	33.1 (1.59)
Black (1296)	6.8 (0.87)	16–17 years old (2177)	17.8 (0.97)
Hispanic/Latino (1933)	10.9 (1.05)	18–20 years old (2819)	9.6 (0.92)
Other (423)	16.8 (2.07)	21+ years old (1707)	5.8 (0.70)
Education Achieved		Heavy Drinking Episodes	
<High school (1076)	19.7 (1.71)	Ever (1911)	34.0 (1.49)
High school (2201)	16.4 (1.09)	Never (6311)	9.0 (0.53)
Some college (2957)	16.8 (1.01)	Alcohol Dependence	
College grad. (2023)	10.3 (0.95)	Ever (1951)	26.0 (1.27)
Smoking Status		Never (6306)	9.0 (0.52)
Current (2978)	26.3 (1.20)	Cannabis Use	
Former (609)	19.0 (1.91)	Ever (2975)	34.9 (1.35)
Never (4624)	7.5 (0.57)	Never (5271)	3.7 (0.32)

TABLE 2. Alcohol-Related "Risk-Taking Behaviors" and "Interpersonal Problems" for NESARC Ever-Drinkers, Ages 18 to 34, by Prescription Drug Misuse (PDM)

Alcohol-related behaviors and problems	"Ever-drinkers" reporting PDM *N* = 1170 % (SE)	"Ever-drinkers" not reporting PDM *N* = 7052 % (SE)
"Risk-Taking Behaviors"		
Drove car more than once while drinking	44.2 (1.68)*	20.8 (.86)*
Drove more than once having drank too much	44.4 (1.79)	20.0 (.70)
Increased chance of injury due to drinking	40.7 (1.81)	14.2 (.56)
Physical fights during or after drinking	34.6 (1.61)	11.6 (.56)
Arrested or legal problems due to drinking	22.6 (1.52)	7.9 (.43)
Any one or more "risk taking behaviors"	72.3 (1.66)	37.2 (.99)
"Interpersonal Troubles"		
Job or school problems due to drinking	19.0 (1.56)	3.0 (.27)
Drank despite problems with family or friends	21.7 (1.49)	4.1 (.32)
Any one or two "interpersonal problems"	28.5 (1.78)	5.9 (.38)

*All differences in prevalence rates comparing items for "ever-drinkers" with PDM versus "ever-drinkers" without PDM were significant at *P* values <.001.

those reporting PDM reported one or more alcohol-related "risk taking behaviors" and 28.5% reported one or more alcohol-related "interpersonal problems." Respondents who reported PDM most commonly reported driving a car while drinking (44.2%), driving more than once when having too much to drink (44.4%), and risking injury due to drinking (40.7%).

We determined the associations of alcohol-related problems with PDM in the entire population of 18- to 34-year-old NESARC "ever-drinkers" ($N = 8257$) and two subsamples,

TABLE 3. Unadjusted and Adjusted Odds Ratios for Alcohol-Related "Risk-Taking Behaviors" and "Interpersonal Troubles" Among NESARC Ever-Drinkers With PDM Compared to Respondents Without PDM for Entire Sample and Alcohol Dependent and Cannabis Using Subsamples, Ages 18 to 34

Alcohol-related problems	Entire 18–34-year-old sample $N = 8257$ OR (95% CI)	Alcohol-dependent (ever) subsample $N = 1951$ OR (95% CI)	Cannabis user (ever) subsample $N = 2975$ OR (95% CI)
"Risk-Taking Behaviors"			
Unadjusted	4.40 (3.72, 5.21)	2.02 (1.45, 2.81)	2.41 (1.96, 2.98)
Adjusted*	1.33 (1.08, 1.63)	1.36 (0.92, 1.99)	1.43 (1.13, 1.80)
"Interpersonal Troubles"			
Unadjusted	6.34 (5.16, 7.81)	2.86 (2.26, 3.62)	3.54 (2.75, 4.56)
Adjusted*	1.82 (1.39, 2.38)	1.96 (1.50, 2.57)	1.71 (1.30, 2.24)

*Adjusted for age, sex, race/ethnicity, education, family history of alcoholism and of drug abuse, smoking status, drinking age-of-onset, "heavy-drinking," alcohol dependence (excluded as a variable in alcohol-dependent subsample), and cannabis use (excluded as a variable in cannabis user subsample).

those who met lifetime criteria for "alcohol dependence" ($n = 1951$) or who reported lifetime marijuana use ($n = 2975$) (Table 3). For the entire cohort, the crude odds for reporting "risk-taking behaviors" and "interpersonal troubles" among those reporting PDM compared to those who did not were 4.4 (95% CI 3.5, 5.2) and 6.3 (95% CI 1.4, 2.4), respectively. In subsamples of respondents with lifetime alcohol dependence or cannabis use, unadjusted odds ratios for reporting one or more alcohol-related "interpersonal problems" and "risk behaviors" remained significantly elevated, roughly two- to fourfold greater for those reporting PDM than for those not reporting PDM. In both the full sample and the two subsamples, the adjusted odds for both domains of alcohol-related problems among drinkers with PDM compared to drinkers without PDM remained increased and, other than the odds associating PDM with "risk-taking behaviors" in the alcohol-dependent subsample (OR = 1.36 [95%CI 0.92, 1.99]), were statistically significant.

In the adjusted analysis comparing drinkers with PDM to drinkers without PDM, the highest increased risks for both alcohol-related "risk-taking behaviors" and "interpersonal troubles" were for respondents meeting lifetime criteria for alcohol dependence, those with drinking age-of-onsets of 15 or younger and for those reporting frequent "heavy drinking" (Table 4). Conversely, the odds of "ever-drinkers" with PDM reporting alcohol-related "risk-taking behaviors" (but not "interpersonal troubles") were substantially reduced compared to those without PDM for respondents ages 18 to 20, women, blacks, and Hispanic/Latinos.

DISCUSSION

The key findings in this analysis of 2001–2002 NESARC data from 18- to 34-year-old respondents is that for those "ever-drinkers" who reported misusing psychoactive drugs (PDM), the odds for reporting alcohol-related problems within each of two domains, "risk-taking behaviors" and "interpersonal troubles," were significantly higher than those for "ever-drinkers" who did not report PDM. These elevated odds were lowered by about threefold, but remained statistically significant, when adjusted for demographics, family histories of alcohol or drug problems, measures of drinking age-of-onset, heavy or dependent drinking, and cannabis use. Among a subset of lifetime "ever-drinkers" who also reported lifetime cannabis use (36% of "ever-drinkers") compared to those not reporting cannabis use, the adjusted odds remained approximately 1.5-fold increased for both domains of alcohol-related problems. For the subsample of "ever-drinkers" meeting alcohol-dependence criteria (23% of "ever-drinkers") compared to drinkers not meeting dependence criteria, the adjusted odds for reporting alcohol-related

TABLE 4. Odds Ratios for Reporting Alcohol-Related "Risk-Taking Behaviors" and "Interpersonal Problems" in NESARC "Ever-Drinkers," Ages 18 to 34, Reporting Prescription Drug Misuse (PDM) Compared to Those Not Reporting PDM

Variables	*N*	Alcohol-Related "Risk-Taking Behaviors" OR (95% CI)	Alcohol-Related "Interpersonal Troubles" OR (95% CI)
PDM, ever			
Yes	1170	1.33 (1.08, 1.63)*	1.82 (1.39, 2.38)*
No		REF	REF
Age, years			
18–20	1035	0.56 (0.44, 0.72)*	0.85 (0.61, 1.18)
21–34	7222	REF	REF
Sex			
Male	4148	REF	REF
Female	4109	0.57 (0.49, 0.66)*	1.03 (0.78, 1.35)
Race/ethnicity			
White	4605	REF	REF
Black	1296	0.66 (0.55, 0.80)*	1.15 (0.79, 1.67)
Hispanic/Latino	1933	0.76 (0.63, 0.92)*	1.68 (1.26, 2.25)*
Family Hx. alcoholism			
Yes	2800	1.47 (1.26, 1.71)*	2.53 (1.89, 3.38)*
No	5012	REF	REF
Age of drinking onset			
≤ 15	1492	3.54 (2.77, 4.53)*	3.21 (1.86, 5.55)*
16–17	2177	3.17 (2.56, 3.92)*	1.54 (0.89, 2.67)
18–20	2819	1.93 (1.55, 2.39)*	1.27 (0.75, 2.14)
21+	1707	REF	REF
Heavy drinking period			
Ever	1191	3.19 (2.77, 3.68)*	3.27 (2.43, 4.39)*
Never	6311	REF	REF
Alcohol dependence			
Ever	1915	4.48 (3.77, 5.34)*	5.79 (4.41, 7.61)*
Never	6306	REF	REF
Cannabis use			
Ever	2975	1.99 (1.73, 2.29)*	1.35. (1.05, 1.73)*
Never	5271	REF	REF

Other variables entered with nonsignificant adjusted odds ratios: Education, Smoking Status, and Family History of Drug Abuse.
REF = Reference group.
*Significant at $P < .05$.

"interpersonal troubles" were approximately twofold increased.

We have not determined in this analysis to what extent periods of problematic drinking preceded, coincided with or followed periods of PDM, but data from NESARC and other studies would suggest that these behaviors typically occurred concurrently (5–9). Among this 18- to 34-year-old NESARC cohort of drinkers who reported PDM, we have previously reported that 80% identified their onsets of drinking to be prior to the onset of PDM and 20% reported PDM either earlier than or at the same time as when first starting to drink (7). Huang et al. (5), for the entire NESARC population, found that the peak hazard rates for nonmedical use of these agents was between ages 17 and 18, with mean duration of use between 2.5 and 3.2 years. McCabe et al. (6) reported that more than one in four 18- to 24-year-old NESARC respondents meeting alcohol dependence criteria reported concurrent PDM. As alcohol use is most likely to persist during these late teen and young adult periods (8,9), PDM and alcohol-related problems, for a great many within this cohort, are most likely occurring concurrently. Forthcoming longitudinal data from the second NESARC should be exceptionally valuable in establishing both the

time sequences and consequences of concurrent alcohol and drug use.

Despite the limitations imposed by our cross-sectional analysis, there is ample evidence that heavy alcohol use and misuse of prescription drugs are closely linked behaviors. McCabe et al. (5) reported a strong co-occurrence of alcohol use disorders and nonmedical use of prescription drugs in 18- to 24-year-old NESARC respondents. Additionally, McCabe et al. (15) reported from a 2001 college survey that the odds for recent high-risk drinking and driving behaviors were from three- to sixfold increased in students reporting nonmedical use of prescription opioid analgesics. Data from these studies indicate that both heavy drinking or dependent drinking and misuse of psychoactive prescription drugs are common, concurrent behaviors among late teens and young adults. These findings support the observation of others and the broad construct that multiple, problematic substance-use behaviors among teenagers and young adults are likely to occur concurrently or in close sequence (16–19).

In that the NESARC relied on recall of events and their causes, our findings are limited by potentially incorrect attribution of problems to either alcohol or drugs by NESARC respondents. Subjects may have readily misattributed behavioral and interpersonal problems to alcohol that were caused either totally or in part by use of prescription or "street" drugs; as well, the converse is also possible. Further, the extent to which misuse of prescription drugs may have been in response to alcohol problems, rather than contributing to their occurrence, was not determined. Longitudinal data from the NESARC in coming years will be critical in defining more precisely sequential or concurrent behaviors and in assigning attribution of specific substance-related problems.

Alcohol-related problems among teenagers and young adults in the United States are prevalent and costly (6–9). Our results indicate that that among young drinkers, misuse of prescription drugs, whether by heavy or dependent drinkers, and even among recreational cannabis users, may be an important risk factor for increasing alcohol problems. To what extent the problems identified in this study that were ascribed to drinking resulting from excessive alcohol use, from drug misuse, or from both has not been fully determined. However, the current data clearly indicate that comprehensive prevention, treatment, and harm-reduction efforts need to target the variety of substance use behaviors common among teenagers and young adults.

REFERENCES

1. National Institute on Drug Abuse Research Report Series. Prescription Drugs: Abuse and Addiction. 2001. www.nida.nih.gov/Research/Prescription/prescription5.html.
2. Volkow ND. NIDA responds to changing drug abuse patterns. NIDA Notes, National Institute on Drug Abuse 2005;20(2):3.
3. National Institute on Drug Abuse. NIDA InfoFacts: High School and Youth Trends. 2006. www.drugabuse.gov/infofacts/HSYouthtrends.htlm.
4. Stinson FS, Grant BF, Dawson DA, Ruan WJ, Huang B, Saha T. Comorbidity between DSM-IV alcohol and specific drug use disorders in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Drug Alcohol Depend.* 2005;80:105–16.
5. 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 Epidemiological Survey on Alcohol and Related Conditions. *J Clin Psychiatry.* 2006;67:1062–1073.
6. McCabe SE, Cranford JA, Boyd CJ. The relationship between past-year drinking behaviors and nonmedical use of prescription drugs: prevalence of co-occurrence in a national sample. *Drug Alchol Depend.* 2006;84:281–288.
7. Hermos J, Winter M, Heeren T, Hingson R. Early Age-of-onset drinking predicts prescription Drug misuse among teenagers and young adults: results from a national survey. *J Addict Med.* 2008;2:22–30.
8. Chen K, Kandel DB. The natural history of drug use from adolescence to the mid-thirties in a general population sample. *Am J Public Health.* 1995;85:41–47.
9. Kandel DB, Chen K, Warner LA, Kessler RC, Grant B. Prevalence and demographic correlates of symptoms of last year dependence on alcohol, nicotine, marijuana and cocaine in the U.S. population. *Drug Alcohol Depend.* 1999;44:11–29.
10. Hingson R, Heeren T, Jamanka A, Howland J. Age of drinking onset and unintentional injury involvement after drinking. *JAMA.* 2000;284:1527–1533.
11. Hingson R, Heeren T, Zakocs R. Age of drinking onset and involvement in physical fights. *Pediatrics.* 2001;108:872–877.

12. Foster SE, Vaughan RD, Foster WH, Califano JA. Alcohol consumption and expenditures for underage drinking and adult excessive drinking. *JAMA.* 2003;289: 989–995.

13. Grant BF, Stinson FS, Dawson DA, Chou SP, Dufour MC, Compton W, Pickering RP, Kaplan K. Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry.* 2004;61:807–816.

14. Hasin DS, Grant BF. The co-occurrence of DSM-IV alcohol abuse in DSM-IV alcohol dependence: results of the National Epidemiologic Survey on Alcohol and Related Conditions on heterogeneity that differ by population subgroup. *Arch Gen Psychiatry.* 2004;61:891–896.

15. Grant BF, Moore TC, Shepard J, Kaplan K. Source and Accuracy Statement: Wave 1 National Epidemiological Survey on Alcohol and Related Conditions (NESARC). National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD. Available at http://miaaa.census.gov/ (accessed March 6, 2008).

16. Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression an psychiatric diagnostic modules in a general population sample. *Drug Alcohol Depend.* 2003;17:7–16.

17. Research Triangle Institute. Software for Survey Data Analysis (SUDAAN) Version 8.1. Research Triangle Park, NC: Research Triangle Institute; 2002.

18. McCabe SE, Teter CJ, Boyd CJ, Knight JR, Wechsler H. Nonmedical use of prescription opoiods among U.S. college students: prevalence and correlates from a national survey. *Addict Behav.* 2005;30:789–805.

19. DuRant RH, Smith JA, Kreiter SR, Krowchuk DP. The relationship between early age of onset of initial substance use and engaging in multiple health risk behaviors among young adolescents. *Arch Pediatr Adolesc Med.* 1999;153:286–291.

20. Clark DB. The natural history of adolescent alcohol use disorders. *Addiction.* 2004;99:5–22

21. Merline AC, O'Malley PM, Schulenberg JE, Bachman JG, Johnston LD. Substance use among adults 35 years of age: prevalence, adulthood predictors, and impact of adolescent substance use. *Am J Public Health.* 2004;94:96–102.

22. Willoughby T, Chalmers H, Busseri MA. Where is the syndrome? Examining co-occurrence among multiple problem behaviors in adolescence. *J Consult Clin Psychol.* 2004;72:1022–1037.

Patient Education and Counseling 75 (2009) 376–380

Contents lists available at ScienceDirect

Patient Education and Counseling

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

Caregiver health literacy and adherence to a daily multi-vitamin with iron regimen in infants

L. Kari Hironaka [a,*], Michael K. Paasche-Orlow [b], Robin L. Young [c], Howard Bauchner [a], Paul L. Geltman [a,d]

[a] Department of Pediatrics, Boston Medical Center/Boston University School of Medicine, Boston, MA 02118, USA
[b] Department of Medicine, Boston Medical Center/Boston University School of Medicine, Boston, MA, USA
[c] Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
[d] Department of Pediatrics, Cambridge Health Alliance, Cambridge, MA, USA

ARTICLE INFO

Article history:
Received 23 September 2008
Received in revised form 9 March 2009
Accepted 18 March 2009

Keywords:
Health literacy
Medication adherence
Iron deficiency

ABSTRACT

Objective: To determine whether or not limited caregiver health literacy is associated with adherence to a daily multi-vitamin with iron regimen in infants.
Methods: 110 caregiver/infant dyads were enrolled in a prospective study to assess the relationship between caregiver health literacy and adherence to a daily multi-vitamin with iron regimen for infants. Households were contacted biweekly over a 3-month period. Adherence was based upon caregiver report. High adherence, our primary outcome, was defined as the administration of the multi-vitamin with iron on 5–7 days over the past week.
Results: As measured by the Short Test of Functional Health Literacy in Adults (S-TOFHLA), 18% of caregivers had limited health literacy skills. Caregivers with limited health literacy skills were more likely to have higher adherence than caregivers with adequate health literacy, after adjusting for a number of possible confounding variables (AOR = 2.13; 95% 1.20–3.78).
Conclusion: Caregivers with limited health literacy were twice as likely to report high adherence to a daily multi-vitamin with iron regimen in infants as caregivers with adequate health literacy in adjusted analysis.
Practice implications: Health literacy may exert a differential influence on adherence depending upon the complexity of the desired health behavior.

1. Introduction

Over the past decade, there has been growing interest in the relationship between literacy and health outcomes. Health literacy is defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" [1]. Limited health literacy has been associated with poverty, limited education, minority status, immigration, and older age [2]. In a nationally representative household survey of 19,000 adults, it was estimated that 36% of the U.S. adult population have limited health literacy skills [2].

A 2004 report by the Institute of Medicine (IOM) concluded that adults with limited health literacy have less knowledge of disease self-management and health-promoting behaviors, report poorer health status, are less likely to use preventive services, are more likely to be hospitalized, and have worse health outcomes for certain chronic conditions [1]. It has been hypothesized that the relationship between health literacy and medication adherence may mediate the effect on specific health outcomes. However, several adult studies assessing the relationship between health literacy and adherence have produced varied results [3–9]. Some studies suggest worse adherence among individuals with limited health literacy [9]; others report no association [7] or even better adherence [4].

Few studies have assessed the relationship between caregiver health literacy and child health outcomes. Limited caregiver literacy skills have been associated with poorer pediatric outcomes in asthma and diabetes [10,11]; however, few data are available regarding caregiver health literacy and medication adherence in children. As part of a larger randomized controlled trial, we had the opportunity to assess the impact of caregiver health literacy on adherence to a daily multi-vitamin with iron regimen in infants.

Reducing iron deficiency among vulnerable populations is one of the objectives of Healthy People 2010 [12], with toddlers being one of the sub-populations of particular concern. Based upon

Abbreviations: GEE, generalized estimating equations; IOM, Institute of Medicine; S-TOFHLA, Short Test of Functional Health Literacy in Adults.

* Corresponding author at: Boston Medical Center, 88 East Newton Street, Vose 4, Boston, MA 02118, USA. Tel.: +1 617 414 7941; fax: +1 617 414 7915.
E-mail address: lynn.hironaka@bmc.org (L.K. Hironaka).

0738-3991/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.pec.2009.03.016

NHANES IV data (1999–2002), the prevalence of iron deficiency among U.S. children aged 1–3 years is estimated to be 8% [13]. Children in lower socioeconomic and minority groups have disproportionately high rates of iron deficiency [14], with some studies reporting rates of almost 20% [15,16]. Iron deficiency anemia has been associated with lower cognitive test scores in numerous studies [17,18] and several studies have demonstrated a persistence of poorer cognitive outcomes at school-age and adolescence [17,19,20].

Both limited health literacy and iron deficiency are associated with many of the same risk factors such as poverty, lower parental education, and minority status [2,17]. Iron deficiency anemia in infants has been found to be negatively associated with both level of maternal knowledge of anemia and adherence to iron supplementation [21]; this provides theoretical support for an association between caregiver health literacy and adherence to iron supplementation in infants. This study sought to determine whether or not limited caregiver health literacy was associated with adherence to a daily multi-vitamin with iron regimen in infants. It was hypothesized that infants of caregivers with limited health literacy skills would demonstrate poorer adherence to a daily multi-vitamin with iron supplement than infants of caregivers with adequate health literacy skills.

2. Methods

This study assessed the association between caregiver health literacy and adherence to a daily multi-vitamin with iron regimen in infants as part of a randomized controlled trial designed to compare levels of adherence among two different multi-vitamin with iron formulations. Adherence was based upon caregiver report and was collected bi-weekly over a 3-month study period. At each follow-up visit, caregivers were asked, "*On how many days in the past week did you give (child's name) the drops/sprinkles*?"

Approval and monitoring for this study was provided through the Boston University Medical Center Institutional Review Board. Caregivers provided written informed consent for themselves and their infant.

2.1. Sample

As part of the larger randomized controlled trial, caregivers of healthy infants, aged 5–7 months, were recruited during their 6-month well child visits at two urban pediatric primary care clinics. Caregiver/infant dyads were randomized to receive one of two formulations of a daily multi-vitamin with iron supplement – either a commonly used liquid formulation or a "sprinkles" formulation which is mixed into a baby's solid food. Interviews were conducted in English or Spanish, depending on caregiver preference. Exclusion criteria included a history of conditions associated with iron deficiency or anemia, use of vitamin or iron supplements within the month prior to enrollment, prematurity, multiple gestations, or low birthweight (<2500 g).

Enrollment for the prospective trial began in June 2005 and continued until March 2006. Of the 150 families who were enrolled in the randomized controlled trial, 40 did not complete the health literacy screen; of these 14 (35%) had completed the study prior to June 2005 when literacy screening was introduced, 20 (50%) were lost to follow-up or withdrew prior to completing the health literacy screen, and six (15%) had missing health literacy data. The cohort for the current analysis included data from the remaining 110 families (Fig. 1).

There were no significant differences between the 110 caregivers who completed the health literacy screen and 40 who did not on the following caregiver variables: age, gender, education level, race/ethnicity, and child's health insurance type. However, compared to caregivers who did not complete the health literacy screen, those who completed the screen were more likely to be born outside of the U.S. (66% vs. 44%; $p = 0.01$).

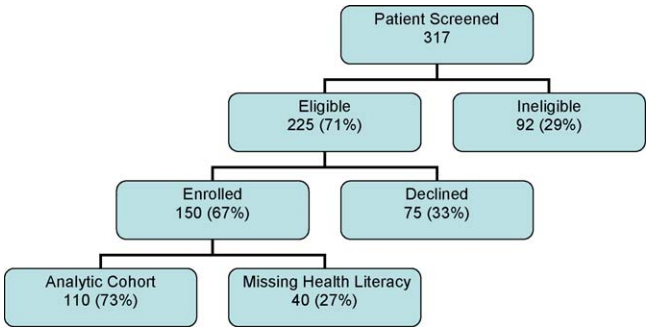


Fig. 1. Flowchart for study enrollment.

2.2. Data collection

Baseline data including demographic information, dietary history, and anthropomorphic measurements were collected via a closed-ended orally administered survey. Bilingual research assistants also provided caregivers with oral and written instructions on how to administer the multi-vitamin with iron supplement. Written instructions were provided in English and Spanish and were written at less than a 5th grade reading level.

Research assistants contacted caregivers on a biweekly basis over a 3-month period between their child's 6-month and 9-month well-child visits, for a total of six possible assessment points. Contact with caregivers alternated between telephone calls and home visits. A closed-ended survey was orally administered to the caregiver at each point of contact and included questions regarding the infant's adherence to the multi-vitamin and iron regimen, ease of administration, and side effects. Attempts were made to collect all survey information from a single caregiver, but this was not always possible. The health literacy screen was administered in-person during one of the home visits or at the 9-month well child check-up, depending on when the research assistants had more time. Families received an additional supply of drops or sprinkles packets during the home visits. A closed-ended survey was administered to caregivers at the end of the study to inquire about any concerns regarding the use of the multi-vitamin with iron supplements.

2.3. Measures

Health literacy, the primary independent variable, was measured using the Short Test of Functional Health Literacy in Adults (S-TOFHLA), a 36-item reading comprehension test [22]. The S-TOFHLA consists of two medically related passages with keywords missing. Utilizing a modified Cloze procedure, subjects select the appropriate word for each omitted word from a list of four choices. The S-TOHFLA has good internal consistency (Cronbach's $\alpha = 0.97$) and is well correlated with the Rapid Estimate of Adult Literacy in Medicine (Spearman correlation 0.81) and the full Test of Functional Health Literacy in Adults (Spearman correlation 0.91), two other measures of health literacy [23]. It is available in English and Spanish. S-TOFHLA scores are categorized as "adequate", "marginal", and "inadequate", based upon standardized cut-off values [22,23]. Prior to analysis, *limited health literacy* was defined as scores in the "marginal" or "inadequate" range; *adequate health literacy* was defined as scores in the "adequate" range.

Adherence, our primary outcome, was measured by caregiver report and defined as high adherence (yes vs. no) for administration of the multi-vitamin with iron supplement on 5–7 days of the preceding week [24].

Other independent covariates were evaluated for inclusion in multivariate models based upon theoretical evidence of possible confounding. These included caregiver age (years), caregiver gender (female or male), caregiver country of birth (US or non-US), language of S-TOFHLA administration (English or Spanish), child's health insurance (public or private), child's race/ethnicity (Black, Hispanic, White/other), duration of time in the study (interview number), caregiver concerns regarding the multi-vitamins (yes or no), side effects (dark stools, constipation, stained teeth, diarrhea, vomiting, or abdominal pain), and randomized assignment to drops or sprinkles.

2.4. Analyses

Bivariate chi-square tests of independence and differences of means *t*-tests were utilized to compare socio-demographic variables with caregiver health literacy levels. As each family contributed between one and six assessments of adherence, generalized estimating equations (GEE) with an exchangeable variance-covariance matrix was used in all analyses of adherence [25]. GEE regression adjusts for a variable number of observations per subject as well as for the correlation between observations contributed by a single subject. First, to evaluate the association between health literacy and high adherence, a simple GEE logistic regression model was performed. An evaluation of a potential interaction between health literacy and interview number was not found to be significant, so an interaction term was not included in the following model.

A GEE multiple logistic regression was used to evaluate the association between high adherence and caregiver health literacy, adjusted for possible confounders. Caregiver age, gender, language of S-TOFHLA administration, insurance type, and interview number were excluded from the final model because they were determined not to be confounders. Caregiver country of birth was highly concordant with health literacy and was not included in the final model. The final model contained the following variables: health literacy, randomization assignment, race/ethnicity, and, caregiver education. Additional analyses were then performed to assess the possibility of confounding by caregiver concerns regarding the multi-vitamin supplements (yes vs. no) and by the presence of specific side effects.

Although attempts were made to collect all follow-up interview data from the primary caregiver enrolled in the study, this was not always feasible given the involvement of multiple caregivers. Of the 110 families enrolled in the study, multiple caregivers provided follow-up information in 11 families (10%). A sensitivity analysis was performed on the multiple logistic regression model limiting our sample to data collected from the caregiver who completed the baseline data collection and S-TOFHLA. This restricted analysis included 394 of the 574 observations.

All statistical tests for association were performed at the 0.05 significance level. Statistical analyses were performed using SAS, version 9.1 (SAS Institute Inc., Cary, NC, 2002–2003).

3. Results

Limited health literacy was observed in 20 of 110 caregivers (18%). Over 90% of the caregivers were women. For the limited health literacy group, 60% completed all six assessment points; for the adequate health literacy group, 70% completed all six assessment points. There was no statistically significant difference in the number of missed assessment points between the two health literacy groups ($p = 0.29$).

Table 1
Sociodemographic characteristics according to health literacy level.

Sociodemographic variables	Total ($n = 110$)	Adequate health literacy ($n = 90$)	Limited health literacy ($n = 20$)	*P*-values
Caregiver age (years)	30.2 (6.55)	30.1 (6.67)	30.2 (6.17)	0.97
Caregiver female	91.8%	91.1%	95.0%	0.57
Caregiver with less than High School Degree	17.3%	15.6%	25.0%	0.31
Caregiver born outside of US	66.4%	61.1%	90.0%	0.01
Spanish S-TOFHLA	14.6%	14.4%	15.0%	0.95
Health insurance – public	86.4%	87.8%	80.0%	0.36
Sprinkles	48.2%	50.0%	40.0%	0.42
Child's race				0.76
Black	48.2%	46.7%	55.0%	
Hispanic	30.0%	32.2%	20.0%	
Other	17.3%	16.7%	20.0%	
White	4.6%	4.4%	5.0%	

Caregivers with limited health literacy were more likely to be born outside of the U.S. than caregivers with adequate health literacy (90% vs. 61%, respectively, $p = 0.01$). Caregiver health literacy had no statistically significant associations with caregiver age, gender, language of S-TOFHLA administration, type of health insurance, race/ethnicity, or randomization assignment (drops vs. sprinkles) (Table 1).

High adherence was reported in 32.7% (188) of the 574 observations. Contrary to the study hypothesis, caregivers with limited health literacy were more likely to report high adherence than caregivers with adequate health literacy on unadjusted analysis (OR = 2.18, 95% CI 1.23–3.88). Caregivers with limited health literacy reported using multi-vitamins with iron on an average of 3.7 days per week compared to caregivers with adequate health literacy who reported using them on an average of 2.4 days per week. Caregivers with limited health literacy consistently reported higher adherence at each assessment point (Fig. 2).

Results from the multiple logistic regression (Table 2) continued to demonstrate increased odds of high adherence among caregivers with limited health literacy skills (AOR = 2.13; 95% 1.20–3.78). The sensitivity analysis revealed a similar trend in all results. There were no differences in terms of race/ethnicity, health literacy, education, being born in the US, randomization assignment, or insurance type between this subgroup and the larger group including data from multiple caregivers. The odds of high adherence among caregivers with limited health literacy skills was somewhat reduced (AOR = 1.88; 95% 0.96–3.67).

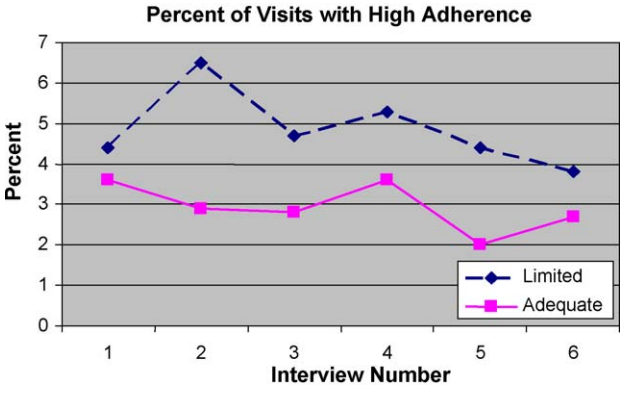


Fig. 2. Percent of visits with high adherence by caregiver literacy.

Table 2
Adjusted odds ratios for high adherence to multi-vitamins and iron.

Variables	Adjusted odds ratio	95% CI	*P*-values
Limited health literacy vs. adequate health literacy	2.13	1.20–3.78	0.01
Randomization: drops vs. sprinkles	1.35	0.82–2.21	0.23
Hispanic vs. White/other	1.03	0.53–2.01	0.93
Black vs. White/other	1.00	0.52–1.94	0.99
Caregiver with less than High School Degree	0.89	0.50–1.59	0.69

To determine whether side effects confounded the relationship between health literacy and adherence, each side effect covariate was added to the model. There were no significant changes in the AOR for health literacy. A similar analysis was performed to determine whether caregiver's concerns about the multi-vitamin with iron affected the results. When caregiver concern was added to the model, the magnitude of the association between limited health literacy and high adherence increased (AOR = 2.4, 95% CI 1.37–4.20).

4. Discussion and conclusion

4.1. Discussion

This study evaluated the relationship between parental health literacy and adherence of infants to a daily multi-vitamin with iron regimen. In a cohort of 110 participants, 18% of participating caregivers demonstrated limited health literacy skills. Contrary to the study hypothesis, in adjusted analysis, caregivers with limited health literacy were twice as likely to report high adherence to a daily multi-vitamin with iron regimen in infants as caregivers with adequate health literacy skills (AOR = 2.13; 95% 1.20–3.78). This relationship was maintained when variables for side effects and concerns about the multi-vitamin with iron were included in the model.

This study adds to the literature by assessing the relationship between caregiver health literacy and medication adherence in pediatric patients. Although previous studies have evaluated the role of caregiver health literacy and pediatric outcomes in the areas of asthma [7], Type 1 diabetes [8], child health care utilization [26], and adherence to recommendations following an acute care visit [27], few studies have assessed the relationship between caregiver health literacy and pediatric medication adherence.

There are several limitations to this study. First of all, adherence was based upon caregiver report of multi-vitamin use over the past 7 days. There was not another measurement or biomarker for multi-vitamin with iron intake. In a sample with higher literacy skills, self-report via questionnaires or diaries may have yielded improved results [28]. However, given the focus on health literacy, it was necessary for all caregiver reported data to be collected via interview. An additional limitation relates to the fact that in 11 of the 110 families, more than one caregiver provided data on adherence. To evaluate the possibility that this introduced bias, we conducted a sensitivity analysis, which was consistent with our main findings. Moreover, as part of the study protocol, caregivers were asked biweekly about adherence to a multi-vitamin with iron regimen. The most likely effect of frequent contact by researchers would be an overall increase in adherence and reporting of adherence. While unlikely, it is possible that frequent contact would differentially influence caregivers with limited health literacy. It is also possible that weakly associated covariates in combination could exert a significant impact on our findings. However, based upon our analyses, there was not adequate justification to include other covariates in the model. The presence of unmeasured confounding could have also affected the results. Finally the S-TOFHLA, the measurement of health literacy skills, assessed reading comprehension but did not assess numeracy skills. Use of a more comprehensive measure of health literacy might have revealed different results.

The overall rate of limited health literacy in this study was significantly lower than other estimates [2,29]. In contrast to results from the 2003 National Assessment of Adult Literacy (NAAL) and other studies, this study observed no association between health literacy and demographic variables such as the level of caregiver education or type of health insurance [1,2,29]. The lack of statistically significant associations may have been due to sample size limitations or lack of variability within a possible confounding variable.

These results must be viewed in the context of a growing literature assessing the complex relationship between health literacy, medication adherence, and disease outcomes. Studies of medication adherence have produced variable results [3–9]. It has been suggested that the relationship between health literacy and adherence may vary depending upon the complexity of the treatment regimen, the amount of self-monitoring required, and even the degree of disease severity [30]. In this study, the daily administration of vitamins was relatively simple. Factors such as self-efficacy and social stigma have also been shown to mediate the relationship between health literacy and medication adherence [5,31]. At a systemic level, the complexities of the medical system, the culture of medical care, and the growing literacy-based demands placed on a patient may have substantial effects on the individual or family with limited health literacy skills – the impact of which could be mitigated by a health care system that is designed to promote patient education and self-care management [32,33].

Overall, relatively few caregivers reported high adherence regardless of their health literacy status. As important as it is to consider the factors that promoted improved adherence in the limited health literacy group, it is equally important to consider which factors contributed to poorer adherence in the adequate health literacy group. In this example, it is possible that improving adherence will be less dependent upon leveling the field between those with limited and adequate health literacy than on utilizing a different approach to promoting health behavior change.

4.2. Conclusions

The relationship between health literacy and medication adherence is likely complex. Limited literacy often coexists with other social vulnerabilities [33], and the relationships between these various social vulnerabilities may explain some of the variability represented in the health literacy and medication adherence research. As health literacy research moves forward, it will be important to gain a greater understanding of these complex interrelationships and how to target interventions at both the individual and community level to improve health outcomes [33].

4.3. Practice implications

Health literacy may exert a differential influence on adherence depending upon the complexity of the desired health behavior. Future interventions to promote adherence should consider the critical points where health literacy may influence these desired behaviors.

Conflict of interest

The authors of this study have no conflicts of interest to disclose.

Role of funding: The larger randomized controlled trial was funded by the U.S. Centers for Disease Control and Prevention through a cooperative agreement with the American Association of Medical Colleges (AAMC-MM 0835 05-05, Geltman PI). The funding source had no involvement with the conduct of the research or preparation of the article for submission.

References

[1] Nielsen-Bohlman LT, Panzer A, Kindig D. Health literacy – a prescription to end confusion. Washington, DC: The National Academies Press; 2004.
[2] 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). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
[3] 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.
[4] Paasche-Orlow MK, Cheng DM, Palepu A, Meli S, Faber V, Samet JH. Health literacy, antiretroviral adherence, and HIV-RNA suppression: a longitudinal perspective. J Gen Intern Med 2006;21:835–40.
[5] Wolf MS, Davis TC, Osborn CY, Skripkauskas S, Bennett CL, Makoul G. Literacy, self-efficacy, and HIV medication adherence. Patient Educ Couns 2007;65:253–60.
[6] Kalichman SC, Catz S, Ramachandran B. Barriers to HIV/AIDS treatment and treatment adherence among African-American adults with disadvantaged education. J Natl Med Assoc 1999;91:439–46.
[7] 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.
[8] Dewalt DA, Pignone M. Health literacy and health outcomes: overview of the literature. In: Schwartzberg J, VanGeest J, Wang C, editors. Understanding health literacy – implications for medicine and public health. American Medical Association; 2005. p. 205–28.
[9] Kalichman SC, Pope H, White D, Cherry C, Amaral CM, Swetzes C, et al. The association between health literacy and HIV treatment Adherence: further evidence from objectively measured medication adherence. J Int Assoc Physicians AIDS Care (Chicago Illinois) 2008;7:317–23.
[10] DeWalt DA, Dilling MH, Rosenthal MS, Pignone MP. Low parental literacy is associated with worse asthma care measures in children. Ambul Pediatr 2007;7:25–31.
[11] Ross LA, Frier BM, Kelnar CJ, Deary IJ. Child and parental mental ability and glycaemic control in children with Type 1 diabetes. Diab Med 2001;18:364–9.
[12] Healthy People 2010: Understanding and Improving Health: US Department of Health and Human Services; 2000.
[13] Brotanek JM, Gosz J, Weitzman M, Flores G. Iron deficiency in early childhood in the United States: risk factors and racial/ethnic disparities. Pediatrics 2007;120:568–75.
[14] Iron deficiency – United States, 1999–2000. Morb Mortal Wkly Rep 2002;51:897–9.
[15] Eden AN, Mir MA. Iron deficiency in 1- to 3-year-old children. A pediatric failure? Arch Pediatr Adolesc Med 1997;151:986–8.
[16] Brugnara C, Zurakowski D, DiCanzio J, Boyd T, Platt O. Reticulocyte hemoglobin content to diagnose iron deficiency in children. J Am Med Assoc 1999;281:2225–30.
[17] Grantham-McGregor S, Ani C. A review of studies on the effect of iron deficiency on cognitive development in children. J Nutr 2001;131:649S–66S (discussion 666S–668S).
[18] McCann JC, Ames BN. An overview of evidence for a causal relation between iron deficiency during development and deficits in cognitive or behavioral function. Am J Clin Nutr 2007;85:931–45.
[19] Lozoff B, Jimenez E, Hagen J, Mollen E, Wolf AW. Poorer behavioral and developmental outcome more than 10 years after treatment for iron deficiency in infancy. Pediatrics 2000;105:E51.
[20] Lozoff B, Jimenez E, Smith JB. Double burden of iron deficiency in infancy and low socioeconomic status: a longitudinal analysis of cognitive test scores to age 19 years. Arch Pediatr Adolesc Med 2006;160:1108–13.
[21] Bilenko N, Yehiel M, Inbar Y, Gazala E. The association between anemia in infants, and maternal knowledge and adherence to iron supplementation in southern Israel. Isr Med Assoc J 2007;9:521–4.
[22] Nurss J, Parker R, Williams MV, Baker D. Test of functional health literacy in adults. Emory University. Peppercorn Books and Press, Inc.; 1995.
[23] Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999;38:33–42.
[24] Lozano P, Finkelstein JA, Hecht J, Shulruff R, Weiss KB. Asthma medication use and disease burden in children in a primary care population. Arch Pediatr Adolesc Med 2003;157:81–8.
[25] Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30.
[26] Sanders LM, Thompson VT, Wilkinson JD. Caregiver health literacy and the use of child health services. Pediatrics 2007;119:e86–92.
[27] Moon RY, Cheng TL, Patel KM, Baumhaft K, Scheidt PC. Parental literacy level and understanding of medical information. Pediatrics 1998;102:e25.
[28] Garber MC, Nau DP, Erickson SR, Aikens JE, Lawrence JB. The concordance of self-report with other measures of medication adherence: a summary of the literature. Med Care 2004;42:649–52.
[29] 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.
[30] Powers BJ, Bosworth HB. Revisiting literacy and adherence: future clinical and research directions. J Gen Intern Med 2006;21:1341–2.
[31] Waite KR, Paasche-Orlow M, Rintamaki LS, Davis TC, Wolf MS. Literacy, social stigma, and HIV medication adherence. J Gen Intern Med 2008;23:1367–72.
[32] Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31:S19–26.
[33] Paasche-Orlow MK, Schillinger D, Greene SM, Wagner EH. How health care systems can begin to address the challenge of limited literacy. J Gen Intern Med 2006;21:884–7.

Antithrombotic Prophylaxis in Elderly Patients with Atrial Fibrillation

Elaine M. Hylek, M.D., M.P.H.[1]

ABSTRACT

The burden of atrial fibrillation (AF) worldwide is projected to increase substantially over the next few decades in part due to an aging population. AF increases the risk of stroke approximately fivefold. The population-attributable risk for stroke by age is considerable: 1.5% for those individuals 50 to 59 years of age compared with 23.5% for those ≥80 years of age. Vitamin K antagonists (VKAs) like warfarin have been shown to greatly reduce the risk of stroke. However, despite their proven efficacy, VKAs remain underused, particularly among elderly patients with AF. The preponderance of evidence from randomized trials and observational studies attests to higher bleeding rates among elderly individuals with AF. Antiplatelet therapy is not effective for stroke prevention in AF and is also associated with significant bleeding risk. Strategies to optimize the effectiveness of VKAs and improve their safety profiles among elderly patients in clinical practice are direly needed. An understanding of the pathological changes that predispose to hemorrhage, hazards of polypharmacy, and factors that contribute to variability in dose response will facilitate a more informed use of these medications in clinical care.

KEYWORDS: Atrial fibrillation, stroke, anticoagulant therapy, vitamin K antagonist, warfarin

The prevalence of atrial fibrillation (AF) is projected to increase markedly over the next few decades. In the United States, the prevalence of AF is expected to exceed 10 million individuals by 2030.[1] Older age is a potent risk factor for AF. Approximately 10% of individuals ≥80 years of age have AF.[2] The morbidity and mortality related to AF is well recognized. In aggregate, AF increases the risk of stroke approximately fivefold. The population-attributable risk for stroke by age is considerable: 1.5% for those individuals 50 to 59 years of age compared with 23.5% for those ≥80 years of age.[3,4] Vitamin K antagonists (VKAs) have been shown to greatly reduce the risk of stroke.[5] However, despite their proven efficacy, VKAs remain underused, particularly among elderly patients with AF.

WARFARIN USE AMONG ELDERLY INDIVIDUALS WITH ATRIAL FIBRILLATION

Multiple studies attest to the underuse of VKA among elderly individuals with AF. The Stockholm Cohort Study on Atrial Fibrillation reported that 54% of individuals without contraindications were prescribed warfarin at hospital discharge. The investigators found a marked decrease in use of warfarin among individuals >80 years of age.[6] Strikingly similar results were reported from a study conducted in the United States of 21 teaching, 13 community, and 4 Veterans Administration hospitals during the same time period. Among patients with AF and no obvious contraindication, warfarin use at discharge was 54%. Age ≥80 years and perceived bleeding risk were negative predictors of warfarin use.[7] The Euro Heart Survey also showed that age >75 years was a

[1]Department of Medicine, Research Unit-Section of General Internal Medicine, Boston University Medical Center, Boston, Massachusetts.

Address for correspondence and reprint requests: Elaine M. Hylek, M.D., Department of Medicine, Research Unit-Section of General Internal Medicine, Boston University Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118 (e-mail: ehylek@bu.edu).

Antithrombotic Management of Atrial Fibrillation; Guest Editor, Marcel Levi, M.D., Ph.D.

Semin Thromb Hemost 2009;35:548–553.
DOI 10.1055/s-0029-1240015. ISSN 0094-6176.

negative predictor of VKA use for stroke prevention in AF.[8]

Understanding the discrepancy between consensus recommendations and clinical practice is critical to optimizing stroke prevention in AF. Aggregate rates of major hemorrhage among participants of randomized trials and observational cohorts have been reassuringly low. However, a full accounting of inclusion and exclusion criteria and end-point definitions is critical for an informed interpretation of these data. Because the validity of a randomized controlled trial is threatened by nonadherence, loss to follow-up, dropouts, and treatment crossovers, trials tend to enroll less sick and more compliant individuals with lower bleeding risk to best ensure internal validity. These realities prompt concerns about the comparability of trial populations and patients cared for in routine practice.[9] In addition, in contrast to early AF trials, more recent trials have preferentially enrolled patients already taking VKA. This is an additional source of bias because these patients tend to be less acutely ill and have already demonstrated an ability to tolerate anticoagulant therapy. Observational cohort studies are also vulnerable to this bias. Inception cohort studies of patients newly taking VKA are time intensive due to their reliance on incident cases. It is also challenging from a logistical perspective to identify and track an individual from the time of the first dose.

MAJOR HEMORRHAGE AMONG ELDERLY INDIVIDUALS WITH ATRIAL FIBRILLATION

The preponderance of evidence from randomized trials and observational studies supports higher bleeding rates among elderly individuals with AF.[10,11] To interpret published rates, it is important to determine the definition of major hemorrhage and whether or not there was adequate representation of the age group in question. Thresholds to classify a bleeding episode as major vary widely. Less subjective definitions require evidence of red blood cell transfusion or change in hemoglobin, but the hemoglobin criterion ranges from a 2 g/dL decrease to a 5 g/dL decrease.[12,13] Definitions that use hospitalization as a sole criterion tend to bias toward higher bleeding event rates in the elderly population. Uniform reporting of hemorrhage is critical to better discern drug safety profiles across different patient subgroups. To determine the degree of selection bias among study participants, it is important to discern the proportion of patients enrolled from the denominator population of potentially eligible subjects. This percentage, in addition to the absolute number, helps to define the external validity of the study population. The spectrum of aging spans the youngest old to the frailest old, and baseline risk varies considerably across these patient strata.[14] Biological age is not necessarily reflected by chronological age, and for this reason, there is no strict age cutoff for use of VKA.

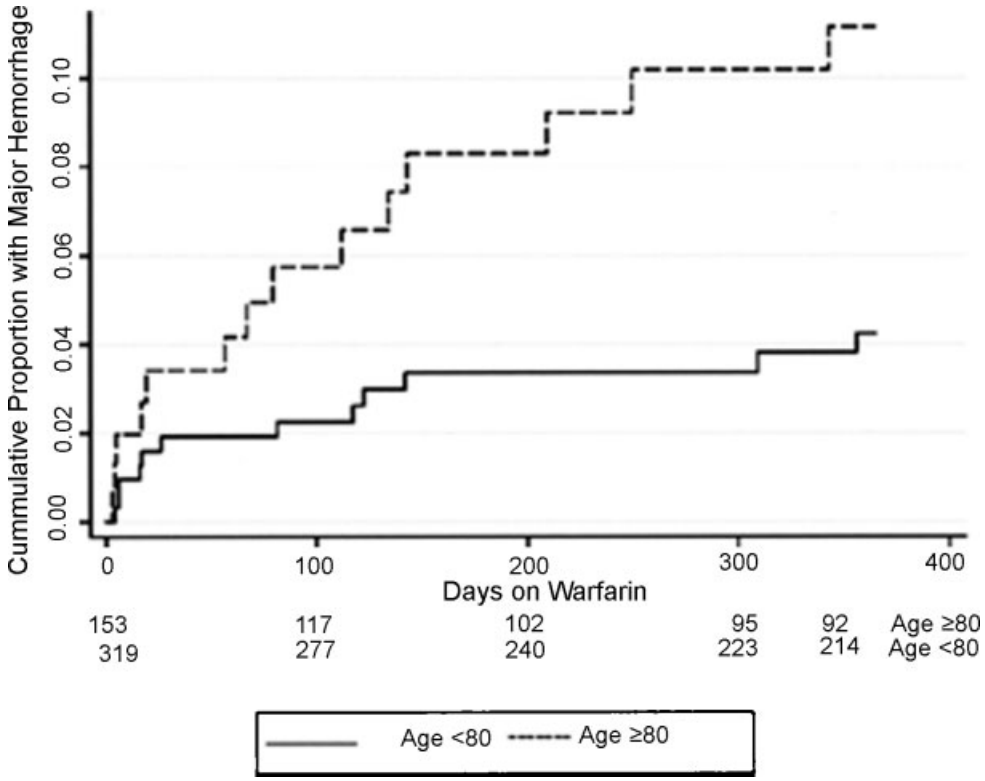


Figure 1 Cumulative incidence of major bleeding among patients age <80 years of age and ≥80 years ($n = 472$). Numbers below graph are the number of patients without bleeding who continued on warfarin at that time point ($p = 0.009$, log-rank test). (Reprinted with permission from Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. Circulation 2007;115:2689–2696).

In the AFFIRM trial (Atrial Fibrillation Follow-up Investigation of Rhythm Management), the risk of major hemorrhage increased by ~5% per year of age.[15] More recently, elderly patients were found to be at greatest risk for bleeding complications in the Amadeus trial that compared VKA to idraparinux, a once-weekly injectable factor Xa inhibitor.[16] Higher rates of major hemorrhage were also reported among patients ≥80 years of age newly taking warfarin for AF (Fig. 1).[17] This inception cohort study confirmed the first 90 days to be the most risk-prone period and highlighted the paradox of warfarin discontinuation among patients at the highest risk for stroke. Of patients who were ≥80 years of age, 26% stopped warfarin within the first year for perceived safety concerns (Fig. 2).

PREDISPOSING FACTORS FOR MAJOR HEMORRHAGE

The etiology of hemorrhage in older individuals is multifactorial. Pathological changes associated with aging render patients susceptible and these biological effects in conjunction with increased vulnerability to polypharmacy and increased rates of invasive procedures underlie much of the bleeding that is experienced in clinical practice. Hemorrhage may herald the presence of an occult pathological lesion that is amenable to intervention. Vascular ectasias are particularly problematic because the bleeding is often recurrent and definitive treatment elusive.

Gastrointestinal Hemorrhage

The rate of lower gastrointestinal bleeding increases precipitously with age by an estimated 200-fold from the third to the ninth decade.[18] The differential diagnosis differs from that of younger individuals and most commonly includes diverticulosis, ischemic colitis, malignancy, and vascular ectasias. Approximately 70% of acute upper gastrointestinal bleeding occurs among individuals >60 years of age. Peptic ulcer disease remains the most common etiology among both young and old adults. Antiplatelet and anti-inflammatory drugs are potent risk factors for gastrointestinal hemorrhage. In addition, there is a differential detrimental effect of aspirin by age with men age ≥80 years the most prone to gastric mucosal complications.[19] To the extent that indications for antiplatelet and/or anti-inflammatory agents increase with age, particularly percutaneous coronary interventions and osteoarthritis, these readily available medications constitute a salient example of the hazardous sequelae of

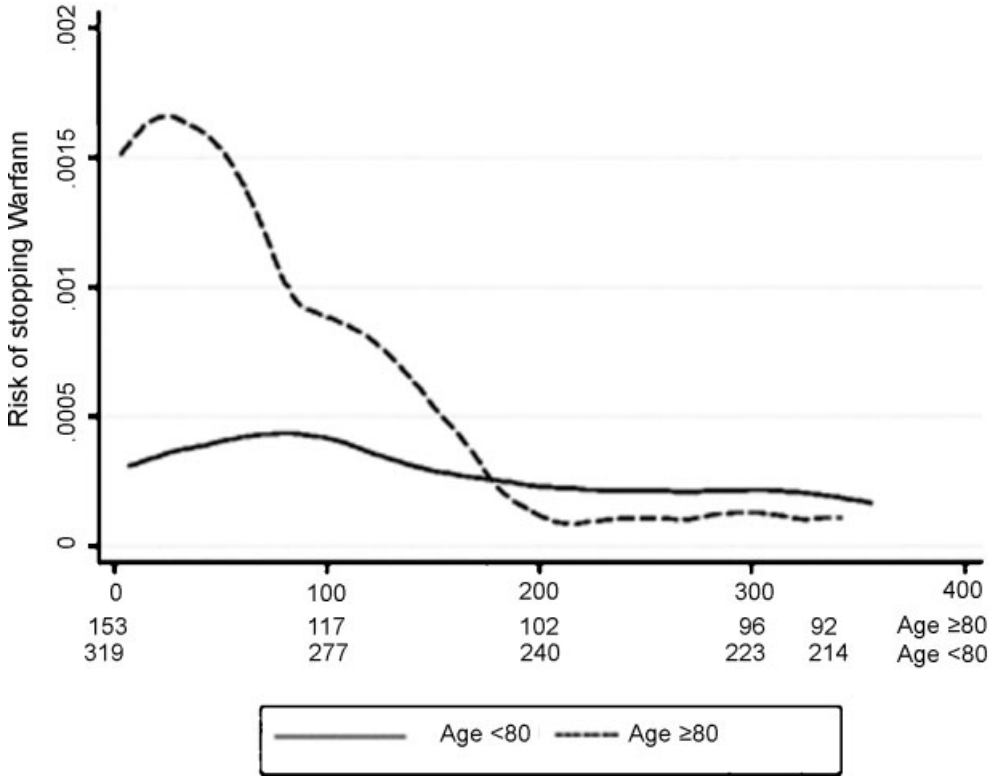


Figure 2 Risk of stopping warfarin in the first year on the basis of perceived safety concerns by age. Numbers below graph are the number of patients on warfarin at that time point ($p = 0.001$, log-rank test). (Reprinted with permission from Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. Circulation 2007;115:2689–2696).

polypharmacy among individuals in the older age group.

Intracranial Hemorrhage

Older age is also a potent risk factor for intracranial hemorrhage.[20] Leukoaraiosis, also termed white matter disease, is a chronic subcortical vasculopathy that was found to be an independent predictor of anticoagulant-related intracranial hemorrhage (hazard ratio, 2.7; 95% confidence interval [CI], 1.4 to 5.3) in the Stroke Prevention In Reversible Ischemia Trial (SPIRIT), in addition to age and anticoagulation intensity.[21] The prevalence of cerebral amyloid angiopathy and leukoaraiosis increase with age and heightens the risk of intracerebral hemorrhage.[22,23]

The contribution of fall risk to intracranial hemorrhage is controversial and a challenging area to study. Underlying mechanisms for the increased propensity for falls among elderly individuals have been well described.[24] Polypharmacy is often implicated because specific medications like α-blockers or tricyclic antidepressants or combinations of these medications with nodal agents, diuretics, or vasodilators may cause orthostatic hypotension and exacerbate underlying autonomic dysfunction. Impaired balance may also be attributable to sarcopenia, decreased proprioception, or loss of visual acuity. Perhaps the most definitive study to date, Gage and colleagues reported rates of intracranial hemorrhage among Medicare beneficiaries deemed to be high fall risk through physician documentation of "frequent falls," "history of falls," "multiple falls," or "tendency for falls" in the medical record.[25] Longitudinal follow-up of this AF cohort revealed intracranial hemorrhage rates of 2.8% (95% CI, 1.9–4.1) among high-fall-risk patients versus 1.1% (95% CI, 1.0–1.3) for those individuals without this designation. Of note, rates of ischemic stroke were 13.7% versus 6.9%, respectively, supporting use of anticoagulant therapy in the presence of multiple stroke risk factors.

OPTIMAL CHOICE OF ANTITHROMBOTIC THERAPY AMONG ELDERLY INDIVIDUALS WITH ATRIAL FIBRILLATION

Current guidelines recommend VKA for individuals with a history of prior stroke or transient ischemic attack and for those patients with two or more of the following risk factors: ≥age 75 years, hypertension, diabetes mellitus, heart failure, or left ventricular dysfunction.[26] Untreated, the estimated risk of stroke in the presence of two of these risk factors is 4% (95% CI, 3.1 to 5.1) per year and increases by ~50% with each additional risk factor.[27] Randomized trials have definitely shown that aspirin, including dual antiplatelet therapy, is not effective for high-risk patients compared with VKA.[28] The rate of vascular events among AF patients randomized to aspirin plus clopidogrel was 5.64% per year versus 3.63% for patients randomized to VKA in the Atrial Fibrillation Clopidogrel Trial with Irbesartan for Prevention of Vascular Events (ACTIVE-W) trial.[29] Rates of major hemorrhage were not significantly different in the two groups: 2.4% versus 2.2%, respectively. More recently, the Birmingham Atrial Fibrillation Treatment of the Aged (BAFTA) study found warfarin to be superior to aspirin in the prevention of stroke (1.8% versus 3.8% per year) among patients age ≥75 years for whom there was clinical pretrial uncertainty about which of the two treatments should be used. As found in ACTIVE-W, rates of major hemorrhage in both arms were similar (1.9% versus 2.2% per year), effectively dispelling the notion that antiplatelet therapy is innocuous and safer than warfarin.[30] Based on the available evidence, there is a limited, if any, role for aspirin therapy for stroke prevention in AF.

STRATEGIES TO REDUCE RISK OF MAJOR HEMORRHAGE AMONG ELDERLY PATIENTS WITH ATRIAL FIBRILLATION

Older patients are at the highest risk of stroke, and AF-related strokes are associated with a 30-day mortality of 24%.[4,31] Given the morbidity and mortality, it is imperative that strategies to minimize hemorrhagic complications be implemented. Attentive management of hypertension with goals of ≤130/80 mm Hg decrease the risk of both ischemic and hemorrhagic stroke.[32] Hazardous drug combinations of VKA plus aspirin should be avoided, when possible, because the risk of gastrointestinal hemorrhage increases dramatically (adjusted relative risk, 6.48; 95% CI, 4.25 to 9.87).[33] Aspirin also increases the risk of intracranial hemorrhage.[34] For elderly patients on VKA, choice of coronary stent, bare metal versus drug eluting, should incorporate bleeding risk parameters to minimize exposure to dual antiplatelet therapy.[35] Addition of proton pump inhibitors as prophylaxis in this setting has recently been questioned because these agents may decrease the effectiveness of clopidogrel.[36,37]

Warfarin dose requirements decline with age.[38] Older age is also associated with a slower rate of recovery following an episode of excessive anticoagulation.[39] Hospitalization induces changes in diet and medications, and discharge transitions are often fraught with fragmentation. For all of these reasons, the initiation period of VKA confers the highest risk of serious bleeding, and vigilant monitoring of anticoagulation intensity is paramount. A recent analysis casts doubt on the overall cost effectiveness of genotype-guided warfarin dosing to offset this risk of early adverse events.[40] Timely dose adjustment and frequent testing during the first month of VKA therapy is absolutely essential to optimize

patient safety. In addition, although paracetamol (acetaminophen) is the preferred antipyretic and pain reliever of choice, it is known to potentiate the anticoagulant effect of VKA with protracted use of larger doses.[41] The touted mechanism is interference with enzymes of the vitamin K cycle. Decompensated heart failure, liver disease, chemotherapy, and amiodarone are other powerful potentiators of warfarin's effect.[42]

Assessment of patient fall risk should be conducted and interventions implemented to minimize this risk. Soft tissue injuries and excessive bruising often precipitate cessation of warfarin therapy. If postural hypotension persists despite medication adjustment, other remedial measures need to be instituted like compression stockings or mineralocorticoid therapy. Cognitive dysfunction is not a contraindication to anticoagulant therapy providing reliable dosing can be achieved with the assistance of a caregiver or other structured medication dispensing plan.

FUTURE DIRECTIONS

Risk stratification among patients with AF remains suboptimal.[43] Whether or not inclusion of biomarker, echocardiographic, and genetic data will improve discrimination awaits large population-based studies. Valid prediction models for intracranial hemorrhage are direly needed, but the rarity of this outcome will continue to challenge this directive. Cerebral microbleeds as detected by gradient-echo magnetic resonance imaging show promise as a possible marker for intracerebral bleeding on anticoagulant therapy.[44,45] Critical questions that will ensue include overall cost effectiveness of this approach and performance threshold of a cutoff score that would justify withholding antithrombotic therapy in high-risk patients with AF.

Whether or not newer anticoagulant drugs with different molecular targets, wider therapeutic windows, shorter half-lives, and less dietary and drug interactions will be safer in elderly patients with AF awaits the conclusion of ongoing clinical trials and subsequent validation in large population-based studies.

REFERENCES

1. Miyasaka Y, Barnes ME, Gersh BJ, et al. Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. Circulation 2006;114(2):119–125
2. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke 1991;22(8):983–988
3. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation 1998;98(10): 946–952
4. Lin HJ, Wolf PA, Kelly-Hayes M, et al. Stroke severity in atrial fibrillation. The Framingham Study. Stroke 1996; 27(10):1760–1764
5. Atrial Fibrillation Investigators. 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(13):1449–1457
6. Friberg L, Hammar N, Ringh M, Pettersson H, Rosenqvist M. Stroke prophylaxis in atrial fibrillation: who gets it and who does not? Report from the Stockholm Cohort-study on Atrial Fibrillation (SCAF-study) Eur Heart J 2006;27(16): 1954–1964
7. Waldo AL, Becker RC, Tapson VF, Colgan KJ; NABOR Steering Committee. Hospitalized patients with atrial fibrillation and a high risk of stroke are not being provided with adequate anticoagulation. J Am Coll Cardiol 2005; 46(9):1729–1736
8. Nieuwlaat R, Capucci A, Lip GY, et al; Euro Heart Survey Investigators. Antithrombotic treatment in real-life atrial fibrillation patients: a report from the Euro Heart Survey on Atrial Fibrillation. Eur Heart J 2006;27(24):3018–3026
9. Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. JAMA 2007;297(11):1233–1240
10. van der Meer FJ, Rosendaal FR, Vandenbroucke JP, Briët E. Bleeding complications in oral anticoagulant therapy. An analysis of risk factors. Arch Intern Med 1993;153(13):1557–1562
11. The Stroke Prevention in Atrial Fibrillation Investigators. Bleeding during antithrombotic therapy in patients with atrial fibrillation. Arch Intern Med 1996;156(4):409–416
12. Seto AC, Kenyon K, Wittkowsky AK. Discrepancies in identification of major bleeding events in patients taking warfarin. Pharmacotherapy 2008;28(9):1098–1103
13. Chesebro JH, Knatterud G, Roberts R, et al. Thrombolysis in Myocardial Infarction (TIMI) Trial, Phase I: a comparison between intravenous tissue plasminogen activator and intravenous streptokinase. Clinical findings through hospital discharge. Circulation 1987;76(1):142–154
14. Bergman H, Ferrucci L, Guralnik J, et al. Frailty: an emerging research and clinical paradigm—issues and controversies. J Gerontol A Biol Sci Med Sci 2007;62(7):731–737
15. DiMarco JP, Flaker G, Waldo AL, et al; AFFIRM Investigators. 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(4):650–656
16. Bousser MG, Bouthier J, Büller HR, et al; Amadeus Investigators. Comparison of idraparinux with vitamin K antagonists for prevention of thromboembolism in patients with atrial fibrillation: a randomised, open-label, non-inferiority trial. Lancet 2008;371(9609):315–321
17. Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. Circulation 2007;115:2689–2696
18. Longstreth GF. Epidemiology and outcome of patients hospitalized with acute lower gastrointestinal hemorrhage: a population-based study. Am J Gastroenterol 1997;92(3):419–424

19. Hernandez-Diaz S, Garcia Rodriguez LA. Cardioprotective aspirin users and their excess risk of upper gastrointestinal complications. BMC Med 2006;20:22
20. Hylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. Ann Intern Med 1994;120(11):897–902
21. Gorter JW. Major bleeding during anticoagulation after cerebral ischemia: patterns and risk factors. Stroke Prevention In Reversible Ischemia Trial (SPIRIT). European Atrial Fibrillation Trial (EAFT) study groups. Neurology 1999; 53(6):1319–1327
22. Rosand J, Hylek EM, O'Donnell HC, Greenberg SM. Warfarin-associated hemorrhage and cerebral amyloid angiopathy: a genetic and pathologic study. Neurology 2000; 55(7):947–951
23. Smith EE, Rosand J, Knudsen KA, Hylek EM, Greenberg SM. Leukoaraiosis is associated with warfarin-related hemorrhage following ischemic stroke. Neurology 2002;59(2):193–197
24. Tinetti ME. Clinical practice. Preventing falls in elderly persons. N Engl J Med 2003;348(1):42–49
25. Gage BF, Birman-Deych E, Kerzner R, Radford MJ, Nilasena DS, Rich MW. Incidence of intracranial hemorrhage in patients with atrial fibrillation who are prone to fall. Am J Med 2005;118(6):612–617
26. Fuster V, Ryden LE, Cannom DS, et al. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (writing committee to revise the 2001 guidelines for the management of patients with atrial fibrillation): developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. Circulation 2006;114:e257–354
27. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA 2001;285(22):2864–2870
28. Stroke Prevention in Atrial Fibrillation Investigators. Warfarin versus aspirin for prevention of thromboembolism in atrial fibrillation: Stroke Prevention in Atrial Fibrillation II Study. Lancet 1994;343(8899):687–691
29. Connolly S, Pogue J, Hart R, et al; ACTIVE Writing Group 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(9526):1903–1912
30. Mant J, Hobbs FD, Fletcher K, et al; BAFTA investigators-Midland Research Practices Network (MidReC). 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(9586):493–503
31. 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(11):1019–1026
32. Delaney JA, Opatrny L, Brophy JM, Suissa S. Drug-drug interactions between antithrombotic medications and the risk of gastrointestinal bleeding. CMAJ 2007;177(4):347–351
33. Arima H, Hart RG, Colman S, et al; PROGRESS Collaborative Group. Perindopril-based blood pressure-lowering reduces major vascular events in patients with atrial fibrillation and prior stroke or transient ischemic attack. Stroke 2005;36(10):2164–2169
34. Hart RG, Tonarelli SB, Pearce LA. Avoiding central nervous system bleeding during antithrombotic therapy: recent data and ideas. Stroke 2005;36(7):1588–1593
35. King SB III, Smith SC Jr, Hirshfeld JW Jr, et al; 2005 WRITING COMMITTEE MEMBERS. 2007 Focused Update of the ACC/AHA/SCAI 2005 Guideline Update for Percutaneous Coronary Intervention: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines: 2007 Writing Group to Review New Evidence and Update the ACC/AHA/SCAI 2005 Guideline Update for Percutaneous Coronary Intervention, Writing on Behalf of the 2005 Writing Committee. Circulation 2008;117(2):261–295
36. Gilard M, Arnaud B, Cornily JC, et al. Influence of omeprazole on the antiplatelet action of clopidogrel associated with aspirin: the randomized, double-blind OCLA (Omeprazole CLopidogrel Aspirin) study. J Am Coll Cardiol 2008;51(3):256–260
37. Simon T, Verstuyft C, Mary-Krause M, et al; French Registry of Acute ST-Elevation and Non-ST-Elevation Myocardial Infarction (FAST-MI) Investigators. Genetic determinants of response to clopidogrel and cardiovascular events. N Engl J Med 2009;360(4):363–375
38. Garcia D, Regan S, Crowther M, Hughes RA, Hylek EM. Warfarin maintenance dosing patterns in clinical practice: implications for safer anticoagulation in the elderly population. Chest 2005;127(6):2049–2056
39. Hylek EM, Regan S, Go AS, Hughes RA, Singer DE, Skates SJ. Clinical predictors of prolonged delay in return of the international normalized ratio to within the therapeutic range after excessive anticoagulation with warfarin. Ann Intern Med 2001;135(6):393–400
40. Eckman MH, Rosand J, Greenberg SM, Gage BF. Cost-effectiveness of using pharmacogenetic information in warfarin dosing for patients with nonvalvular atrial fibrillation. Ann Intern Med 2009;150(2):73–83
41. Thijssen HH, Soute BA, Vervoort LM, Claessens JG. Paracetamol (acetaminophen) warfarin interaction: NAPQI, the toxic metabolite of paracetamol, is an inhibitor of enzymes in the vitamin K cycle. Thromb Haemost 2004; 92(4):797–802
42. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G; American College of Chest Physicians. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest 2008; 133(6, suppl):160S–198S
43. Stroke Risk in Atrial Fibrillation Working Group. Comparison of 12 risk stratification schemes to predict stroke in patients with nonvalvular atrial fibrillation. Stroke 2008; 39(6):1901–1910
44. Koennecke HC. Cerebral microbleeds on MRI: prevalence, associations, and potential clinical implications. Neurology 2006;66(2):165–171
45. Copenhaver BR, Hsia AW, Merino JG, et al. Racial differences in microbleed prevalence in primary intracerebral hemorrhage. Neurology 2008;71(15):1176–1182

Review article

FIC Federazione Italiana di Cardiologia
Italian Federation of Cardiology

The need for new oral anticoagulants in clinical practice

Elaine M. Hylek

Atrial fibrillation, the most clinically important arrhythmia, is a considerable independent risk factor for the development of stroke. Vitamin K antagonists, primarily warfarin, are the only oral anticoagulants currently available for the long-term prevention of stroke in patients with atrial fibrillation. Although there is considerable evidence that warfarin reduces the risk of stroke in patients with atrial fibrillation, it is associated with various challenges to its use in routine clinical practice. Numerous studies have shown that it is underused in patients with atrial fibrillation, particularly elderly patients who would seem to benefit the most. Many physicians appear hesitant to prescribe warfarin for patients with atrial fibrillation because they are unconvinced that the benefits seen in clinical trials will actually translate into their everyday practice. As a result, there is a pressing need for convenient new, well-tolerated and effective oral anticoagulants that do not require frequent dose adjustment and routine coagulation monitoring.
***J Cardiovasc Med* 10:605–609 © 2009 Italian Federation of Cardiology.**

Journal of Cardiovascular Medicine 2009, 10:605–609

Keywords: atrial fibrillation, oral anticoagulants, stroke prevention, warfarin

Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

Correspondence to Elaine M. Hylek, MD, MPH, Department of Medicine, Boston University Medical Center, Boston, MA 02118, USA
Tel: +1 617 414 3743; fax: +1 617 414 4676; e-mail: ehylek@bu.edu

Received 18 May 2009 Accepted 18 May 2009

Introduction

The burden of thromboembolic disease is increasing throughout the world. An example of this is the increasing incidence of venous thromboembolism (VTE) in Europe and the USA. It was estimated that there were more than 900 000 occurrences of pulmonary embolism and deep vein thrombosis (DVT), and 296 000 VTE-related deaths in the USA in 2002 [1]; and 762 000 pulmonary embolisms and DVTs, and 370 000 VTE-related deaths in six European countries in 2004 [2]. The incidence of VTE is expected to continue to rise because age is an independent risk factor for VTE [3] and the elderly population in the developed world is growing [4]. In addition, arthritis is more prevalent in the elderly [5] and so the number of total hip and knee replacement procedures will continue to increase [6].

The burden of thromboembolic disease is not confined to VTE. Atrial fibrillation is the most frequently occurring, clinically important arrhythmia, and an important independent risk factor for stroke [7]. The rate of stroke in patients with atrial fibrillation is approximately five times higher than in those without atrial fibrillation [8,9], and the incidence of atrial fibrillation increases with age [9–13]. The Framingham study [9] has shown that the attributable risk of stroke for patients with atrial fibrillation increases from 1.5% at age 50–59 years to 23.5% at age 80–89 years. Indeed, at over 80 years of age, atrial fibrillation is the only cardiovascular condition associated with an increased risk of stroke [9]. The incidence of atrial fibrillation is expected to continue to increase, and it has been projected that there will be more than 10 million people with atrial fibrillation in the USA by 2035 (Fig. 1) [14]. The increased prevalence of atrial fibrillation is also due to improvements in patients' survival from conditions that predispose to atrial fibrillation such as coronary heart disease. Because the lifetime risk for the development of atrial fibrillation in Europe is similar to that in the USA [11], we can expect the prevalence of atrial fibrillation to also increase within Europe.

Atrial fibrillation is associated with substantial morbidity and mortality. Approximately 15% of all strokes occur in those with atrial fibrillation [15], and this patient group has a 50–90% increase in mortality after adjusting for coexisting cardiovascular conditions [16]. Stroke is the third most frequent cause of death in developed countries; each year 15 million people have a stroke and, of these, 5 million die and an additional 5 million are permanently disabled [17]. Stroke is, therefore, a major problem throughout the developed world and effective therapies for its prevention are required.

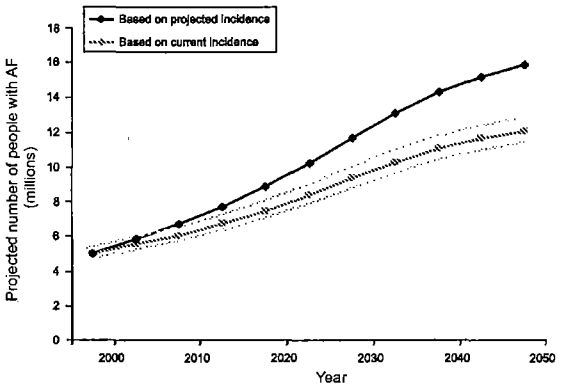
Vitamin K antagonists (VKAs) are the only oral anticoagulants currently available for the long-term prevention of stroke in patients with atrial fibrillation, and the primary VKA is warfarin. However, although effective, reducing the overall risk of stroke by 62% [18], warfarin is associated with various challenges to its use in routine clinical practice.

Challenges inherent with the use of warfarin in routine clinical practice

Warfarin is very effective when used as a long-term anticoagulant for the prevention of stroke in atrial fibrillation [19]. However, a number of physiological and

1558-2027 © 2009 Italian Federation of Cardiology DOI:10.2459/JCM.0b013e32832e4954

Fig. 1



Projected increase in the incidence of atrial fibrillation in the USA. Assumes no further increase in the incidence of age-adjusted atrial fibrillation (grey curve with 95% CI) and continued increase in incidence rate as evident in 1980–2000 (black curve). CI, confidence interval. Reproduced with permission from [14].

pharmacological factors influence its therapeutic efficacy and safety, and the rate of adverse events associated with its use may result in an unfavourable risk–benefit ratio.

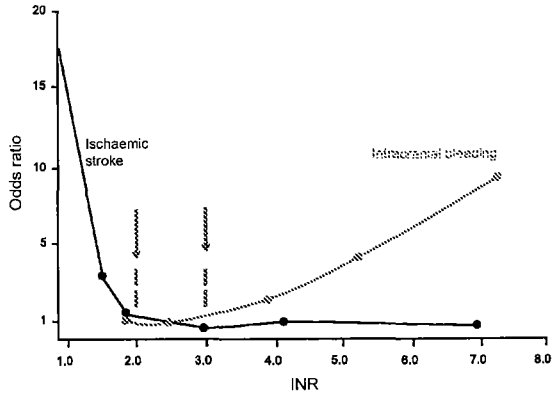
Warfarin has an unpredictable dose response, which is highly variable between individuals: elderly patients being particularly sensitive to its anticoagulant effect [20]. Several studies [20–23] have demonstrated that patients require a lower maintenance dose of warfarin as they age. In a cross-sectional analysis [22] of 2305 patients receiving warfarin, the decrease in dose was an average of 0.5 mg per decade, and there was a 21% decrease in dose over a 15-year period in a longitudinal study [23]. Data from a large anticoagulation clinic showed that, among patients with an international normalized ratio (INR) target range of 2.0–3.0, only 25% of those aged above 80 years required a weekly warfarin maintenance dose of more than 30 mg, compared with approximately 70% of those aged below 65 years [20]. Garcia *et al.* [21] demonstrated that not only did warfarin requirement decrease greatly with age but also that women required lower doses than men of the same age. Consequently, the frequently used starting warfarin dose of 5 mg/day could lead to overcoagulation in many elderly patients, particularly women [21].

Several other factors can also contribute to the variable dose response of warfarin. Interference with the cyclic interconversion of vitamin K, impaired absorption of vitamin K and fluctuations in dietary vitamin K can all induce instability in the anticoagulant effect of warfarin [20]. Warfarin is also associated with many drug interactions, including some of the most frequently used prescription-only medications (propranolol, amiodarone, ceftriaxone, doxycycline, erythromycin, metronidazole and esomeprazole) and over-the-counter medications (such as acetaminophen, fenoprofen, lovastatin and clarithromycin) [20]. Genetic factors, including a frequently occurring mutation in the gene coding for the cytochrome P450 2C9 hepatic enzyme, a hereditary resistance to warfarin and a mutation in the factor IX propeptide, also affect the dose response of warfarin [24]. In addition, comorbid conditions, such as congestive heart failure and cancer, can also influence its dose response.

Another challenge for physicians when using warfarin is its narrow therapeutic window, which makes it difficult to maintain a balance between preventing a stroke and avoiding bleeding complications. In earlier studies it was shown that the prothrombin time ratio was the dominant risk factor for intracranial haemorrhage and that it is important to keep the INR within the range of 2.0–3.0 [25]. Data from clinical studies suggest that optimal protection against ischaemic stroke is achieved when patients have an INR between 2.0 and 3.0 [25,26], and the American College of Cardiology/American Heart Association and European Society of Cardiology guidelines for the management of patients with atrial fibrillation [19] now recommend this as the target INR range (Fig. 2). Because of these challenges, patients receiving oral anticoagulation with warfarin require careful monitoring, dose adjustment, clinical surveillance and continuous patient education [24].

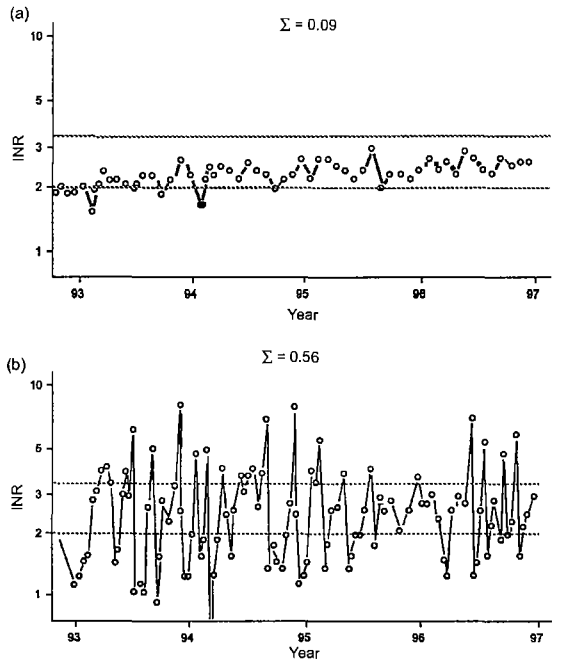
After the initiation of warfarin, patients should have their INR checked daily until stable anticoagulation has been achieved. The INR should then be checked weekly for several weeks to confirm the stability of anticoagulation. The frequency of testing thereafter will depend on the variability of the INR, which can differ markedly across individual patients (Fig. 3) (Hylek, unpublished data).

Fig. 2



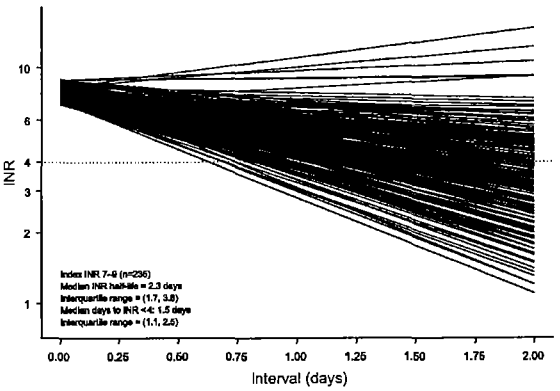
Anticoagulation intensity for stroke prevention in atrial fibrillation. INR, international normalized ratio. Reproduced with permission from [19].

Fig. 3



Patients with (a) minimal variability in the normalized ratio, (b) high international normalized ratio variability. Σ, mathematical representation of standard deviation; INR, international normalized ratio. Adapted from Hylek (unpublished data).

Fig. 4



Patients' international normalized ratio (INR) decay curves over 48 h. Adapted from Hylek ([20] and unpublished data).

Once the INR response is stable, the frequency of testing can be reduced to once in every 4 weeks [24].

The long half-life of warfarin (36–42 h) [20] offers physicians yet another challenge if, for example, therapy has to be interrupted for invasive procedures, such as elective surgery, to allow recovery from an elevated INR or in the case of major haemorrhage. If warfarin is stopped for elective surgery, it will take approximately 4 days for the INR to reach 1.5, when surgery can be performed safely [27]. It will also take approximately 3 days for the INR to return to 2.0 after warfarin therapy is restarted [27]. However, during the period of warfarin withdrawal, the patient will still have partial thromboembolic protection. It has, therefore, been estimated that this temporary withdrawal of warfarin exposes the patient to a risk of thromboembolism equivalent to 1 day without anticoagulation before and another day after surgery [27].

Despite the extensive use of warfarin for several decades, little has been published on the patient-specific factors that influence normalization of the INR after withdrawal of therapy [28]. This is important, as a better understanding of INR decay patterns after cessation of warfarin would help to better predict time to normalization and optimize the use of vitamin K to treat excessive anticoagulation [28]. The variability of patients' INR decay curves over the 48-h period after withholding two doses of warfarin is shown in Fig. 4. In this study of 633 patients with an INR of at least 6.0, 37% still had an INR of at least 4.0 after two doses of warfarin had been withheld [28]. Moreover, patients who required lower warfarin maintenance doses were more likely to have an INR of at least 4.0 2 days after the withdrawal of treatment, compared with those who required higher maintenance doses of warfarin [adjusted odds ratio (OR) per 10 mg = 0.87, 95% confidence interval (CI) 0.79–0.97]. Additional risk factors identified for an INR of at least 4.0 included age (OR per decade of life = 1.18, 95% CI 1.01–1.38), initial INR (OR per unit = 1.25, 95% CI 1.14–1.37), decompensated congestive heart failure (OR = 2.79, 95% CI 1.30–5.98) and active cancer (OR = 2.48, 95% CI 1.11–5.57) (Fig. 4) [28].

Rapid reversal of the anticoagulant effect of warfarin is critical in patients on long-term anticoagulation who require urgent surgical procedures or who experience major haemorrhage. Vitamin K_1 can be given to reverse the effects of warfarin. INR can usually be normalized within 24 h. However, the use of vitamin K_1 can make subsequent INR stabilization difficult [29]. Immediate reversal can be achieved with fresh frozen plasma or factor concentrates. However, vitamin K_1 is still essential for sustaining the reversal achieved by these agents [29].

Underutilization of warfarin in patients with atrial fibrillation

Regardless of the evidence that warfarin is effective, numerous studies have shown that it is underused in patients with atrial fibrillation, particularly in those elderly patients who would seem to benefit the most. Many physicians remain hesitant to prescribe warfarin for

patients with atrial fibrillation because they are not convinced that the benefits of warfarin seen in clinical trials will translate into their everyday practice [30,31]. A survey of 1189 randomly selected USA office-based practitioners in primary care, cardiology and neurology found that most were 'very' or 'somewhat' likely to use anticoagulants in a 65-year-old patient, but fewer were likely to do so in a 75-year-old patient with left atrial enlargement (71 vs. 63%), intermittent or paroxysmal atrial fibrillation (68 vs. 56%), recent-onset atrial fibrillation (86 vs. 80%) or embolic stroke (96 vs. 93%) [30]. In addition, even among those physicians equally likely to use anticoagulation for both 65- and 75-year-old patients, anticoagulation intensity (measured by INR or prothrombin time ratio) was significantly lower ($P < 0.04$) among the older group.

Data from Europe and the USA show that only approximately half of the high-risk patients with atrial fibrillation receive warfarin [32,33]. In a retrospective study [32] of 945 patients from the USA with atrial fibrillation, of which 86% were stratified as high risk for stroke, only 55% received warfarin. Similar results were found in another retrospective study [33] in 2796 Swedish patients with atrial fibrillation: although 68% had indications for, and no apparent contraindications to, warfarin, only 54% of patients received it.

Importantly, warfarin use does not appear to increase significantly in those at the highest risk of stroke [33,34]. For example, both studies [32,33] from the USA and Sweden referred to above found that underutilization was particularly prevalent in patients aged over 80 years. The Euro Heart Survey [34] also found that atrial fibrillation therapy was not generally tailored to a patient's stroke risk profile. The availability of an anticoagulation monitoring clinic was an important factor in addition to perceived bleeding risk and patient preference [32,33]. Perceived bleeding risk is also one of the major reasons why warfarin therapy is discontinued in the elderly. One study in 472 patients showed that physician-perceived safety issues accounted for 81% of the patients aged at least 80 years who discontinued warfarin within the first year [35].

Therefore, it is clear that patients with atrial fibrillation who are at increased risk of stroke, and particularly elderly patients with atrial fibrillation, need new oral anticoagulants that will overcome the limitations associated with warfarin. New anticoagulants that could provide well-tolerated and effective prophylaxis without the need for routine coagulation monitoring would help to optimize stroke prevention in patients with atrial fibrillation.

Conclusion

The burden of thromboembolic disease is expected to continue to increase, and oral anticoagulants will have a critical role in the long-term management of chronic conditions such as the prevention of stroke in patients with atrial fibrillation. Despite being effective, warfarin is limited by a variable dose response, a narrow therapeutic window, the need for frequent INR monitoring and a long half-life, all of which contribute to its underutilization.

There is a pressing need for new, well-tolerated and effective oral anticoagulants that do not require frequent dose adjustment and routine coagulation monitoring, thereby offering both patients and clinicians the benefit of convenient, long-term, effective anticoagulation. Several new oral anticoagulants are in development for thromboembolic diseases, including dabigatran etexilate, a direct thrombin inhibitor, and inhibitors of factor Xa such as rivaroxaban, apixaban, PRT054021, YM150 and DU-176b. To date, two of these, dabigatran etexilate and rivaroxaban, have been proven to be effective for the prevention of VTE after elective total hip and knee replacement. Both dabigatran and rivaroxaban, in addition to apixaban, are now being evaluated in phase III trials for the prevention of stroke in patients with atrial fibrillation. The results of these studies are eagerly awaited.

Acknowledgements

The authors would like to acknowledge Christine Gardner of Chameleon Communications International who provided editorial assistance with funding from Bayer Schering Pharma AG.

Elaine M. Hylek did not receive any honoraria or consultancy fees for writing this manuscript. The author reports receiving grant support from AstraZeneca, serving as an advisor to Boehringer Ingelheim, Bristol-Myers Squibb, Sanofi-Aventis and The Medicine Company; and participating in clinical symposia sponsored by Bayer HealthCare and Bristol-Myers Squibb.

References

1 Heit JA, Cohen AT, Anderson FA. on behalf of the VTE Impact Assessment Group. Estimated annual number of incident and recurrent, nonfatal and fatal venous thromboembolism (VTE) events in the US [abstract #910]. *Blood* 2005; 106.
2 Cohen AT, Agnelli G, Anderson FA, Arcelus JI, Bergqvist D, Brecht JG, *et al.* Venous thromboembolism (VTE) in Europe. The number of VTE events and associated morbidity and mortality. *Thromb Haemost* 2007; **98**:756–764.
3 Geerts WH, Bergqvist D, Pineo GF, Heit JA, Samama CM, Lassen MR, *et al.* Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th ed.). *Chest* 2008; **133**:381S–453S.
4 World Health Organization. The world is fast ageing: have we noticed? 2006. http://www.who.int/ageing/en. [Accessed 12 May 2009].
5 Reginster JY. The prevalence and burden of arthritis. *Rheumatology (Oxford)* 2002; **41 (Suppl 1)**:3–6.
6 Lavernia C, Lee DJ, Hernandez VH. The increasing financial burden of knee revision surgery in the United States. *Clin Orthop Relat Res* 2006; **446**:221–226.
7 Laupacis A, Albers G, Dalen J, Dunn M, Feinberg W, Jacobson A. Antithrombotic therapy in atrial fibrillation. *Chest* 1995; **108**:352S–359S.
8 Wolf PA, Dawber TR, Thomas HE Jr, Kannel WB. Epidemiologic assessment of chronic atrial fibrillation and risk of stroke: the Framingham study. *Neurology* 1978; **28**:973–977.

9 Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* 1991; **22**:983–988.
10 Furberg CD, Psaty BM, Manolio TA, Gardin JM, Smith VE, Rautaharju PM. Prevalence of atrial fibrillation in elderly subjects (the Cardiovascular Health Study). *Am J Cardiol* 1994; **74**:236–241.
11 Heeringa J, van der Kuip DAM, Hofman A, Kors JA, van Herpen G, Stricker BHC, *et al.* Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. *Eur Heart J* 2006; **27**:949–953.
12 Lake FR, Cullen KJ, de Klerk NH, McCall MG, Rosman DL. Atrial fibrillation and mortality in an elderly population. *Aust N Z J Med* 1989; **19**:321–326.
13 Phillips SJ, Whisnant JP, O'Fallon WM, Frye RL. Prevalence of cardiovascular disease and diabetes mellitus in residents of Rochester, Minnesota. *Mayo Clin Proc* 1990; **65**:344–359.
14 Miyasaka Y, Barnes ME, Gersh BJ, Cha SS, Bailey KR, Abhayaratna WP, *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.
15 Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation: a major contributor to stroke in the elderly. The Framingham Study. *Arch Intern Med* 1987; **147**:1561–1564.
16 Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. *Circulation* 1998; **98**:946–952.
17 World Health Organization. The atlas of heart disease and stroke. 2004. http://www.who.int/cardiovascular_diseases/resources/atlas/en/. [Accessed 12 May 2009].
18 Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. *Ann Intern Med* 1999; **131**:492–501.
19 Fuster V, Ryden LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, *et al.* ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation: executive summary – a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients With Atrial Fibrillation) – developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. *Circulation* 2006; **114**:700–752.
20 Hylek EM. Oral anticoagulants. Pharmacologic issues for use in the elderly. *Clin Geriatr Med* 2001; **17**:1–13.
21 Garcia D, Regan S, Crowther M, Hughes RA, Hylek EM. Warfarin maintenance dosing patterns in clinical practice: implications for safer anticoagulation in the elderly population. *Chest* 2005; **127**:2049–2056.
22 James AH, Britt RP, Raskino CL, Thompson SG. Factors affecting the maintenance dose of warfarin. *J Clin Pathol* 1992; **45**:704–706.
23 Wynne HA, Kamali F, Edwards C, Long A, Kelly P. Effect of ageing upon warfarin dose requirements: a longitudinal study. *Age Ageing* 1996; **25**:429–431.
24 Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G, *et al.* Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians evidence-based clinical practice guidelines (8th ed.). *Chest* 2008; **133**:160S–198S.
25 Hylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. *Ann Intern Med* 1994; **120**:897–902.
26 Hylek EM, Skates SJ, Sheehan MA, Singer DE. An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. *N Engl J Med* 1996; **335**:540–546.
27 Kearon C, Hirsh J. Management of anticoagulation before and after elective surgery. *N Engl J Med* 1997; **336**:1506–1511.
28 Hylek EM, Regan S, Go AS, Hughes RA, Singer DE, Skates SJ. Clinical predictors of prolonged delay in return of the international normalized ratio to within the therapeutic range after excessive anticoagulation with warfarin. *Ann Intern Med* 2001; **135**:393–400.
29 Baker RI, Coughlin PB, Gallus AS, Harper PL, Salem HH, Wood EM, *et al.* Warfarin reversal: consensus guidelines, on behalf of the Australasian Society of Thrombosis and Haemostasis. *Med J Aust* 2004; **181**:492–497.
30 McCrory DC, Matchar DB, Samsa G, Sanders LL, Pritchett EL. Physician attitudes about anticoagulation for nonvalvular atrial fibrillation in the elderly. *Arch Intern Med* 1995; **155**:277–281.
31 Albers GW, Yim JM, Belew KM, Bittar N, Hattemer CR, Phillips BG, *et al.* Status of antithrombotic therapy for patients with atrial fibrillation in university hospitals. *Arch Intern Med* 1996; **156**:2311–2316.
32 Waldo AL, Becker RC, Tapson VF, NABOR Steering Committee. 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.
33 Friberg L, Hammar N, Ringh M, Pettersson H, Rosenqvist M. Stroke prophylaxis in atrial fibrillation: who gets it and who does not? Report from the Stockholm Cohort-study on Atrial Fibrillation (SCAF-study). *Eur Heart J* 2006; **27**:1954–1964.
34 Nieuwlaat R, Capucci A, Lip GY, Olsson SB, Prins MH, Nieman FH, *et al.* Antithrombotic treatment in real-life atrial fibrillation patients: a report from the Euro Heart Survey on Atrial Fibrillation. *Eur Heart J* 2006; **27**:3018–3026.
35 Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. *Circulation* 2007; **115**:2689–2696.

Editorial Focus

Understanding low INR in clinical practice

Elaine M. Hylek[1]; Adam J. Rose[1,2]

[1]Department of Medicine, Research Unit-Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachussetts, USA; [2]Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, Massachussetts, USA

By human nature we focus more intently upon the harm that we may cause by doing too much than the harm that we may allow by doing too little (1, 2). It is well known that while vitamin K antagonists (VKA) are highly effective in preventing thromboembolic events, their use can also lead to serious haemorrhagic complications (3–7), some of which may be disabling or fatal. This very real risk is almost certainly part of the reason why VKA are underutilized in many patients with atrial fibrillation (8). Heightened fear of haemorrhage also evokes a heightened avoidance of elevated International Normalised Ratio (INR) as evidenced by the fact that, in most studies, patients spend more time below than above the target INR range (9–14). The research literature also reflects this tendency to focus on overanticoagulation, with previous studies contributing to a much greater understanding of the causes of high INR than low INR. This tendency may also reflect a difference in perceived risk: the daily risk of an adverse event is almost certainly greater when the INR is high (15) than when it is low (16). Nevertheless, subtherapeutic anticoagulation is associated with more frequent and more severe strokes and represents a more important phenomenon than our limited understanding of it might suggest (17).

In this issue, Rombouts et al. (18) report their findings on the frequency of low INR and risk factors for low INR among patients cared for by the Leiden Thrombosis Service. Of 13,443 patients initiating VKA therapy, 7,419 met the study eligibility criteria for stability defined as four consecutive INR determinations within the target range. Within four weeks of this stable period, 12% of patients had a subtherapeutic INR, a proportion that approximately doubled by eight weeks, and reached 50% after 40 weeks. Use of acenocoumarol (22% of the cohort) doubled the risk of a subtherapeutic INR compared to phenprocoumon and shortened the time to occurrence. The median time to first low INR was 13 weeks versus 51 weeks, respectively. Higher target intensity and use of VKA therapy for prophylaxis of venous thromboembolism (VTE) were also associated with increased risk. The authors additionally found that 30% of low INR episodes were preceded by an invasive procedure, haemorrhage, or elevated INR.

This study highlights the common occurrence of subtherapeutic INR levels and provides a clear estimate of the incidence of low INR in a meticulously constructed inception cohort of patients deemed stable on anticoagulant therapy. Nearly one-fourth of patients in this study experienced a low INR within two months of a period marked by stability (four consecutive INR determinations in the target range). It is notable that 45% of the initial cohort never achieved stable INR which emphasises the challenges inherent to VKA and the gross underestimate of low INR that occurs in routine practice. Importantly, the authors also found that nearly one-third of low INR episodes resulted from clinically justified interventions to minimise risk of haemorrhage, and therefore, reflective of informed clinical care rather than substandard anticoagulation management.

In this study, acenocoumarol was associated with a twofold increase in risk for low INR (adjusted hazard ratio [HR] 2.14) compared to phenprocoumon. Fihn et al. had previously reported more time in the therapeutic range with phenprocoumon compared to acenocoumarol, and phenprocoumon has been shown to exhibit less INR variability over a 24-hour period (19, 20). Potential mechanisms for these observations include differences in pharmacokinetics (the half-life of acenocoumarol is 8–11 hours versus 5–6 days for phenprocoumon), pharmacogenetics, and timing of blood sample collection in relation to dose. Phenprocoumon is less affected by CYP2C9 polymorphisms compared to other VKA (21).

Extrapolating from time-in-range analyses and known effects on INR variability, the authors suggest preferential use of phenprocoumon in clinical practice. However, the study was not

Correspondence to:
Elaine M. Hylek
Research Unit-Section of General Internal Medicine
Boston University School of Medicine
801 Massachusetts Avenue, Crosstown Center – Second Floor
Boston, MA 02118 USA
Tel.: +1 617 414 3743, Fax: +1 617 414 4676
E-mail: ehylek@bu.edu

Financial support:
Dr. Hylek reports receiving grant support from AstraZeneca, serving as an advisor to Boehringer Ingelheim, Bristol-Myers Squibb, Sanofi-Aventis, and The Medicines Company; and participating in clinical symposia sponsored by Bayer and Bristol-Myers Squibb. Dr. Rose is supported by a career development award from the United States Department of Veterans Affairs, Health Services Research and Development Service. The opinions expressed in this commentary are those of the authors and do not represent the official views of the Department of Veterans Affairs.

Received: January 30, 2009
Accepted: February 2, 2009

Prepublished online: February 9, 2009
doi:10.1160/TH09-01-0067

Thromb Haemost 2009; 101: 417–418

designed to assess differences in clinical outcomes (two thromboembolic events occurred) and published data are conflicting on the overall safety of phenprocoumon compared to acenocoumarol. Widely disparate results range from increased major bleeding with phenprocoumon to an isolated increase only in minor bleeding, to no difference in bleeding, and to decreased bleeding compared to acenocoumarol (10, 22–24). Without definitive data on efficacy and safety, treatment recommendations based on surrogate endpoints should be interpreted with caution.

As acknowledged by the authors, the retrospective design of the study prohibited a detailed assessment of other potential risk factors for low INR, particularly medication adherence and dietary change. The authors invoke non-adherence due to patient perception of risk as a potential explanation for the differential rates of low INR by indication for therapy. After adjustment for covariates, patients receiving a VKA for thromboprophylaxis had the highest rate of first low INR (HR 1.88), followed by secondary prevention of VTE (HR 1.36), atrial fibrillation (reference category), and mechanical heart valves (HR 0.69). A more comprehensive accounting of the reasons for unintentional low INR values in routine practice is needed to facilitate interventions to improve time in the therapeutic range.

Understanding the precipitants of low INR is long overdue, and from that standpoint alone, this study is an important contribution. However, many pivotal issues regarding thromboembolism and low INR remain unexplored. Is risk affected by the length of the subtherapeutic period? What is the association between low INR, factor VIIa and molecular markers of thrombin activation? What ultimately drives thrombus formation and embolisation? Is the risk of low INR modified by the clinical context? Do concomitant medications attenuate or magnify the risk of low INR and what is the effect of temporal changes in thrombogenicity? Answers to these questions will provide important insights to fundamental mechanisms. In the interim, optimization of anticoagulant therapy as guided by innovative investigation such as that by Rombouts et al. remains a pressing need.

References

1. Spranca M, Minsk E, Baron J. Omission and comission in judgement and choice. J Exp Soc Psychol 1991; 27: 76–105.

2. Hayward RA, Asch SM, Hogan MM, et al. Sins of omission: getting too little medical care may be the greatest threat to patient safety. J Gen Intern Med 2005; 20: 686–691.

3. Beyth RJ, Quinn LM, Landefeld CS. Prospective evaluation of an index for predicting the risk of major bleeding in outpatients treated with warfarin. Am J Med 1998; 105: 91–99.

4. Fihn SD, Callahan CM, Martin DC, et al. The risk for and severity of bleeding complications in elderly patients treated with warfarin. The National Consortium of Anticoagulation Clinics. Ann Intern Med 1996; 124: 970–979.

5. Gage BF, Yan Y, Milligan PE, et al. Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF). Am Heart J 2006; 151: 713–719.

6. Schulman S, Beyth RJ, Kearon C, et al. Hemorrhagic complications of anticoagulant and thrombolytic treatment: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest 2008; 133: 257S-298S.

7. 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.

8. Choudhry NK, Soumerai SB, Normand SL, et al. Warfarin prescribing in atrial fibrillation: the impact of physician, patient, and hospital characteristics. Am J Med 2006; 119: 607–615.

9. 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.

10. 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.

11. 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.

12. 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.

13. McCormick D, Gurwitz JH, Goldberg RJ, et al. Prevalence and quality of warfarin use for patients with atrial fibrillation in the long-term care setting. Arch Intern Med 2001; 161: 2458–2463.

14. Matchar DB, Samsa GP, Cohen SJ, et al. Improving the quality of anticoagulation of patients with atrial fibrillation in managed care organizations: results of the managing anticoagulation services trial. Am J Med 2002; 113: 42–51.

15. Kucher N, Connolly S, Beckman JA, et al. International normalized ratio increase before warfarin-associated hemorrhage: brief and subtle. Arch Intern Med 2004; 164: 2176–2179.

16. 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.

17. Hylek EM, Go AS, Chang Y, et al. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. New Engl J Med 2003; 349: 1019–1026.

18. Rombouts E, Rosendaal FR, Van der Meer FJ. Subtherapeutic oral anticoagulant therapy: Frequency and risk factors. Thromb Haemost 2009; 101: 552-556.

19. Fihn SD, Gadisseur AA, Pasterkamp E, et al. Comparison of control and stability of oral anticoagulant therapy using acenocoumarol versus phenprocoumon. Thromb Haemost 2003; 90: 260–266.

20. van Geest-Daalderop JH, Hutten BA, Péquériaux NC, et al. The influence on INRs and coagulation factors of the time span between blood sample collection and intake of phenprocoumon or acenocoumarol: consequences for the assessment of the dose. Thromb Haemost 2007; 98: 738–746.

21. Beinema M, Brouwers JR, Schalekamp T, Wilffert B. Pharmacogenetic differences between warfarin, acenocoumarol and phenprocoumon. Thromb Haemost 2008; 100: 1052–1057.

22. van der Meer FJ, Rosendaal FR, Vandenbroucke JP, Briët E Bleeding complications in oral anticoagulant therapy. An analysis of risk factors. Arch Intern Med 1993; 153: 1557–1562.

23. Gadisseur AP, van der Meer FJ, Adriaansen HJ, et al. Therapeutic quality control of oral anticoagulant therapy comparing the short-acting acenocoumarol and the long-acting phenprocoumon. Br J Haematol 2002; 117: 940–946.

24. Visser LE, van Schaik RH, van Vliet M, et al. The risk of bleeding complications in patients with cytochrome P450 CYP2C9*2 or CYP2C9*3 alleles on acenocoumarol or phenprocoumon. Thromb Haemost 2004; 92: 61–66.

A Reengineered Hospital Discharge Program to Decrease Rehospitalization

A Randomized Trial

Brian W. Jack, MD; Veerappa K. Chetty, PhD; David Anthony, MD, MSc; Jeffrey L. Greenwald, MD; Gail M. Sanchez, PharmD, BCPS; Anna E. Johnson, RN; Shaula R. Forsythe, MA, MPH; Julie K. O'Donnell, MPH; Michael K. Paasche-Orlow, MD, MA, MPH; Christopher Manasseh, MD; Stephen Martin, MD, MEd; and Larry Culpepper, MD, MPH

Background: Emergency department visits and rehospitalization are common after hospital discharge.

Objective: To test the effects of an intervention designed to minimize hospital utilization after discharge.

Design: Randomized trial using block randomization of 6 and 8. Randomly arranged index cards were placed in opaque envelopes labeled consecutively with study numbers, and participants were assigned a study group by revealing the index card.

Setting: General medical service at an urban, academic, safety-net hospital.

Patients: 749 English-speaking hospitalized adults (mean age, 49.9 years).

Intervention: A nurse discharge advocate worked with patients during their hospital stay to arrange follow-up appointments, confirm medication reconciliation, and conduct patient education with an individualized instruction booklet that was sent to their primary care provider. A clinical pharmacist called patients 2 to 4 days after discharge to reinforce the discharge plan and review medications. Participants and providers were not blinded to treatment assignment.

Measurements: Primary outcomes were emergency department visits and hospitalizations within 30 days of discharge. Secondary outcomes were self-reported preparedness for discharge and frequency of primary care providers' follow-up within 30 days of discharge. Research staff doing follow-up were blinded to study group assignment.

Results: Participants in the intervention group ($n = 370$) had a lower rate of hospital utilization than those receiving usual care ($n = 368$) (0.314 vs. 0.451 visit per person per month; incidence rate ratio, 0.695 [95% CI, 0.515 to 0.937]; $P = 0.009$). The intervention was most effective among participants with hospital utilization in the 6 months before index admission ($P = 0.014$). Adverse events were not assessed; these data were collected but are still being analyzed.

Limitation: This was a single-center study in which not all potentially eligible patients could be enrolled, and outcome assessment sometimes relied on participant report.

Conclusion: A package of discharge services reduced hospital utilization within 30 days of discharge.

Funding: Agency for Healthcare Research and Quality and National Heart, Lung, and Blood Institute, National Institutes of Health.

Ann Intern Med. 2009;150:178-187. www.annals.org
For author affiliations, see end of text.
ClinicalTrials.gov registration number: NCT00252057.

One in 5 hospitalizations is complicated by postdischarge adverse events (1, 2), some of which may lead to preventable emergency department visits or readmissions. Despite this finding, hospital discharge procedures have not been standardized (3). In addition, the declining presence of primary care providers (PCPs) in hospitals has not been adequately accompanied by systems to ensure that patient data are transferred to subsequent caregivers (4, 5). For example, discharge summaries frequently lack critical data and are not sent to the PCP in a timely fashion (6, 7), resulting in outpatient clinicians being unaware of test results that were pending at discharge (8) and evaluations that were scheduled to be done after discharge not being completed (9). Similarly, patients are often left unprepared at discharge; many do not understand their discharge medications and cannot recall their chief diagnoses (10). With more than 32 million adult discharges in the United States each year (11), these deficiencies in the transition of care increase illness, unnecessary hospital utilization, and cost.

Some peridischarge interventions have shown a reduction in hospital readmission rates and cost (12–14), emergency department visits (15), and postdischarge adverse events (16), whereas some have shown little or no effect (17–20). Peridischarge interventions have also shown improved PCP follow-up and outpatient work-ups (21) and higher patient satisfaction (15). Most of these studies have focused on specific diagnoses (14, 22, 23) or highly selected populations, such as geriatric adults (12, 13, 19, 24). Some have focused on specific aspects of the discharge, such as increasing access to primary care follow-up (25), connecting with transitional nursing services (26), or improving patients' ability to advocate for themselves after discharge (12). To date, no study has evaluated a standard-

See also:

Print
Editors' Notes 179

Web-Only
Appendix Table
Conversion of graphics into slides

Context

Emergency department visits and rehospitalizations are common after hospital discharge.

Contribution

This trial demonstrated that a nurse discharge advocate and clinical pharmacist working together to coordinate hospital discharge, educate patients, and reconcile medications led to fewer follow-up emergency visits and rehospitalizations than usual care alone.

Caution

The trial was conducted at a single center, and not all eligible patients were enrolled.

Implication

A systematic approach to hospital discharges can reduce unnecessary health service use.

—The Editors

ized discharge intervention that includes patient education, comprehensive discharge planning, and postdischarge telephone reinforcement in a general medical population.

In 2004, we began an in-depth examination of hospital discharge, for which we designed a package of services to minimize discharge failures—a process called *reengineered discharge* (RED) (**Table 1**) (3, 27). We did a randomized, controlled trial to evaluate the clinical effect of implementing RED among patients admitted to a general medical service.

Methods

Setting and Participants

We conducted a 2-group, randomized, controlled trial of English-speaking patients 18 years of age or older who were admitted to the medical teaching service of Boston Medical Center, Boston, Massachusetts—a large, urban, safety-net hospital with an ethnically diverse patient population. Patients had 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. We did not enroll patients if they were admitted from a skilled nursing facility or other hospital, transferred to a different hospital service before enrollment, admitted for a planned hospitalization, were on hospital precautions or suicide watch, or were deaf or blind. Boston University's institutional review board approved all study activities.

Randomization

Each morning, a list of admitted patients was reviewed for initial eligibility (hospital location, age, date and time of admission, and previous enrollment). Last names of potential participants were ranked by using a random-number sequence to determine the order in which to approach patients for enrollment. A trained research assistant then approached each patient and further determined eligibility according to inclusion and exclusion criteria (**Figure 1**).

By using block randomization (28) with varying block sizes of 6 and 8, we randomly arranged index cards indicating either the usual care or intervention group. We placed the cards in opaque envelopes labeled consecutively with study numbers. We assigned eligible participants who consented to enrollment to a study group by revealing the concealed index card. This process continued until 2 participants were enrolled each day of the week (or 3 participants if the first 2 participants were randomly assigned to the usual care group). This protocol ensured that research assistants could not selectively choose potential participants for enrollment or predict assignment. Participants randomly assigned to usual care received no further intervention. There were 40 participants in the usual care group and 38 in the intervention group who were enrolled but no longer met inclusion criteria at discharge (most commonly because they were discharged to a nursing facility). Because the primary analysis was by intention to treat, we included these participants in the analysis, with the exception of those who died before index discharge, requested to be removed, or were previously enrolled (**Figure 1**).

Interventions

Nurse discharge advocates (DAs) carried out all aspects of the in-hospital intervention. We hired 6 part-time DAs to work with intervention participants to ensure coverage by 1 DA 7 days a week, 5 hours a day. We trained all DAs to deliver the RED intervention by using a manual containing detailed scripts, observation of relevant clinical interactions, and simulated practice sessions. The primary goals of the DA were to coordinate the discharge plan with the hospital team and educate and prepare the participant for discharge. At admission, the DA completed the RED intervention components outlined in **Table 1**. Additional information about the DA training manual is published elsewhere (3) and can be found on our Web site (www.bu.edu/fammed/projectred/index.html).

With information collected from the hospital team and the participant, the DA created the after-hospital care plan (AHCP), which contained medical provider contact information, dates for appointments and tests, an appointment calendar, a color-coded medication schedule, a list of tests with pending results at discharge, an illustrated description of the discharge diagnosis, and information about what to do if a problem arises. Information for the AHCP was manually entered into a Microsoft Word (Microsoft, Redmond, Washington) template, printed, and spiral-bound to produce an individualized, color booklet designed to be accessible to individuals with limited health literacy. By using scripts from the training manual, the DA used a teach-back methodology (29) to review the contents of the AHCP with the participant. On the day of discharge, the AHCP and discharge summary were faxed to the PCP.

Table 1. **Components of Reengineered Hospital Discharge**

In-hospital component (discharge advocate)

1. Educate patient about relevant diagnoses throughout hospital stay.
2. Make appointments for clinician follow-up and postdischarge testing.
   - Solicit input from patient about convenient date(s) and time(s) for appointments.
   - Coordinate appointments with physicians, testing, and other services.
   - Discuss reason for and importance of physician appointments.
   - Confirm that patient knows location and transportation plan and review barriers to keeping appointments.
3. Discuss with patient any pending in-hospital tests or studies completed and who will follow-up with results.
4. Organize postdischarge services.
   - Be sure patient understands the importance of such services.
   - Make appointments at times convenient for patient.
   - Discuss the details about how to receive each service.
5. Confirm medication plan.
   - Reconcile the discharge medication regimen.
   - Explain what medications to take, emphasizing any changes in the regimen.
   - Review each medication's purpose, how to take it correctly, and important side effects.
   - Be sure the patient has a realistic plan about how to obtain medications.
6. Reconcile the discharge plan with national guidelines and critical pathways.
7. Review appropriate steps for what to do if a problem arises.
   - Instruct how to contact the primary care provider (or coverage) by providing contact numbers for evenings and weekends.
   - Instruct on what constitutes an emergency and what to do in the case of an emergency.
8. Transmit discharge summary to physicians and services accepting responsibility of patient's care that contains the following:
   - Reason for hospitalization with specific principal diagnosis.
   - Important findings.
   - Procedures done and care, treatment, and services provided to patient.
   - Patient's condition at discharge.
   - Complete and reconciled medication list (including allergies).
   - List of acute medical issues, tests, and studies for which confirmed results are pending at the time of discharge and require follow-up.
   - Information about input from consultative services, including rehabilitation therapy.
   - When creating this document, the original source documents—laboratory, radiology, operative reports, and medication administration records—should be in the transcriber's immediate possession and be visible when it is necessary to transcribe information from 1 document to another.
9. Assess the degree of understanding by asking the patient to explain in his or her own words the details of the plan.
   - May require contacting family members who will share in the caregiving responsibilities.

After-hospital care plan

10. Give the patient a written discharge plan at the time of discharge that contains the following:
    - Reason for hospitalization (discharge diagnosis and significant comorbid conditions).
    - Discharge medication list (how and when to take each medication and how to obtain medication).
    - Contact information and picture of primary care provider and discharge advocate.
    - Information for follow-up primary care, specialty care, and outpatient test appointments.
    - Calendar, labeled with scheduled appointments and tests.
    - Information for tests and studies for which confirmed results are not available at the time of discharge.

Pharmacist postdischarge telephone component

11. Call the patient to reinforce discharge plan, review medications, and solve problems.

A clinical pharmacist telephoned the participants 2 to 4 days after the index discharge to reinforce the discharge plan by using a scripted interview. The pharmacist had access to the AHCP and hospital discharge summary and, over several days, made at least 3 attempts to reach each participant. The pharmacist asked participants to bring their medications to the telephone to review them and address medication-related problems; the pharmacist communicated these issues to the PCP or DA.

Outcomes Measures and Follow-up

At the time of recruitment, research assistants collected baseline data, including sociodemographic characteristics; the Short Form-12 Health Survey, Version 2 (30); the depression subscale from the Patient Health Questionnaire-9 (31); and the Rapid Estimate of Adult Literacy in Medicine (32). We calculated the Charlson Comorbidity Index score by using primary and secondary diagnoses recorded on the index admission discharge summary (33). We determined the number of hospital admissions and emergency department visits in the 6 months before index admission through medical record review (Boston Medical Center hospital utilization) and participant report (all other hospital utilization).

The primary end point was the rate of hospital utilization—the total number of emergency department visits and readmissions per participant within 30 days of the index discharge. Any emergency department visit in which a participant was subsequently hospitalized was counted as a readmission. Secondary end points were self-reported preparedness for discharge, rate of primary care follow-up visits, and knowledge of discharge diagnosis. We collected outcome data by review of the hospital's electronic medical records (EMRs) and by contacting participants by telephone 30 days after discharge. We obtained dates of subsequent emergency department visits and readmissions at Boston Medical Center from the EMRs and collected those at other hospitals through participant report. For participants who could not be reached within 60 days after discharge, we assumed that they were alive and relied on hospital EMRs for primary outcomes. Research staff doing follow-up telephone calls and reviewing hospital records were blinded to study group assignment. Discharge advocates and pharmacists recorded time spent working with each participant.

Statistical Analyses

On the basis of unpublished pilot data from the general medical service at Boston Medical Center from July 2003 to June 2004, we estimated that with a readmission incidence rate of 0.197 visit per person per month and an emergency department visit incidence rate of 0.17 visit per person per month (combined hospital utilization rate of 0.367 visit per person per month), we needed to enroll 750 participants to detect an incidence rate reduction of 0.25 visit per person per month in the primary outcome and achieve 80% power, with a 2-sided α level of 0.05.

For outcome data, we followed each participant for 30 days after index discharge. We measured person-time in months, making total person-months equal to the number of participants in each study group. We used the Poisson test and proportions test to test for significance of primary outcomes and secondary outcomes, respectively. We conducted a sensitivity analysis and excluded outliers with high subsequent hospital utilization.

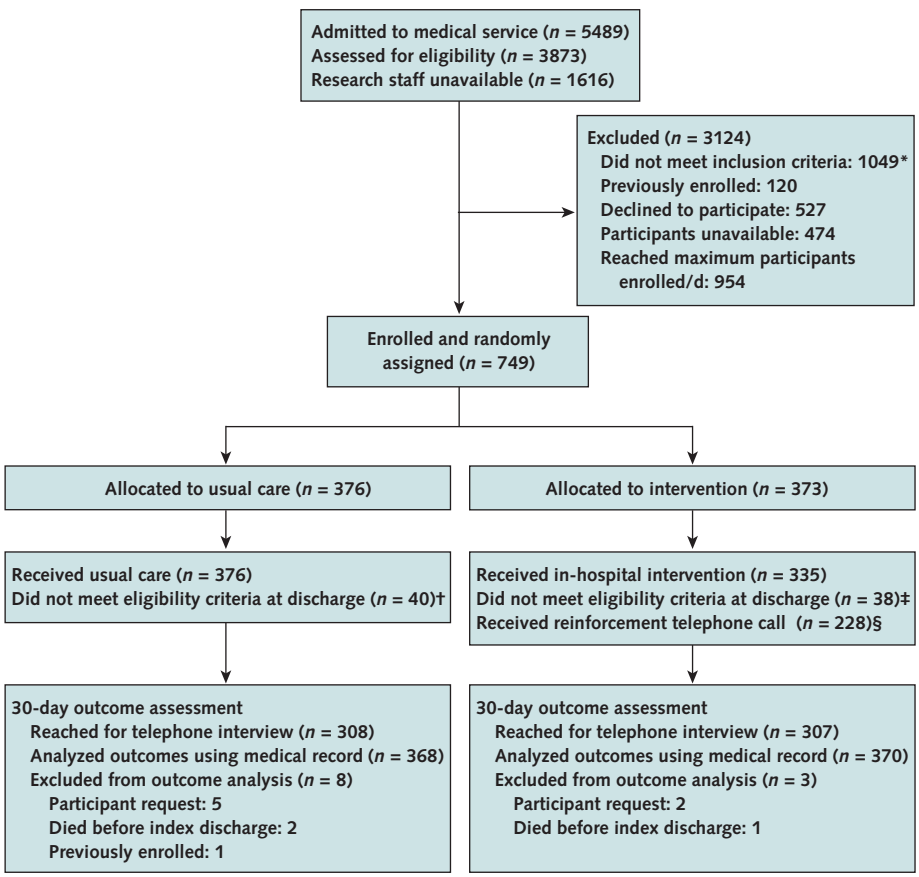
We generated cumulative hazard curves for time to multiple events (emergency department visits and readmissions) and compared them by using a log-rank test. We measured the time-to-event from the index discharge date. This method corresponds to the Wei, Lin, and Weissfeld (34) marginal data model for ordered multiple events, which allows each event to have a separate underlying hazard (35).

We did subgroup analysis with Poisson regression by using total hospital utilization number per participant as the dependent variable. We determined subgroups a priori and included depression diagnosis (36), previous hospital utilization (37), health literacy level (38), sex, and age. To evaluate potential interactions between these variables and the intervention, we included interaction terms in the Poisson regression. We used 2-sided significance tests. We considered P values less than 0.05 to be statistically significant. All data were analyzed with S-Plus, version 8.0 (Insightful, Seattle, Washington), and Intercooled Stata, version 10 (StataCorp, College Station, Texas).

Role of the Funding Source

The Agency for Healthcare Research and Quality and the National Heart, Lung, and Blood Institute, National

Figure 1. **Study flow diagram.**



* Patients did not meet inclusion criteria if they were admitted from or planned discharge to an institutional setting ($n = 74$), planned hospitalization ($n = 3$) or discharge to a non-U.S. community ($n = 5$), were transferred to different hospital service ($n = 8$), did not speak English ($n = 371$) or have a telephone ($n = 71$), were on hospital precautions ($n = 274$) or suicide watch with a sitter ($n = 10$), were unable to consent ($n = 181$), had sickle cell disease as the admitting diagnosis ($n = 38$), had privacy status ($n = 8$), were deaf or blind ($n = 2$), or other ($n = 4$).
† Usual care participants did not meet eligibility criteria if they were discharged to a nursing facility ($n = 28$), were transferred to another hospital service ($n = 1$), were previously enrolled ($n = 1$), died during index admission ($n = 2$), requested to be removed ($n = 5$), or other ($n = 3$).
‡ Intervention participants did not meet eligibility criteria if they were discharged to a nursing facility ($n = 21$), were transferred to another hospital service ($n = 6$), died during index admission ($n = 1$), requested to be removed ($n = 2$), or other ($n = 8$).
§ 107 intervention participants did not receive a reinforcement call because they could not be reached by telephone ($n = 93$), they were readmitted the same or next day ($n = 2$), there was no staffing coverage ($n = 8$), or other ($n = 4$).

Institutes of Health, funded this work. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, and approval of the manuscript.

RESULTS

Study Sample

During the study period from 3 January 2006 to 18 October 2007, we admitted 5489 patients and assessed 3873 for eligibility (**Figure 1**). Because of a lack of available research staff, we did not assess 1616 patients. Of those assessed for eligibility, 1049 did not meet eligibility criteria, 120 were previously enrolled, 527 declined to participate, 474 were unavailable in their hospital room at the time of enrollment, and 954 were not approached because the maximum number of enrolled participants was reached that day. We enrolled and randomly assigned 749 participants: 376 in the usual care group and 373 in the intervention group (**Figure 1**). For primary and secondary outcome analyses, we excluded 11 participants on the basis of participant request ($n = 7$), death before index discharge ($n = 3$), and previous enrollment ($n = 1$), which left 368 in the usual care group and 370 in the intervention group. Baseline demographic and clinical characteristics were similar across study groups (**Table 2**).

Process Measures

In the intervention group, we discharged 346 of 370 (94%) participants with a primary care appointment, 306 (83%) left with an AHCP, 197 (53%) had their medications reconciled with the ambulatory EMR and had their updated medication list included in their AHCP, and 336 (91%) had their discharge information sent to their PCP within 24 hours after discharge. The pharmacist reached 228 (62%) of the intervention participants a median of 4 days (interquartile range [IQR], 3 to 6 days) after discharge and completed medication review with 195 (53%) intervention participants. The pharmacist found that 126 of 195 (65%) intervention participants who completed medication review had at least 1 medication problem and 103 (53%) needed corrective action by the pharmacist, such as contacting the participant's PCP.

In the usual care group, we discharged 127 of 368 (35%) participants with a primary care appointment; data on medication reconciliation and discharge summary transfer to the PCP were unavailable.

Outcome Follow-up

We obtained participant-reported outcome data by telephone for 615 of 738 (83%) participants a median of 32 days (IQR, 30 to 36 days) after discharge. We reached similar proportions of intervention (307 [83%]) and usual care (307 [83%]) group participants ($P = 0.87$). Likewise, similar proportions of intervention (12 [3%]) and usual care (7 [2%]) group participants reported hospital utilization at hospitals other than Boston Medical Center ($P = 0.36$).

Hospital Utilization

In the intervention group, 56 (15.1%) participants had 1 hospital utilization and 24 (6.5%) had more than 1 hospital utilization. These 80 (21.6%) participants had 116 hospital utilizations (61 emergency department visits and 55 readmissions) during 370 person-months of follow-up (0.314 visit per person per month). In the usual care group, 69 (18.8%) participants had 1 hospital utilization and 30 (8.1%) had more than 1 hospital utilization. These 99 (26.9%) participants had 166 visits (90 emergency department visits and 76 readmissions) during 368 person-months of follow-up (0.451 visit per person per month) (**Table 3**). Intervention participants had a lower rate of hospital utilization than usual care participants (incidence rate ratio, 0.695 [95% CI, 0.515 to 0.937]); $P = 0.009$). After we repeated the analysis excluding 1 usual care participant with more than 8 hospital utilizations, hospital utilization between study groups remained statistically significant ($P = 0.028$). Approximately 30% of participants in each study group with any subsequent hospital utilization had more than 1 subsequent hospital utilization.

Figure 2 shows the cumulative hazard curves comparing hospital utilization in the 2 groups over the 30 days after discharge ($P = 0.004$).

Subgroup analyses revealed that the intervention was more effective at reducing hospital utilization for participants with greater hospital utilization in the previous 6 months (P for interaction $= 0.014$).

Secondary Outcomes

Participants receiving the intervention could identify their index discharge diagnosis (242 [79%] vs. 217 [70%] participants; $P = 0.017$) and PCP name (292 [95%] vs. 275 [89%] participants; $P = 0.007$) more often than usual care participants. Intervention participants also reported a higher PCP follow-up rate than usual care participants (190 [62%]) vs. 135 [44%]; $P < 0.001$). Intervention group participants reported being more prepared for discharge at 30 days (**Table 3**). Each component of the AHCP tool was highly rated by intervention participants (**Appendix Table**, available at www.annals.org).

Time Spent Providing Intervention

The DA spent a median of 42.5 minutes (IQR, 30 to 60 minutes) speaking directly with each participant, both collecting participant information and teaching the AHCP booklet. The DA made a median of 3 attempts (IQR, 2 to 5 attempts) per participant to call or page interns. An additional estimated 45 minutes was spent reviewing the participant's EMR, communicating with the medical team, and preparing the AHCP. Therefore, total DA time was estimated to be 87.5 minutes per participant. Estimated weekly DA time (following 14 participants per week) was 20.4 hours or approximately 0.5 full-time equivalent.

The pharmacist postdischarge telephone calls took a

Table 2. **Baseline Participant Characteristics***

Characteristic	Usual Care Group (*n* = 376)	Intervention Group (*n* = 373)
Men, *n (%)*	176 (47)	195 (52)
Mean age, (SD), *y*	49.6 (15.3)	50.1 (15.1)
Race, *n (%)*		
White non-Hispanic	103 (27)	106 (28)
Black non-Hispanic	197 (52)	191 (51)
Hispanic	38 (10)	38 (10)
Other race or mixed race	38 (10)	38 (10)
Annual personal income, *n (%)*		
<$10 000	119 (32)	118 (32)
$10 000–$19 999	61 (16)	73 (20)
$20 000–$49 999	74 (20)	58 (16)
≥$50 000	24 (6.4)	19 (5.1)
Health insurance, *n (%)*		
Private	64 (17)	58 (16)
Medicaid	184 (49)	174 (47)
Medicare	49 (13)	51 (14)
Free Care†	72 (19)	86 (23)
Education level, *n (%)*		
Less than high school	33 (8.8)	22 (5.9)
Some high school	69 (18)	66 (18)
High school graduate or GED	131 (35)	151 (40)
Some college	94 (25)	84 (23)
4-year college graduate or higher	45 (12)	48 (13)
Health literacy level, *n (%)*‡		
Grade 3 or below	56 (15)	58 (16)
Grade 4 to 6	37 (9.8)	39 (10)
Grade 7 to 8	119 (32)	110 (29)
Grade 9 or above	154 (41)	153 (41)
Current employment status, *n (%)*		
Full-time	96 (26)	83 (22)
Part-time	40 (11)	48 (13)
Retired	65 (17)	69 (18)
Disabled	88 (23)	78 (21)
Unemployed	68 (18)	75 (20)
Other	16 (4.3)	16 (4.2)
Homeless in past 3 mo, *n (%)*	40 (11)	35 (9.4)
Mean previous hospital admissions (SD), *n*§	0.71 (1.4)	0.64 (1.1)
Mean previous emergency department visits (SD), *n*§	1.0 (1.8)	0.86 (1.6)
Mean length of stay (SD), *d*	2.6 (3.0)	2.8 (3.4)
PCP at enrollment, *n (%)*	303 (81)	299 (80)
Mean Charlson Comorbidity Index score (SD)‖	1.2 (2.0)	1.2 (1.8)
Mean Physical Component Summary score (SD)¶	40.7 (7.4)	40.1 (7.3)
Mean Mental Component Summary score (SD)¶	46.3 (9.8)	46.7 (9.3)
Major depressive disorder, *n (%)***	52 (14)	69 (18)
Minor depressive disorder, *n (%)***	60 (16)	58 (16)

PCP = primary care provider; REALM = Rapid Estimate for Adult Literacy in Medicine.
* Not all column percentages sum to 100% because of missing values.
† Free Care is a Massachusetts state program for uninsured patients.
‡ Health literacy categories correspond to total REALM scores (32) of grade 3 or below (REALM score, 0–18), grade 4 to 6 (REALM score, 19–44), grade 7 to 8 (REALM score, 45–60), and grade 9 or above (REALM score, 61–66).
§ Previous hospital admissions and emergency department visits include those that occurred within 6 mo before index admission.
‖ The Charlson Comorbidity Index (33) score reflects the cumulative increased likelihood of 1-year mortality. The higher the score, the more severe the comorbid condition. A 35% increase in risk for death is reflected in a 1-point increase in weights. The minimum score is zero; there is no maximum score.
¶ From the Short Form-12 Health Survey (30). The Physical Component Summary score range is 0–100. Mean score for U.S. population is 50 (SD, 10). Higher scores suggest greater physical functional status. The Mental Component Summary score range is 0–100. Mean score for U.S. population is 50 (SD, 10). Higher scores suggest greater mental functional status.
** Determined by using the Patient Health Questionnaire-9, a 9-item, 4-point Likert scale, standard scoring algorithm to screen for major and minor depression (31).

median of 14 minutes (IQR, 10 to 19 minutes), with 10 (6 to 18) additional minutes spent on call preparation, missed calls, and resolving problems identified during calls. It took the pharmacist a median of 2 attempts (IQR, 1 to 3 attempts) to reach participants by telephone. Median total pharmacist time was approximately 26 minutes (IQR, 18 to 36 minutes) per participant. Estimated weekly pharmacist time (following 14 participants per week) was 6.1 hours or approximately 0.15 full-time equivalent.

Outcome Cost Analysis

The actual cost of emergency department visits totaled $21 389 for the usual care group and $11 285 for the intervention group. The actual cost of hospital visits totaled

Table 3. **Primary and Secondary Outcomes**

Variable	Usual Care Group	Intervention Group	*P* Value
Primary outcomes ≤30 d after index hospitalization			
Patients, *n*	368	370	–
Hospital utilizations, *n (visits/patient/mo)**	166 (0.451)	116 (0.314)	0.009
IRR (95% CI)	1.0	0.695 (0.515–0.937)	–
Emergency department visits, *n (visits/patient/mo)*	90 (0.245)	61 (0.165)	0.014
IRR (95% CI)	1.0	0.674 (0.476–0.955)	–
Readmissions, *n (visits/patient/mo)*	76 (0.207)	55 (0.149)	0.090
IRR (95% CI)	1.0	0.720 (0.445–1.164)	–
Secondary outcomes†			
Patients, *n*	308	307	–
Able to identify discharge diagnosis, *n (%)*	217 (70)	242 (79)	0.017
Able to identify PCP name, *n (%)*	275 (89)	292 (95)	0.007
Visited PCP, *n (%)*	135 (44)	190 (62)	<0.001
How well were your questions answered before you left the hospital?‡	108 (62)	129 (77)	0.002
How well did you understand your appointments after you left the hospital?‡	219 (79)	254 (86)	0.025
How well did you understand how to take your medications after leaving the hospital?‡	233 (83)	264 (89)	0.049
How well did you understand your main problem or diagnosis when you left the hospital?‡	167 (57)	198 (66)	0.014
How prepared were you to leave the hospital?‡	163 (55)	197 (65)	0.013

IRR = incidence rate ratio; PCP = primary care provider.
* Defined as the sum of emergency department visits plus rehospitalizations. An emergency department visit that leads to a rehospitalization is counted only as a rehospitalization.
† Denominators were participants who were reached at the 30-day follow-up phone call and those who answered questions.
‡ Questions were answered on a 5-point Likert scale. The percentage reflects participants who responded with either of the top 2 categories on the scale ("very prepared" or "prepared").

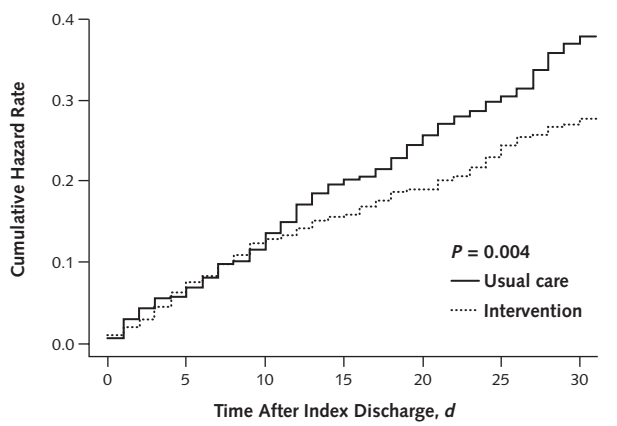
$412 544 for the usual care group and $268 942 for the intervention group. Follow-up PCP appointments were given an estimated cost of $55, on the basis of costs from an average hospital follow-up visit at Boston Medical Center. The estimated cost of primary care outpatient visits within 30 days after discharge totaled $8906 for 44% of 368 usual care participants and $12 617 for 62% of 370 intervention participants. The difference between study groups in total cost (combining actual hospital utilization cost and estimated outpatient cost) for 738 participants was $149 995—an average of $412 per person who received the intervention. This represents a 33.9% lower observed cost for the intervention group.

DISCUSSION

The RED intervention decreased hospital utilization (combined emergency department visits and readmissions) within 30 days of discharge by about 30% among patients on a general medical service of an urban, academic medical center. More intervention group participants reported seeing their PCP for follow-up within 30 days and reported higher levels of preparedness for discharge. In addition, the intervention was successful in reducing hospital utilization among participants who frequently used hospital services. These data support implementation of a comprehensive program for hospital discharge among similar hospitals.

Our intervention includes patient-centered education, comprehensive discharge planning, and postdischarge rein-

Figure 2. **Cumulative hazard rate of hospital utilization for 30 days after index hospital discharge.**



Cumulative Events*						
Usual care	30	59	87	111	132	164†
Intervention	30	51	63	75	97	110†

* The denominators for the events were 433 for usual care and 397 for intervention. This represents the number of discharges for each group, which includes index discharges and discharges from all subsequent admissions. At each discharge, the participant is returned to the risk pool. The denominator is thus constant during the entire 30 days.
† Two events for the usual care group and 6 events for the intervention group were removed from this analysis because the date of admission was missing.

forcement and is practical and easily applied to general medical patients. The RED intervention has 3 core elements: the DA, the AHCP, and the follow-up telephone call by those of the pharmacist. Because these elements were bundled, we could not clearly determine the degree that each part contributed to the effects demonstrated. No previous studies have evaluated this trio of interventions together, although the roles of the DA and the pharmacist build on previous literature (12, 15, 16, 19). For example, peridischarge nursing support services have been shown to improve discharge for patients with heart failure (14, 23, 39, 40). Coleman and colleagues (12) used a nurse "transition coach" to demonstrate reduced readmissions at 30 and 90 days among elderly patients. Naylor and coworkers (13, 19) found that nurse specialists involved during and after discharge also effectively reduced acute readmissions.

Several studies have analyzed pharmacist interventions. Dudas and colleagues (15) randomly assigned patients to receive a telephone call by a pharmacist after discharge and demonstrated fewer emergency department visits. Schnipper and coworkers (16) used pharmacist counseling before and after discharge and showed reductions in preventable adverse drug events and medication-related readmissions and emergency department visits. Al-Rashed and colleagues (41) found that predischarge pharmacist-based counseling for elderly patients followed by a postdischarge home visit resulted in fewer unplanned primary care visits and fewer readmissions.

The techniques used to teach the AHCP, its content, and its format (for example, pictures, color, and large font) were informed by the literature on limited health literacy (42, 43). Overall, the intervention improved patient comprehension of key elements of self-care: 30 days after discharge, intervention participants were better able to identify their primary diagnosis and reported better understanding of their diagnosis, medications, and appointments. The content, format, and teaching of discharge preparation tools deserve further attention because few studies have assessed the effect of patient education on subsequent hospital utilization.

Because intervention group participants were more likely to report seeing their PCPs after discharge and we transmitted discharge information to PCPs promptly after discharge, the intervention optimized the chance that PCPs could identify and address outstanding issues. In addition, the pharmacist follow-up telephone call identified any problems that a patient was having after discharge and relayed those issues to the PCP. Previous studies have suggested that improved access to community-based follow-up alone may not be enough to reduce hospital readmissions (18, 25). We provide evidence that when combined with other elements of RED, improving PCP follow-up may help reduce hospital utilization.

Implementing this discharge intervention required about 1.5 hours of nursing time and 30 minutes of pharmacist time per participant. Because some of the DA activities were redundant with those of existing hospital personnel, implementation of the RED intervention using existing hospital staff would require less time per patient. Also, because information was manually entered to create each AHCP, hospital information technology solutions could be developed to make this process more efficient. Despite this, we demonstrated hospital utilization cost savings averaging $412 per discharge. These figures do not include the cost of the intervention, which involved 0.5 full-time equivalent for a nurse and 0.15 full-time equivalent for a clinical pharmacist. If adopted broadly, this intervention could produce substantial effects on health care financing (44). However, an important challenge for programs like RED is that health providers, who are best situated to implement such a program, may have no financial incentive to do so. Hospitals serving capitation-based patient populations may benefit financially from reducing unneeded rehospitalization. Under the fee-for-service scheme, the payer will benefit even after paying the full cost of the intervention. Hospitals will also benefit from decreasing the rehospitalization rate as an important quality-improvement target, and investment in strategies proven to work will be attractive to payers. The National Quality Forum is reviewing new metrics of quality care surrounding readmission rates (45), and programs like RED may be used to improve health care organizations' quality ratings.

Our study has limitations. Because of staffing limitations, we were only able to enroll 2 to 3 participants per day, and we could not enroll participants on some weekends and holidays. Because of the nature of our urban, underserved patient population and exclusion of patients coming from nursing homes, the study sample was younger and had fewer comorbid conditions than those in other studies; thus, our results may not be generalizable to all patient groups. Also, we relied on participant self-report for outcomes that we could not gather from EMRs, notably data on PCP follow-up and visits at hospitals other than Boston Medical Center. Previous studies have suggested that patient reports of emergency department and hospital use correlate well with electronic records from 6 months to 1 year (46, 47). Ritter and colleagues (48) demonstrated that patients tended to underreport outpatient visits over 6 months compared with electronic charts and found no demographic or health-related predictors of underreporting. In our case, recall bias should be expected to be nondifferential because our study was randomized, we reached both study groups equally, and outcome assessors were blinded to study assignment. We assumed that study participants not reached by telephone for an outcome assessment were alive for 30 days after the index discharge, and we relied on hospital EMRs to gather primary outcomes. Therefore, we did not capture deaths or hospital utilizations at institutions other than Boston Medical Center for this limited number of participants. For the cost analysis, we could not determine a generalizable cost for

the intervention because costs vary widely by institution and location. Similarly, we could not estimate the downstream cost implications of avoided emergency department visits and readmissions. Still, we present the actual costs for 3 important types of directly related medical utilization. The cost of hospital utilization and outpatient visits also cannot be easily generalized. Our goal is to provide the direct comparison that can be made for these key costs between study groups, and we observed a 33.9% reduction in these costs.

In summary, the RED program successfully reduced hospital utilization, improved patient self-perceived preparation for discharge, and increased PCP follow-up. In 2007, the National Quality Forum Consensus Standards Maintenance committee identified hospital discharge as a critical area for improvement. The resulting National Quality Forum "Safe Practice" was based largely on the principles of the RED program (49). Our study provides data supporting the implementation of the discharge standards promoted by the National Quality Forum.

From Boston University School of Medicine, Boston Medical Center, Boston, Massachusetts, and Brown University Alpert Medical School, Pawtucket, Rhode Island.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Acknowledgment: The authors thank Caroline Hesko, MPH, for data collection and Lynn Schipelliti, RN, and Mary Goodwin, RN, for implementing the intervention. The RED study employed Ms. Hesko, Ms. Schipelliti, and Ms. Goodwin.

Grant Support: By Agency for Healthcare Research and Quality grants 1UC1HS014289-01 and 1U18HS015905-01 (Dr. Jack) and National Heart, Lung, and Blood Institute, National Institutes of Health, grant 1 R01 HL081307-01 (Dr. Jack).

Potential Financial Conflicts of Interest: *Grants received:* B.W. Jack (Agency for Healthcare Research and Quality; National Heart, Lung, and Blood Institute, National Institutes of Health).

Reproducible Research Statement: *Study protocol and statistical code:* Available from Dr. Jack (e-mail, brian.jack@bmc.org). *Data set:* Available through written agreement with Dr. Jack (e-mail, brian.jack@bmc.org).

Requests for Single Reprints: Brian W. Jack, MD, Department of Family Medicine, Boston University School of Medicine, Boston Medical Center, 1 BMC Place, Dowling 5 South, Boston, MA 02118; e-mail, brian.jack@bmc.org.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW.** The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med. 2003;138:161-7. [PMID: 12558354]

2. **Forster AJ, Clark HD, Menard A, Dupuis N, Chernish R, Chandok N, et al.** Adverse events among medical patients after discharge from hospital. CMAJ. 2004;170:345-9. [PMID: 14757670]

3. **Greenwald JL, Denham CR, Jack BW.** The hospital discharge: a review of a high risk care transition with highlights of a reengineered discharge process. J Patient Saf. 2007;3:97-106.

4. **Wachter RM.** Hospitalists in the United States—mission accomplished or work in progress? N Engl J Med. 2004;350:1935-6. [PMID: 15128892]

5. **Moore C, Wisnivesky J, Williams S, McGinn T.** Medical errors related to discontinuity of care from an inpatient to an outpatient setting. J Gen Intern Med. 2003;18:646-51. [PMID: 12911647]

6. **Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW.** Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA. 2007;297:831-41. [PMID: 17327525]

7. **van Walraven C, Seth R, Austin PC, Laupacis A.** Effect of discharge summary availability during post-discharge visits on hospital readmission. J Gen Intern Med. 2002;17:186-92. [PMID: 11929504]

8. **Roy CL, Poon EG, Karson AS, Ladak-Merchant Z, Johnson RE, Maviglia SM, et al.** Patient safety concerns arising from test results that return after hospital discharge. Ann Intern Med. 2005;143:121-8. [PMID: 16027454]

9. **Moore C, McGinn T, Halm E.** Tying up loose ends: discharging patients with unresolved medical issues. Arch Intern Med. 2007;167:1305-11. [PMID: 17592105]

10. **Makaryus AN, Friedman EA.** Patients' understanding of their treatment plans and diagnosis at discharge. Mayo Clin Proc. 2005;80:991-4. [PMID: 16092576]

11. **Levit K, Ryan K, Elixhauser A, Stranges E, Kassed C, Coffey R.** HCUP facts and figures: Statistics on hospital-based care in the United States in 2005. Accessed at www.hcup-us.ahrq.gov/reports.jsp on 26 May 2008.

12. **Coleman EA, Parry C, Chalmers S, Min SJ.** The care transitions intervention: results of a randomized controlled trial. Arch Intern Med. 2006;166:1822-8. [PMID: 17000937]

13. **Naylor MD, Brooten D, Campbell R, Jacobsen BS, Mezey MD, Pauly MV, et al.** Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA. 1999;281:613-20. [PMID: 10029122]

14. **Anderson C, Deepak BV, Amoateng-Adjepong Y, Zarich S.** Benefits of comprehensive inpatient education and discharge planning combined with outpatient support in elderly patients with congestive heart failure. Congest Heart Fail. 2005;11:315-21. [PMID: 16330907]

15. **Dudas V, Bookwalter T, Kerr KM, Pantilat SZ.** The impact of follow-up telephone calls to patients after hospitalization. Am J Med. 2001;111:26S-30S. [PMID: 11790365]

16. **Schnipper JL, Kirwin JL, Cotugno MC, Wahlstrom SA, Brown BA, Tarvin E, et al.** Role of pharmacist counseling in preventing adverse drug events after hospitalization. Arch Intern Med. 2006;166:565-71. [PMID: 16534045]

17. **Siu AL, Kravitz RL, Keeler E, Hemmerling K, Kington R, Davis JW, et al.** Postdischarge geriatric assessment of hospitalized frail elderly patients. Arch Intern Med. 1996;156:76-81. [PMID: 8526700]

18. **Holland R, Lenaghan E, Harvey I, Smith R, Shepstone L, Lipp A, et al.** Does home based medication review keep older people out of hospital? The HOMER randomised controlled trial. BMJ. 2005;330:293. [PMID: 15665005]

19. **Naylor M, Brooten D, Jones R, Lavizzo-Mourey R, Mezey M, Pauly M.** Comprehensive discharge planning for the hospitalized elderly. A randomized clinical trial. Ann Intern Med. 1994;120:999-1006. [PMID: 8185149]

20. **Shepperd S, Parkes J, McClaren J, Phillips C.** Discharge planning from hospital to home. Cochrane Database Syst Rev. 2004:CD000313. [PMID: 14973952]

21. **Balaban RB, Weissman JS, Samuel PA, Woolhandler S.** Redefining and redesigning hospital discharge to enhance patient care: a randomized controlled study. J Gen Intern Med. 2008;23:1228-33. [PMID: 18452048]

22. **Andersen HE, Schultz-Larsen K, Kreiner S, Forchhammer BH, Eriksen K, Brown A.** Can readmission after stroke be prevented? Results of a randomized clinical study: a postdischarge follow-up service for stroke survivors. Stroke. 2000;31:1038-45. [PMID: 10797163]

23. **Phillips CO, Wright SM, Kern DE, Singa RM, Shepperd S, Rubin HR.** Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: a meta-analysis. JAMA. 2004;291:1358-67. [PMID: 15026403]

24. **Evans RL, Hendricks RD.** Evaluating hospital discharge planning: a random-

ized clinical trial. Med Care. 1993;31:358-70. [PMID: 8464252]
25. **Weinberger M, Oddone EZ, Henderson WG.** Does increased access to primary care reduce hospital readmissions? Veterans Affairs Cooperative Study Group on Primary Care and Hospital Readmission. N Engl J Med. 1996;334:1441-7. [PMID: 8618584]
26. **Naylor MD, Brooten D.** The roles and functions of clinical nurse specialists. Image J Nurs Sch. 1993;25:73-8. [PMID: 8449535]
27. **Anthony D, Chetty VK, Kartha A, McKenna K, DePaoli RM, Jack B.** Patient safety at time of discharge—an example of a multifaceted process evaluation. In: Henriksen K, Battles JB, Marks ESL, Lewin DI, eds. Advances in Patient Safety: From research to Implementation. vol. 2. Concepts and Methodology. Rockville, MD: Agency for Healthcare Research and Quality; 2005.
28. **Lachin JM, Matts JP, Wei LJ.** Randomization in clinical trials: conclusions and recommendations. Control Clin Trials. 1988;9:365-74. [PMID: 3203526]
29. **Paasche-Orlow MK, Riekert KA, Bilderback A, Chanmugam A, Hill P, Rand CS, et al.** Tailored education may reduce health literacy disparities in asthma self-management. Am J Respir Crit Care Med. 2005;172:980-6. [PMID: 16081544]
30. **Ware J Jr, Kosinski M, Keller SD.** A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34:220-33. [PMID: 8628042]
31. **Kroenke K, Spitzer RL, Williams JB.** The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16:606-13. [PMID: 11556941]
32. **Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al.** Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med. 1993;25:391-5. [PMID: 8349060]
33. **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;40:373-83. [PMID: 3558716]
34. **Wei LJ, Lin DY, Weissfeld L.** Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc. 1989;84:1065-73.
35. **Therneau TM, Grambsch PM.** Modeling Survival Data: Extending the Cox Model. New York: Springer Science and Business Media; 2000.
36. **Büla CJ, Wietlisbach V, Burnand B, Yersin B.** Depressive symptoms as a predictor of 6-month outcomes and services utilization in elderly medical inpatients. Arch Intern Med. 2001;161:2609-15. [PMID: 11718593]
37. **van Walraven C, Mamdani M, Fang J, Austin PC.** Continuity of care and patient outcomes after hospital discharge. J Gen Intern Med. 2004;19:624-31. [PMID: 15209600]
38. **Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, et al.** Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. Am J Public Health. 2002;92:1278-83. [PMID: 12144984]
39. **Blue L, Lang E, McMurray JJ, Davie AP, McDonagh TA, Murdoch DR, et al.** Randomised controlled trial of specialist nurse intervention in heart failure. BMJ. 2001;323:715-8. [PMID: 11576977]
40. **Tsilimingras D, Bates DW.** Addressing postdischarge adverse events: a neglected area. Jt Comm J Qual Patient Saf. 2008;34:85-97. [PMID: 18351193]
41. **Al-Rashed SA, Wright DJ, Roebuck N, Sunter W, Chrystyn H.** The value of inpatient pharmaceutical counselling to elderly patients prior to discharge. Br J Clin Pharmacol. 2002;54:657-64. [PMID: 12492615]
42. **Paasche-Orlow MK, Schillinger D, Greene SM, Wagner EH.** How health care systems can begin to address the challenge of limited literacy. J Gen Intern Med. 2006;21:884-7. [PMID: 16881952]
43. **Paasche-Orlow MK, Parker RM.** Improving the effectiveness of patient education: a focus on limited health literacy. In: King TE, Wheeler MB, eds. Medical Management of Vulnerable and Underserved Patients: Principles, Practice, and Populations. New York: McGraw Hill; 2007:101-9.
44. **Friedman B, Basu J.** The rate and cost of hospital readmissions for preventable conditions. Med Care Res Rev. 2004;61:225-40. [PMID: 15155053]
45. **National Quality Forum.** National quality forum endorses consensus standards for quality hospital care: Patient safety in hospitals focus of 48 NQF-endorsed measures. Accessed at www.qualityforum.org/news/releases/051508-endorsed-measures.asp on 21 May 2008.
46. **Marquis KH, Marquis MS, Newhouse JP.** The measurement of expenditures for outpatient physician and dental services: methodological findings from the health insurance study. Med Care. 1976;14:913-31. [PMID: 824510]
47. **Roberts RO, Bergstralh EJ, Schmidt L, Jacobsen SJ.** Comparison of self-reported and medical record health care utilization measures. J Clin Epidemiol. 1996;49:989-95. [PMID: 8780606]
48. **Ritter PL, Stewart AL, Kaymaz H, Sobel DS, Block DA, Lorig KR.** Self-reports of health care utilization compared to provider records. J Clin Epidemiol. 2001;54:136-41. [PMID: 11166528]
49. **National Quality Forum.** Safe practices for better healthcare 2006 update. A consensus report. Washington, DC: National Quality Forum; 2007.

Current Author Addresses: Drs. Jack, Chetty, Manasseh, Martin, and Culpepper; Ms. Johnson; Ms. Forsythe; and Ms. O'Donnell: Department of Family Medicine, Boston University School of Medicine, Boston Medical Center, 1 BMC Place, Dowling 5 South, Boston, MA 02118.

Dr. Anthony: Department of Family Medicine, Brown University Alpert Medical School, 111 Brewster Street, Pawtucket, RI 02860.

Dr. Greenwald: Hospital Medicine Unit, Department of Medicine, Boston University School of Medicine, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118.

Dr. Sanchez: Boston Medical Center, 88 East Newton Street, Boston, MA 02118.

Dr. Paasche-Orlow: Section of General Internal Medicine, Boston University School of Medicine, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118.

Author Contributions: Conception and design: B.W. Jack, V.K. Chetty, D. Anthony, J.L. Greenwald, G.M. Sanchez, C. Manasseh, L. Culpepper.

Analysis and interpretation of the data: B.W. Jack, V.K. Chetty, D. Anthony, J.L. Greenwald, A.E. Johnson, S.R. Forsythe, J.K. O'Donnell, M.K. Paasche-Orlow, S. Martin, L. Culpepper.

Drafting of the article: B.W. Jack, D. Anthony, J.L. Greenwald, G.M. Sanchez, A.E. Johnson, S.R. Forsythe, J.K. O'Donnell, M.K. Paasche-Orlow.

Critical revision of the article for important intellectual content: B.W. Jack, V.K. Chetty, J.L. Greenwald, G.M. Sanchez, A.E. Johnson, S.R. Forsythe, J.K. O'Donnell, M.K. Paasche-Orlow, L. Culpepper.

Final approval of the article: B.W. Jack, V.K. Chetty, J.L. Greenwald, G.M. Sanchez, A.E. Johnson, S.R. Forsythe, J.K. O'Donnell, M.K. Paasche-Orlow, C. Manasseh, S. Martin, L. Culpepper.

Provision of study materials or patients: B.W. Jack, A.E. Johnson, M.K. Paasche-Orlow, C. Manasseh.

Statistical expertise: V.K. Chetty, M.K. Paasche-Orlow.

Obtaining of funding: B.W. Jack, D. Anthony.

Administrative, technical, or logistic support: B.W. Jack, V.K. Chetty, D. Anthony, A.E. Johnson, J.K. O'Donnell, C. Manasseh.

Collection and assembly of data: B.W. Jack, G.M. Sanchez, A.E. Johnson, S.R. Forsythe, J.K. O'Donnell.

Appendix Table. **Evaluation of the AHCP by Intervention Participants 30 Days After Discharge**

Question	Participant Response, *n (%)**
In the past 4 weeks, how often did you refer to your AHCP?†	
Daily	31 (12)
Frequently	39 (14)
Occasionally	75 (28)
Once or twice	78 (29)
Never	21 (7.8)
How useful was the AHCP booklet?†	
Extremely useful	46 (17)
Very useful	92 (34)
Moderately useful	50 (19)
A little bit useful	40 (15)
Not at all useful	10 (3.7)
What was the most helpful part of the AHCP?†	
RED medication schedule	51 (19)
Appointment page	41 (15)
Medical provider contact information	26 (9.7)
Appointment calendar	24 (8.9)
Diagnosis information	29 (11)
Other	29 (11)
How helpful was the RED medication calendar?‡	
Extremely helpful	26 (17)
Very helpful	46 (30)
Moderately helpful	15 (9.7)
A little bit helpful	10 (6.5)
Not at all helpful	4 (2.6)

AHCP = after-hospital care plan; RED = reengineered discharge.

* Not all percentages sum to 100% because of missing values (participants did not answer the question—they either declined or ended the call early).

† The denominator was intervention participants who were reached for the 30-day follow-up telephone call and received an AHCP (n = 269).

‡ The denominator was intervention participants who were reached for the 30-day follow-up telephone call and received an RED medication calendar in their AHCP (n = 155).

Cost-Effectiveness of Screening for Unhealthy Alcohol Use with %Carbohydrate Deficient Transferrin: Results From a Literature-Based Decision Analytic Computer Model

Alok Kapoor, Kevin L. Kraemer, Kenneth J. Smith, Mark S. Roberts, and Richard Saitz

Background: The %carbohydrate deficient transferrin (%CDT) test offers objective evidence of unhealthy alcohol use but its cost-effectiveness in primary care conditions is unknown.

Methods: Using a decision tree and Markov model, we performed a literature-based cost-effectiveness analysis of 4 strategies for detecting unhealthy alcohol use in adult primary care patients: (i) Questionnaire Only, using a validated 3-item alcohol questionnaire; (ii) %CDT Only; (iii) Questionnaire followed by %CDT (Questionnaire-%CDT) if the questionnaire is negative; and (iv) No Screening. For those patients screening positive, clinicians performed more detailed assessment to characterize unhealthy use and determine therapy. We estimated costs using Medicare reimbursement and the Medical Expenditure Panel Survey. We determined sensitivity, specificity, prevalence of disease, and mortality from the medical literature. In the base case, we calculated the incremental cost-effectiveness ratio (ICER) in 2006 dollars per quality-adjusted life year ($/QALY) for a 50-year-old cohort.

Results: In the base case, the ICER for the Questionnaire-%CDT strategy was $15,500/QALY compared with the Questionnaire Only strategy. Other strategies were dominated. When the prevalence of unhealthy alcohol use exceeded 15% and screening age was <60 years, the Questionnaire-%CDT strategy costs less than $50,000/QALY compared to the Questionnaire Only strategy.

Conclusions: Adding %CDT to questionnaire-based screening for unhealthy alcohol use was cost-effective in our literature-based decision analytic model set in typical primary care conditions. Screening with %CDT should be considered for adults up to the age of 60 when the prevalence of unhealthy alcohol use is 15% or more and screening questionnaires are negative.

Key Words: Carbohydrate Deficient Transferring, Alcohol Use, Primary Care.

THE UNITED STATES Preventive Services Task Force (USPSTF) recommends screening for unhealthy alcohol use, including at-risk drinking, problem drinking, alcohol abuse, and alcohol dependence (U.S. Preventive Services Task Force, 2004). The National Institute on Alcohol Abuse and Alcoholism (NIAAA) defines at-risk drinking as >14 drinks per week or 5 or more drinks on a single occasion for men and >7 drinks per week or 4 or more drinks on a single occasion for women or those aged over 65 (National Institute of Alcohol and Alcoholism, 2005). Among the multiple questionnaires available to screen for unhealthy alcohol use, AUDIT-Consumption (AUDIT-C) offers a 3-item inventory of the quantity and frequency of unhealthy alcohol use (Fiellin et al., 2000). It is generally sensitive (81 to 94%) and specific (82 to 86%) but can be subjected to inaccurate or untruthful responses.(Bradley et al., 2003, 2007; Gordon et al., 2001).

Serum biomarkers such as the %carbohydrate deficient transferrin (%CDT) test can provide objective evidence of unhealthy alcohol use. Heavy daily consumption of alcohol for 2 weeks or more triggers a positive test. Studies have found that %CDT has high specificity (77 to 100%) but variable sensitivity (10 to 85%) (Berner et al., 2006; Koch et al., 2004; Miller and Anton, 2004). Performance estimates vary depending on whether the goal of screening is to detect very heavy drinking (>60 to 80 g of ethanol or more than 5 to 7 drinks per day) or the at-risk amounts defined above.

%CDT has been widely used in Europe (Miller and Anton, 2004) and the United States Food and Drug Administration

From the Hospital Medicine Unit, Section of General Internal Medicine (AK), Boston Medical Center, Boston, Massachusetts; Division of General Internal Medicine, University of Pittsburgh School of Medicine (KLK, KJS, MSR), Pittsburgh, Pennsylvania; Section of Decision Sciences and Clinical Systems Modeling, University of Pittsburgh School of Medicine (KJS, MSR), Pittsburgh, Pennsylvania; and Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Boston Medical Center, and Department of Epidemiology and Youth Alcohol Prevention Center, Boston University School of Public Health (RS), Boston, Massachusetts.

Received for publication October 28, 2008; accepted March 16, 2009.

Reprint requests: Alok Kapoor, 801 Massachusetts Avenue – Second Floor, Boston, MA 02118; Fax: 617-414-4676; E-mail: alok.kapoor@BMC.org

DOI: 10.1111/j.1530-0277.2009.00974.x

approved a %CDT assay in 2001 for detecting chronic heavy alcohol consumption (Food and Drug Administration, 2007). Although %CDT has the advantage of being an objective test, it has a low positive predictive value if used as the sole screening tool (Aertgeerts et al., 2001) and is expensive (Coulton et al., 2006). Nevertheless, multiple experts have suggested that alcohol biomarkers like %CDT can be useful in clinical settings including primary care (Miller et al., 2006). There are few published data about who should be tested with %CDT, how %CDT should be integrated with questionnaires such as AUDIT-C, and for which patient groups %CDT screening is most cost-effective.

Decision analysis is a systematic explicit, quantitative way of making decisions in health care that can lead to both enhanced decisions and better outcomes for patients (Hunink et al., 2001). In its most basic form, the modeler builds a decision tree and inputs the probability and value of each outcome derived from some combination of original data and the published medical literature. The modeler then associates costs and effects with each outcome. Cost-effectiveness may then be calculated as the cost divided by the benefit, the latter being expressed in disease-specific units such as the number of strokes averted, quality-adjusted life years (QALYs) gained, or in monetary units itself. When monetary units are used to calculate benefit, the analysis is termed a cost-benefit analysis. Dillie and colleagues (2005) assessed the cost-benefit of %CDT screening in primary care but focused on diabetic and hypertensive patients and did not assess the value of adding %CDT to established screening questionnaires. We conducted a comprehensive, literature-based decision analysis computer model to evaluate the cost-effectiveness of %CDT testing both alone and combined with questionnaire to screen for unhealthy alcohol use in primary care.

METHODS

Framework and Decision Model

We conducted our analysis following the recommendations of the Panel on Cost-Effectiveness in Health and Medicine (Panel on Cost-Effectiveness), (Russell et al., 1996; Siegel et al., 1996; Weinstein et al., 1996). We adopted a societal perspective, including costs and effects incurred both by patients receiving care and institutions providing care. The target population included adult men and women (ages 18 to 100 years) in primary care. The time horizon, or period over which costs and effects were aggregated, was from screening until death or age 100 years.

We modeled 4 strategies for detecting unhealthy alcohol use in primary care using TreeAge Pro 2007 Suite software (TreeAge Software Inc., Williamstown, MA). The 4 strategies were: (i) Questionnaire Only, using AUDIT-C; (ii) %CDT only; (iii) Questionnaire followed

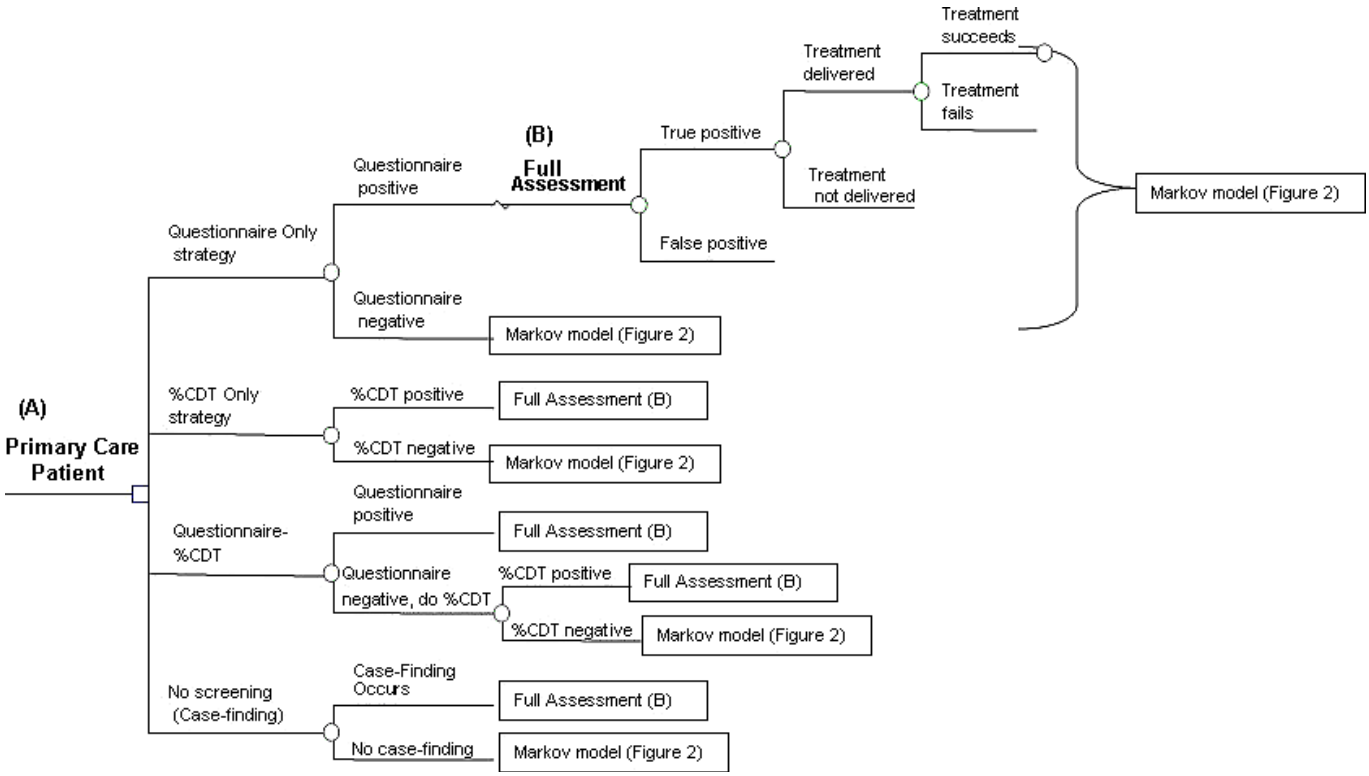


Fig. 1. Decision tree of four strategies to screen for unhealthy alcohol use in primary care. (**A**) A clinician can screen a primary care patient once for unhealthy alcohol by one of four strategies. (**B**) Once a patient tests positive by a screening test, he or she moves into the full assessment phase. In the full assessment, clinicians ask questions to determine if the test result is a true or false positive and determine if there is an alcohol disorder. Then, there is a probability that the clinician delivers a treatment (brief intervention for at-risk drinking or abuse) or refers to specialty alcohol treatment for alcohol dependence. Finally, there is a chance that the treatment succeeds, placing the unhealthy drinker into a safer health state. Patients then enter the Markov model in one of six health states (see Fig. 2).

by %CDT (Questionnaire-%CDT) if the questionnaire is negative; and (iv) No Screening (case-finding only in which the clinician does not screen but discovers unhealthy use through the course of caring for a patient; Fig. 1). The Questionnaire Only strategy models current guidelines from national organizations including the USPSTF and NIAAA. The Questionnaire-%CDT strategy allows direct assessment of the cost-effectiveness of adding %CDT to the current recommended questionnaire-based screening strategy.

The initial part of the decision model simulated one-time screening, assessment, and intervention for the spectrum of unhealthy alcohol use, including at-risk drinking, alcohol abuse, and dependence. For AUDIT-C, we considered a score of ≥5 (out of a possible 12 points) positive for a man and ≥2 positive for a woman (Bradley et al., 2003, 2007; Gordon et al., 2001). The cut-off for %CDT was 2.6% as recommended by the manufacturers (Axis Shield ASA, Oslo, Norway; Berner et al., 2006). We assumed all screen-positive patients completed a full clinical assessment (i.e., the gold standard) for unhealthy alcohol use. Following this assessment, we modeled the probability that a patient would receive a brief intervention and the probability that a delivered intervention was successful. For alcohol dependence, we also modeled the probability that a patient would receive formal alcohol treatment which includes a course of cognitive behavioral (or similarly effective) therapy. If brief intervention were successful, a patient with at-risk drinking or alcohol abuse converted to safe drinking. If formal alcohol treatment worked, a patient with alcohol dependence converted to a recovery state. Such conversions are also possible in untreated groups. This "screening effect" is a beneficial reduction in drinking that occurs from the mere detection and verification of disease. Because we are uncertain if this effect would occur in real world (as opposed to research) conditions, we only applied this effect to the No Screening strategy in the base case, biasing the analysis against screening strategies.

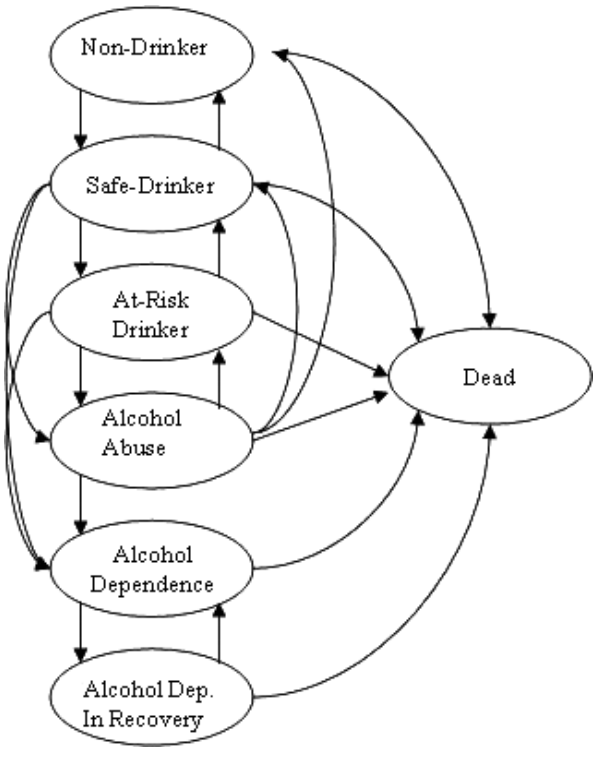


Fig. 2. Markov model of health states defined by alcohol consumption (Non-Drinker, Safe Drinker, At-Risk Drinker) or the presence of an alcohol diagnosis (Alcohol Abuse, Alcohol Dependence, Alcohol Dependence in Recovery).

Patients finished the initial alcohol screening and intervention part of the model in 1 of 6 mutually exclusive alcohol-related health states (Fig. 2). We then used a Markov model to track the transitions among these 6 health states from the time of screening/intervention until death.

The time frame of the first part of the model (Fig. 1) is the time it takes for screening, assessment, and treatment to occur (i.e., ranging from a single clinic visit for an individual who screens negative or screens positive and receives brief intervention at the screening visit, and up to multiple visits for an individual who is alcohol-dependent and receives alcohol treatment). The "time frame" of the second part of the model, the Markov model (Fig. 2), is from the time immediately after screening/assessment/intervention until death or the age of 100 years.

Probabilities and Health State Utilities

For each probability estimate we searched Medline (1950 to spring 2007), spoke with experts, and consulted the documents of authorities such as the Centers for Disease Control and Prevention. For each parameter, we chose the highest quality evidence available but when there was uncertainty about the true value among equally good data, we made a conservative choice that biased against %CDT screening. For %CDT performance, we included data from a study of primary care patients in Germany screened for at-risk drinking (Berner et al., 2006). There was no such study from a population in the U.S. Because research on %CDT has mostly involved testing for very heavy alcohol use (e.g., >80 g ethanol or 6 drinks/day for a man and >40 g ethanol or 3 drinks/day for a woman), we also calculated the cost-effectiveness of %CDT testing using discrete diagnostic performance estimates for detecting very heavy drinkers compared with the remaining unhealthy drinking population. For these %CDT performance estimates, we used a large, multi-center international trial of patients from a range of recruitment settings (not primary care; Holder, 1998) (Table 1).

For efficacy of brief intervention, we used estimates from a 5 to 10 minute brief intervention trial (Ockene et al., 1999). We operationalized efficacy with 2 variables for the transition from the at risk drinking or abuse state to safe state. The transition rate in the group receiving screening and brief intervention was 39%. The same transition in the group receiving screening alone was 28% indicating a net effect of 11%.

We derived health state transition probabilities from 2 well-established longitudinal studies conducted in the U.S. (Kerr et al., 2002; Schuckit et al., 2001). To calculate survival, we consulted the published literature (Arias, 2006; Dawson, 2000). To calculate quality-adjusted survival, we multiplied survival by health state utilities previously measured by our group (Kraemer et al., 2005). Utilities represent a degree of preference for 1 health state (scored between 0 and 1) versus a perfect health state (utility of 1).

Costs

We calculated initial costs for screening and treatment in 2006 U.S. dollars. Our estimates for screening costs represent current Medicare reimbursement for physician time and lab testing (Centers for Medicare & Medicaid, 2006a,b). To aggregate direct health care costs in the future, we used data from the Medical Expenditure Panel Survey (MEPS) (Agency for Healthcare Research and Quality, 2004). We also did not include cost incurred to people injured by the index patient nor productivity gains for treated patients experiencing improved health. For alcohol dependence and abuse, we assigned a multiplier to the baseline annual costs provided by MEPS. There are conflicting reports about the costs for at-risk drinking and so we assumed at-risk drinking had no effect on direct health care costs (multiplier = 1.0; Dillie et al., 2005; Holder, 1998; Mertens et al., 2005).

Table 1. Parameter List, Baseline Estimate, Range for Sensitivity Analysis, and Comment

Parameter	Baseline	Range	Comments/citations
Probabilities			
Demographics			
Initial age of cohort	50	18–80	Implies all individuals at the same initial age
Prevalence of unhealthy drinking (%)			Includes at-risk drinkers, alcohol abuse, and
50-year-old cohort	22	10–40	alcohol dependence; prevalence; values are
25-year-old cohort	28	15–45	a weighted average assuming 50% of
75-year-old cohort	6	1–20	cohort are male (Manwell et al., 1998)
Prevalence of abuse (%)	11	7–15	Prevalence for the base case, 50-year-old patient
Prevalence of at-risk drinking (%)	4	2–10	
Prevalence of dependence (%)	8	4–12	
Test performance and prevalence			
Sensitivity of AUDIT-C questionnaire in women	81	50–99	Sensitivity and specificity to detect >7
Specificity of AUDIT-C questionnaire in women	86	50–100	drinks/wk or 4 or more drinks/day ± DSM IV disorder at a specificity of 86% (Bradley et al., 2003)
Sensitivity of AUDIT-C questionnaire in men	94	50–99	Sensitivity and specificity to detect >16
Specificity of AUDIT-C questionnaire in men	82	50–100	drinks/wk (Gordon et al., 2001)
Sensitivity of %CDT (men and women combined)	34	10–99	Sensitivity and specificity to detect individuals
Specificity of %CDT (men and women combined)	94	50–100	at-risk or more unhealthy drinking (reference standard AUDIT > 8) (Berner et al., 2006)
% delivery, treatment, and screening effects			
Delivery of brief intervention (BI) (%)			
At-risk	39	0–59	Probability of BI delivery by primary care
Abuse	59	39–71	provider after positive screen by
Dependence	71	59–100	questionnaire or %CDT; in the source publication (Burman et al., 2004), the 10-item AUDIT was used to categorize disease severity
% individuals with dependence who follow up for alcohol treatment after brief intervention or usual care	40	10–90	Preliminary data from our own work for receipt of "alcohol assistance" (ASAP Study Clinical Trials Identifier NCT00183105) after brief intervention or usual care. (Note: In the model, 0% of alcohol dependents reduce their drinking after brief intervention alone; alcohol dependents must proceed to alcohol treatment before any benefit occurs)
% at-risk drinkers or drinkers with alcohol abuse achieving low risk drinking after brief intervention	39	0–75	Percentage transitioning from at-risk or alcohol abuse to low risk drinking (e.g., within guidelines/suggested limits) after BI (Ockene et al., 1999)
% dependent drinkers achieving low risk drinking after alcohol treatment	41	0–80	Percentage transitioning from dependence to recovery after alcohol treatment (Project MATCH authors, 1997)
Screening effect parameters (%)			Percentage transitioning from an unhealthy to healthy state after detection but without treatment; in the base case, the screening effects only applied to the No Screening strategy but was explored in sensitivity analyses.
Abuse or at-risk to safe	28	0–50	From control arm of a randomized controlled trial for BI (Ockene et al., 1999)
Dependence to recovery	14	0–50	No trial data for this parameter found; 1 year spontaneous probability of transition used (Schuckit et al., 2001)
Percent follow up of a positive %CDT result	50	10–90	Composite probability that provider notifies patient and patient returns for full assessment
Percent refusal of %CDT	0	0–100	Assumed this value is zero; refusal implies no change in cost-effectiveness
Utilities			
Nondrinker (age <65)	0.91	0.74–1.00	For all unhealthy states, we used standard gamble utilities measured in the community (Kraemer et al., 2005). For the utility of individuals with age > 65 in Nondrinker or Safe state, we used a generic, published utility for the well elderly (Gold et al., 1998)

Table 1. (Continued)

Parameter	Baseline	Range	Comments/citations
Safe drinker (age <65)	0.86	0.74–100	
Nondrinker or safe drinker (age 65 or more)	0.84	0.74–1.00	
At-risk drinker (all ages)	0.80	0.74–1.00	
Abuse drinker (all ages)	0.74	0.65–0.80	
Dependent drinker (all ages)	0.65	0.40–0.80	
Recovery (all ages)	0.81	0.74–0.86	
Hazard ratios			
Nondrinker	1	1	Hazard ratio (of dying) for drinking state compared with nondrinker reference state (Dawson, 2000)
Safe drinker	0.8	0.33–2.00	
At-risk	0.92	0.50–4.00	
Abuse drinker	1.07	0.50–1.00	
Dependent drinker	1.42	0.50–5.00	
Recovery	1.18	0.50–4.00	
Costs			
Initial costs (in $US)			
Questionnaire	3	0–50	CPT 99203 - 1 of 30 minutes assuming physician billed by time (Centers for Medicare & Medicaid, 2006b)
%CDT	38	20–150	Medicare reimbursement CPT 82373 + venipuncture + 0.5 hour wages = $25 (Centers for Medicare & Medicaid, 2006a)+ $3 (Centers for Medicare & Medicaid, 2006a) + $9.50 (Bureau of Labor Statistics, 2005)
Full assessment following positive questionnaire or case-finding	33	0–250	CPT 99203 - 10 of 30 minutes assuming physician billed based on time (Centers for Medicare & Medicaid, 2006b)
Full assessment following positive %CDT	128	0–250	Follow-up visit (CPT 99213) + 3 hours wage + daily travel for patient (Centers for Medicare & Medicaid, 2006b)
Brief intervention following positive questionnaire or case-finding	26	0–200	CPT 99203 – 7.5 of 30 minutes (Centers for Medicare & Medicaid, 2006b)
Brief intervention following positive %CDT	0	0–200	No additional cost after new visit cost which permits 15–25 minutes of provider time
Cost of alcohol dependence treatment	1,077	200–10,000	Includes provider costs as discussed by Cisler and colleagues (1998) + lost wages for 6 sessions, total 18 hours (Bureau of Labor Statistics, 2005) + 6 days of travel
Cost of hourly wages lost for patient	19	5–30	National mean wage adjusted for inflation (Bureau of Labor Statistics, 2005)
Cost of daily travel paid by patient	15	0–30	Estimated by authors
Future costs (in $US)			Mean annual cost including out of pocket and third party disbursements for individuals without alcohol disorder (only listed for men) (Agency for Healthcare Research and Quality, 2004)
Age 18–44	1,942	500–4,710	
Ages 45–64	4,710	1,942–9657	
Age 65 and over	9,657	4,710–20,00	
Future cost multipliers			
Dependent drinker	2	1–3	Implies that future annual cost will be twice that of nondrinkers for each year lived with the disorder (Blose and Holder, 1991; Holder, 1998)
Abuse	1.5	0.5–2	Implies that future annual cost will be 1.5 fold that of nondrinkers (estimated by authors)
Recovery, at-risk, safe	1.0	1.0–2.0	Implies that future annual cost will be the same as that of nondrinkers (estimated by authors)

Markov Model Calibration

Transitions out of the Nondrinker state or into dependence become infrequent after the third decade of life. To calibrate transition rates for the above health states, we performed 1,000 model simulations starting with the previously mentioned transition rates. From the simulations, we calculated the proportion transitioning out of abstinence over a lifetime in 4 age and gender strata for which there was published information available (Adams and Schoenborn, 2006). Similarly, we calculated the proportions transitioning into dependence in these strata and compared these values to the published information (Dawson et al., 2005). In both cases the published data reflects the "background" rate of discovery, treatment, and transition to safer health states (either Safe,

Abstinent, or Recovery depending on the starting point) and represents the natural history of unhealthy alcohol diagnosis and treatment prior to availability of %CDT screening. We then repeated simulations, each time adjusting transition rates until the proportions approximated the published data. We did not find information about the age of transition for other health states and used the published (uncalibrated) rates in those cases.

Analysis

We calculated the incremental cost-effectiveness ratio (ICER) as the difference in costs between the least expensive strategy and the next least expensive strategy divided by the difference in their effectiveness (measured both in unadjusted life years and QALYs). The ratio is expressed as how much additionally it costs (in dollars) to achieve an additional QALY. Policy makers are interested in the ICER value because it accounts for the fact that there was a less expensive option when making a selection from competing programs (Hunink et al., 2001). Interpreting the results of cost-effectiveness analysis can be problematic, making it difficult to decide whether to adopt a diagnostic test or treatment. The threshold for adoption in the Unites States is thought to be somewhere between $20,000/QALY and $100,000/QALY, with a threshold of $50,000/QALY frequently proposed (Bell et al., 2006). In the base case, we examined the cost-effectiveness in a hypothetical cohort of 50 year olds. We then repeated this analysis in 25 year olds and 75 year olds. We discounted all future health costs and QALYs by 3%.

We conducted 1-way and selected 2-way sensitivity analyses to assess the influence of uncertainty in individual parameter values on the ICER for the Questionnaire-%CDT strategy compared with the Questionnaire Only strategy. We also performed probabilistic sensitivity analysis (a process that involves specifying distributions for model input parameters and then sampling simultaneously from these distributions to assess the joint effect of input parameter uncertainty).

RESULTS

Model predicted proportions of transition out of the nondrinker state and of dependence onset calibrated well, falling within 2% of published probabilities (Table A1). The base-case model results for cost, effectiveness, and ICER are shown in Table 2. The No Screening and %CDT Only strategies were both more costly and less effective than (i.e., they were dominated by) the other strategies in the base case and the other scenarios described in Table 2. In the base case 50-year-old cohort, the Questionnaire-%CDT strategy cost $15,500/QALY compared with the Questionnaire Only strategy. The ICER for the same comparison in a 25-year-old cohort was substantially lower at $3,380/QALY and in a 75-year-old cohort was substantially higher at $243,000/QALY (Table 3).

Table 3. Incremental Cost-Effectiveness Ratio for Questionnaire-%CDT Strategy Versus Questionnaire Only as a Function of Age, Use of Life Years, Patient Costs Inclusion, and Screening Effects

Age	ICER ($/QALY)	ICER ($/LY)	ICER without patient costs included ($/QALY)	ICER with screening effects included in intervention strategies[a]
25	3,380	18,200	Dominates[b]	Dominates
50	15,500	58,600	5,030	2,290
75	243,000	441,000	164,000	128,000

ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; LY, life year.

[a]Screening effect pertains to the transition from unhealthy to healthy state after positive screening but without formal intervention. In the base case, we applied this effect only to the No Screening strategy. In sensitivity analysis, we applied this to all strategies.

[b]Dominates implies that the Questionnaire-%CDT strategy cost less and gained more QALYs compared with the Questionnaire Only strategy.

Sensitivity Analyses

The baseline ICER estimate was sensitive to the percentage of at-risk drinkers or drinkers with alcohol abuse achieving safe drinking levels after brief intervention, questionnaire and %CDT sensitivity, age at screening, prevalence of unhealthy alcohol use, and the follow-up rate of positive %CDT results (Fig. 3). The Questionnaire-%CDT strategy dominated the Questionnaire Only strategy when questionnaire sensitivity was less than 76% or %CDT sensitivity was greater than 64%. In order for the ICER to cross the $50,000/QALY threshold, the % at-risk drinkers or drinkers with alcohol abuse achieving low risk drinking after brief intervention would have to drop from 39% to 17%, the sensitivity of %CDT would have to drop from 34% to 17%, or follow-up after ordering the %CDT test would have to drop from 50% to 23%. In the analysis that looked at the effect of using discrete %CDT diagnostic performance data for very heavy alcohol use, the ICER for the Questionnaire-%CDT strategy increased to $27,800/QALY and the %CDT only strategy was still dominated.

Table 2. Cost and Effectiveness of 4 Strategies for Alcohol Screening in a Cohort of 50-Year-Old Primary Care Patients

Strategy	Cost (in $)	Incremental cost (in $)	Effectiveness (in QALYs)	Incremental effectiveness (in QALYs)	ICER (in $/QALY)
Questionnaire Only	143,568		16.013		
%CDT Only	144,104	523	15.984	−0.031	(Dominated[a])
Questionnaire-%CDT	143,581	13	16.014	0.001	15,500
No Screening	143,780	199	15.999	−0.015	(Dominated)

ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

[a]Dominated implies this strategy cost more and is less effective.

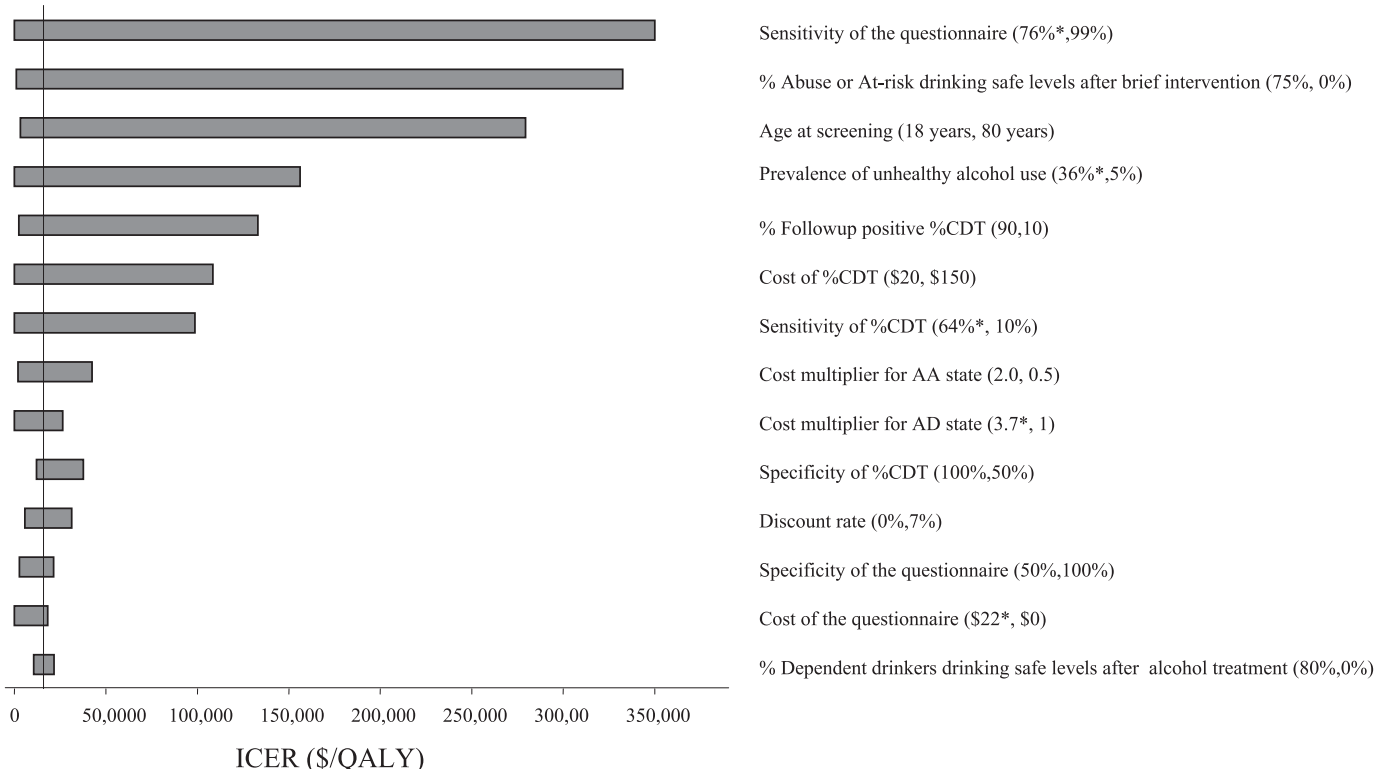


Fig. 3. Tornado diagram of one-way sensitivity analyses on important model parameters. The horizontal bars indicate the incremental cost-effectiveness ratio (ICER) of the Questionnaire-%CDT strategy compared with the Questionnaire Only strategy. Values in parentheses for each variable represent the range over which sensitivity analysis was performed as shown in Table 1. If the Questionnaire-%CDT strategy the Questionnaire Only strategy, then one end of the range is replaced by the value at which dominance occurs and is shown by an asterisk; the vertical line represents the ICER using the baseline value. An asterisk denotes the value for which the Questionnaire-%CDT strategy dominates the Questionnaire Only strategy. ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year.

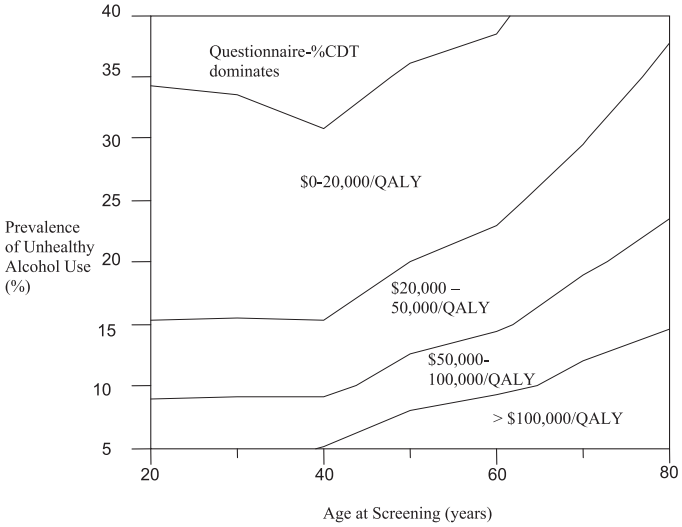


Fig. 4. Two-way sensitivity analysis on the incremental cost-effectiveness ratio (ICER) as a function of the prevalence of unhealthy alcohol use and age at screening. The $/QALY values indicate the ICER range for the Questionnaire-%CDT strategy compared to Questionnaire Only strategy at specific combinations. QALY, quality adjusted life year.

In a 2-way sensitivity analysis varying age and prevalence of unhealthy alcohol use (Fig. 4), the ICER for Questionnaire-%CDT remains below a $50,000/QALY threshold if unhealthy alcohol use is ≥15% and screening age is ≤60. The Questionnaire Only strategy dominated the No Screening strategy in virtually all age cohorts. Probabilistic sensitivity analysis indicates that at the $50,000/QALY threshold, the Questionnaire-%CDT strategy was favored in 64% of the simulations compared to the Questionnaire Only which was favored 35% of the time (Fig. 5).

DISCUSSION

Our analysis indicates that adding %CDT to questionnaire based one-time screening is cost-effective in typical primary care conditions. In 50 year olds, the Questionnaire-%CDT strategy costs $15,500 per QALY gained when compared to the Questionnaire Only strategy. Compared with the Questionnaire Only strategy, the Questionnaire-%CDT strategy was favored at a threshold of $50,000/QALY when the prevalence of unhealthy alcohol use exceeded 15% and the age at screening was < 60 years. The Questionnaire Only strategy dominated the No Screening strategy in virtually all age cohorts. Screening with the %CDT test alone was not cost-effective.

We provide evidence for intensifying screening to detect unhealthy alcohol use in primary care by adding a %CDT test when questionnaire screening is negative. Our analysis

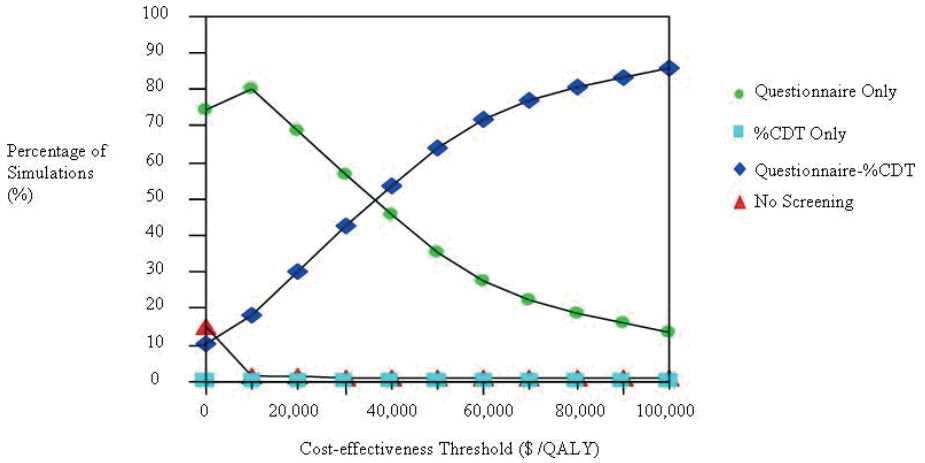


Fig. 5. Percentage of simulations for which four strategies to screen for unhealthy alcohol use are cost-effective in a 50-year-old cohort of primary care patients. QALY, quality adjusted life year.

differs from a cost-benefit analysis conducted by Dillie and colleagues (2005) which suggested that adding %CDT to physician interview is cost saving, meaning that it achieved better outcome at a lower cost. We found that adding %CDT to questionnaire screening was cost-effective (achieved better outcome but at a higher, though generally acceptable, cost) but not cost-saving. Some people informally summarize such interventions as "good value." Unlike the previous analysis, we accounted for incomplete follow-up of %CDT, potential need for a second visit to address the positive result, and poor provider performance in delivering treatment (Burman et al., 2004; Saitz et al., 2003). Our analysis also included the long term cost and effects of screening with %CDT and included patient time costs and out of pocket expenses (i.e., the societal perspective). The base case ICER value of $15,500/QALY compares favorably with the cost-effectiveness of other currently accepted screening programs—e.g. one-time HIV screening ($33,000/QALY) (Paltiel et al., 2006) or colonoscopy every 10 years compared with annual fecal occult blood testing or no screening ($12,000 to 18,000/life year; Pignone et al., 2002).

Our conclusion also differed from Coulton and colleagues (2006) who found the cost per patient screened was 20-fold greater for %CDT compared with questionnaire based screening in Welsh males. This group did not, however, analyze the incremental cost-effectiveness of adding %CDT to questionnaire-based screening as in the current study and did not account for potential downstream costs saved, mortality avoided, and quality of life improved.

There are several limitations to this work. There is no single estimate for the prevalence of unhealthy alcohol use in primary care. Prevalence varies by gender, race, ethnicity, geography, and duration but has been reported in multiple studies (Manwell et al., 1998; Taj et al., 1998) to be more than 20% using the current NIAAA definition we adopted for our analysis. We chose prevalence estimates from a study by Manwell and colleagues (1998) in which 21,282 patients in Wisconsin were screened for unhealthy alcohol use. That study reported a *90-day prevalence* of unhealthy alcohol use of 23%, combined for all ages and both genders. The study included one of the largest U.S. primary care samples available and it provided data about the spectrum of unhealthy alcohol use. Our sensitivity analysis suggests the Questionnaire-%CDT strategy would still be cost-effective (at the $50,000/QALY threshold) in a lower prevalence scenario when the age at screening is less than 60 years.

There is also no single way to administer brief intervention and therefore no single estimate for the transition rate from at-risk drinking to safe drinking levels. We believe our choice for the value of the transition rate (i.e., 39%) after brief intervention was conservative. Other studies such as Project Treat (Fleming et al., 1997) using a longer initial BI and incorporating follow-up contacts have described the effect to be larger but we believe a one-time, 5 to 10 minute intervention was the one most likely to resemble how physicians actually conduct brief intervention. Comparisons with other brief intervention trials such as those included in a recent systematic review (Beich et al., 2003) are limited by exclusion of subjects with lower levels of risky alcohol.

Another limitation of the Markov modeling technique we used is that the transition probabilities depend only on the current state and not on the history of past states. For example, individuals in the at-risk drinking state in a given 1-year cycle had the same probability of transitioning into other states regardless of their drinking state in prior cycles. We did not have information about the rate of transition from safe to at-risk drinking for an individual with a prior history of at-risk drinking compared with someone without this history. We obtained information about transitions in drinking behavior from a study by Kerr and colleagues (2002) based on the National Health Nutrition Examination Survey. Transition rates provided by Kerr and colleagues represent the rate of transitions at the aggregate level. This includes individuals with and without a prior history of at-risk drinking. We

therefore believe that the transition rates we used are an accurate representation of the transitions from safe to at-risk drinking, at the aggregate level. For individuals with a history of alcohol dependence, this "amnestic" property of Markov models was mitigated by the high rate of relapse built into the Recovery state.

Other limitations include absence of conditional diagnostic test performance data for %CDT (i.e., the sensitivity and specificity in a population already having tested negative by questionnaire). We believe biomarker screening has a diagnostic performance that is independent from questionnaire performance. Our estimate for %CDT performance to detect unhealthy alcohol use was a conservative choice from the limited trials set in general primary care. Had we chosen to use discrete diagnostic performance data for detecting very heavy drinkers, as in the previously mentioned sensitivity analysis, the economic implications would not have changed substantially.

We also did not have information about the effectiveness of brief intervention or alcohol treatment in a group testing negative by questionnaire. Brief intervention is likely to be less successful in a group testing negative by questionnaire. Such individuals may be feigning low risk use or they may be infrequent risky drinkers, and in either case less likely to change, although the exact magnitude of the differential effectiveness is not known.

We did not have information about the clinical effect of ordering a blood test in patients denying unhealthy alcohol use. Patients who take offense from being asked to confirm their reported drinking behavior with %CDT may decide not to discuss their alcohol use or other medical problems as freely with their provider. They may even decide to sever relations with this provider. We believed the frequency of these untoward consequences would be low and therefore did not model any costs for the deterioration or discontinuation in the patient–provider relationship. We feel the decision to not to model these costs, however, was still a conservative choice given that mention of objective corroboration of a person's report with %CDT will likely prime an admission of unhealthy use for a large percentage of primary care patients, thereby obviating the need and cost for the test. In addition, at least 1 study suggests that the use of %CDT can provide motivation for some patients to reduce their alcohol use (Fleming et al., 2004). The exact direction of the bias imposed by our balanced modeling assumptions (i.e., that all patients who screened negative by the questionnaire would undergo the blood test and that no patient would voluntarily disclose their drinking status upon broaching the issue of biomarker screening) is unknown and represents area for future inquiry.

Lastly, we did not model all possible consequences of a false-positive %CDT result. There is no consensus for the workup of elevated %CDT results and false-positive results may occur in patients underreporting alcohol use (i.e., the gold standard interviews used to assess performance are imperfect). Future research should assess %CDT performance and treatment effectiveness in a cohort testing negative by questionnaire, patient and provider acceptability of the Questionnaire-%CDT strategy, and the implications of false-positive %CDT results.

In conclusion, adding %CDT to questionnaire based one-time screening for unhealthy alcohol use was cost-effective in typical primary care conditions and, at minimum, clinicians should screen all patients with a questionnaire. Some clinicians may consider ordering %CDT after a negative screening questionnaire for adults up to age 60 when the prevalence of unhealthy alcohol use is 15% or more. However, despite its cost-effectiveness, issues around effectiveness of brief intervention in a questionnaire negative group, patient acceptability of blood testing in this same group, and management of false-positive results should be better studied before we can recommend widespread use of %CDT.

APPENDIX

Table A1. Predicted versus published proportions for transition out of nondrinker state and transition into dependence in 4 age and gender strata

	Proportion transitioning out of nondrinker state[a]		Proportion transitioning into dependence in the future compared with the total ever being dependent[b]	
Stratum	Model predicted proportion	Published proportion	Model predicted proportion	Published proportion
25-year-old men	0.56	0.55	0.33	0.33
25-year-old women	0.35	0.37		
50-year-old men	0.05	0.06	0.06	0.05
50-year-old women	0	0		

[a]Analysis adjusting for gender as published in the National Household Interview Survey (Adams and Schoenborn, 2006).

[b]Analysis unadjusted for gender as published in the National Epidemiologic Survey of Alcohol Related Conditions (Dawson et al., 2005).

ACKNOWLEDGMENTS

This study was supported by National Research Service Award T32 HP 10028-08 and the National Institute on Alcohol Abuse and Alcoholism R01 AA12617.

REFERENCES

Adams PF, Schoenborn CA (2006) Health behaviors of adults: United States, 2002–2004. Vital Health Stat Ser 10:1–140.

Aertgeerts B, Buntinx F, Ansoms S, Fevery J (2001) Screening properties of questionnaires and laboratory tests for the detection of alcohol abuse or dependence in a general practice population. Br J Gen Pract 51:206–217.

Agency for Healthcare Research and Quality (2004) Medical Expenditure Panel Survey. Available at: http://www.meps.ahrq.gov/mepsweb/. Accessed June 13, 2007.

Arias E (2006) United States life tables, 2003. Natl Vital Stat Rep 54:1–40.

Beich A, Thorsen T, Rollnick S (2003) Screening in brief intervention trials targeting excessive drinkers in general practice: systematic review and meta-analysis. BMJ 327:536–542.

Bell CM, Urbach DR, Ray JG, Bayoumi A, Rosen AB, Greenberg D, Neumann PJ (2006) Bias in published cost effectiveness studies: systematic review. BMJ 332:699–703.

Berner MM, Bentele M, Kriston L, Manz C, Clement H-W, Harter M, Mundle G (2006) DOVER and QUVER-new marker combinations to detect and monitor at-risk drinking. Alcohol Clin Exp Res 30:1372–1380.

Blose JO, Holder HD (1991) The utilization of medical care by treated alcoholics: longitudinal patterns by age, gender, and type of care. J Sub Abuse 3:13–27.

Bradley KA, Bush KR, Epler AJ, Dobie DJ, Davis TM, Sporleder JL, Maynard C, Burman ML, Kivlahan DR (2003) Two brief alcohol-screening tests from the Alcohol Use Disorders Identification Test (AUDIT): validation in a female veterans affairs patient population. Arch Intern Med 163:821–829.

Bradley KA, Debenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR (2007) AUDIT-C as a brief screen for alcohol misuse in primary care. Alcohol Clin Exp Res 31:1208–1217.

Bureau of Labor Statistics (2005) National Compensation Survey. Available at: http://data.bls.gov/PDQ/outside.jsp?survey=nc. Accessed June 13, 2007.

Burman ML, Kivlahan D, Buchbinder M, Broglio K, Zhou XH, Merrill JO, Mcdonell MB, Fihn SD, Bradley KA, Ambulatory Care Quality Improvement Project I (2004) Alcohol-related advice for Veterans Affairs primary care patients: who gets it? Who gives it? J Stud Alcohol 65:621–630.

Centers for Medicare & Medicaid (2006a) Clinical Diagnostic Laboratory Fee Schedule. Available at: http://www.cms.hhs.gov/ClinicalLabFeeSched. Accessed June 6, 2007.

Centers for Medicare & Medicaid (2006b) Physician Fee Schedule. Available at: http://www.cms.hhs.gov/pfslookup/02_PFSearch.asp. Accessed June 6, 2007.

Cisler R, Holder HD, Longabaugh R, Stout RL, Zweben A (1998) Actual and estimated replication costs for alcohol treatment modalities: case study from Project MATCH. J Stud Alcohol 59:503–512.

Coulton S, Drummond C, James D, Godfrey C, Bland JM, Parrott S, Peters T, Stepwice Research T. (2006) Opportunistic screening for alcohol use disorders in primary care: comparative study. BMJ 332:511–517.

Dawson DA (2000) Alcohol consumption, alcohol dependence, and all-cause mortality. Alcohol Clin Exp Res 24:72–81.

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.

Dillie KS, Mundt M, French MT, Fleming MF (2005) Cost-benefit analysis of a new alcohol biomarker, carbohydrate deficient transferrin, in a chronic illness primary care sample. Alcohol Clin Exp Res 29:2008–2014.

Fiellin DA, Reid MC, O'Connor PG (2000) Screening for alcohol problems in primary care: a systematic review. Arch Intern Med 160:1977–1989.

Fleming MF, Barry KL, Manwell LB, Johnson K, London R (1997) Brief physician advice for problem alcohol drinkers. A randomized controlled trial in community-based primary care practices. JAMA 277:1039–1045.

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.

Food and Drug Administration (2007) Device Listing. Available at: http://www.fda.gov/search/databases.html. Accessed August 1, 2007.

Gold MR, Franks P, McCoy KI, Fryback DG (1998) Toward consistency in cost-utility analyses: using national measures to create condition-specific values. Medical Care 36:778–792.

Gordon AJ, Maisto SA, Mcneil M, Kraemer KL, Conigliaro RL, Kelley ME, Conigliaro J (2001) Three questions can detect hazardous drinkers. J Fam Pract 50:313–320.

Holder HD (1998) Cost benefits of substance abuse treatment: an overview of results from alcohol and drug abuse. J Ment Health Policy Econ 1:23–29.

Hunink M, Glasziou P, Siegel J (2001) Decision Making in Health and Medicine. Cambridge University Press, Cambridge UK.

Kerr WC, Fillmore KM, Bostrom A (2002) Stability of alcohol consumption over time: evidence from three longitudinal surveys from the United States. J Stud Alcohol 63:325–333.

Koch H, Meerkerk GJ, Zaat JO, Ham MF, Scholten RJ, Assendelft WJ (2004) Accuracy of carbohydrate-deficient transferrin in the detection of excessive alcohol consumption: a systematic review. Alcohol Alcohol 39:75–85.

Kraemer KL, Roberts MS, Horton NJ, Palfai T, Samet JH, Freedner N, Tibbetts N, Saitz R (2005) Health utility ratings for a spectrum of alcohol-related health states. Med Care 43:541–550.

Manwell LB, Fleming MF, Johnson K, Barry KL (1998) Tobacco, alcohol, and drug use in a primary care sample: 90-day prevalence and associated factors. J Addict Dis 17:67–81.

Mertens JR, Weisner C, Ray GT, Fireman B, Walsh K (2005) Hazardous drinkers and drug users in HMO primary care: prevalence, medical conditions, and costs. Alcohol Clin Exp Res 29:989–998.

Miller PM, Anton RF (2004) Biochemical alcohol screening in primary health care. Addict Behav 29:1427–1437.

Miller PM, Spies C, Neumann T, Javors MA, Hoyumpa AM, Roache J, Webb A, Kashi M, Sharkey FE, Anton RF, Egan BM, Basile J, Nguyen S, Fleming MF, Dillie KS (2006) Alcohol biomarker screening in medical and surgical settings. Alcohol Clin Exp Res 30:185–193.

National Institute of Alcohol and Alcoholism (2005) Assessing Alcohol Problems: A Guide for Clinicians and Researchers. Available at: http://pubs.niaaa.nih.gov/publications/Practitioner/PocketGuide/pocket_guide5.htm. Accessed April 6, 2007.

Ockene JK, Adams A, Hurley TG, Wheeler EV, Hebert JR (1999) Brief physician- and nurse practitioner-delivered counseling for high-risk drinkers: does it work? Arch Intern Med 159:2198–2205.

Paltiel AD, Walensky RP, Schackman BR, Seage GR III, Mercincavage LM, Weinstein MC, Freedberg KA (2006) Expanded HIV screening in the United States: effect on clinical outcomes, HIV transmission, and costs. Ann Intern Med 145:797–806.

Pignone M, Saha S, Hoerger T, Mandelblatt J, Pignone M, Saha S, Hoerger T, Mandelblatt J (2002) Cost-effectiveness analyses of colorectal cancer screening: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 137:96–104.

Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC (1996) The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA 276:1172–1177.

Saitz R, Horton NJ, Sullivan LM, Moskowitz MA, Samet JH (2003) 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 138:372–382.

Schuckit MA, Smith TL, Danko GP, Bucholz KK, Reich T, Bierut L (2001) Five-year clinical course associated with DSM-IV alcohol abuse or dependence in a large group of men and women. Am J Psychiatry 158:1084–1090.

Siegel JE, Weinstein MC, Russell LB, Gold MR (1996) Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. JAMA 276:1339–1341.

Taj N, Devera-Sales A, Vinson DC (1998) Screening for problem drinking: does a single question work? J Fam Pract 46:328–335.

U.S. Preventive Services Task Force (2004) Screening for Alcohol Misuse. Available at: http://www.ahcpr.gov/clinic/uspstf/uspsdrin.htm. Accessed April 6, 2007.

Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB (1996) Recommendations of the panel on cost-effectiveness in health and medicine. JAMA 276:1253–1258.

Journal of Prevention & Intervention in the Community, 37:129–142, 2009

ISSN: 1085-2352 print/1540-7330 online
DOI: 10.1080/10852350902735734

Routledge
Taylor & Francis Group

Post-Hospital Medical Respite Care and Hospital Readmission of Homeless Persons

STEFAN G. KERTESZ

Birmingham Veterans Affairs Medical Center Research Enhancement Award Program, and Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA

MICHAEL A. POSNER

Department of Mathematical Sciences, Villanova University, Villanova, Pennsylvania, USA

JAMES J. O'CONNELL

Harvard Medical School and Boston University School of Medicine, Boston, Massachusetts, USA

STACY SWAIN

Boston Health Care for the Homeless Program, Boston, Massachusetts, USA

ASHLEY N. MULLINS

The Wright Institute, Berkely, California, USA

MICHAEL SHWARTZ

School of Management, Boston University, Boston, Massachusetts, USA

ARLENE S. ASH

Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

This report is dedicated to the memories of Nurse Barbara McInnis of Pine Street Inn and Ms. Ellen Dailey, whose longstanding advocacy on behalf of Boston's homeless spurred the development of novel models of service to this population. The authors acknowledge the kind assistance of Mr. Walter Ferranti of Boston Medical Center, and Teresa Everson in the conduct of this study. The Lister Hill Center on Health Policy (University of Alabama at Birmingham), the National Institute on Drug Abuse (K23-DA-15487), and Boston Health Care for the Homeless Program supported this work. These entities had no role in the collection, analysis, or interpretation of the data or in the decision to submit the manuscript for publication. Professor Ash is associated with the company that licenses the software to implement Diagnostic Cost Group (DCG)–based casemix adjustment (Versick Health, lnc.).

Address correspondence to Stefan G. Kertesz, University of Alabama at Birmingham, 1530 3rd Avenue South MT608, Birmingham, AL 35294-0001, USA. E-mail: skertesz@uab.edu

Medical respite programs offer medical, nursing, and other care as well as accommodation for homeless persons discharged from acute hospital stays. They represent a community-based adaptation of urban health systems to the specific needs of homeless persons. This article examines whether post-hospital discharge to a homeless medical respite program was associated with a reduced chance of 90-day readmission compared to other disposition options. Adjusting for imbalances in patient characteristics using propensity scores, respite patients were the only group that was significantly less likely to be readmitted within 90 days compared to those released to Own Care. Respite programs merit attention as a potentially efficacious service for homeless persons leaving the hospital.

KEYWORDS discharge planning, health services, homeless, readmission, retrospective studies

The homeless, estimated to number 744,000 at a single point in time (National Alliance to End Homelessness, 2007), are subject to poor health status and excess mortality (Levy & O'Connell, 2004), and are also more likely to report being unable to obtain needed health care (Kushel, Vittingoff, & Haas, 2001). Homeless individuals experience high rates of hospitalization and prolonged length of stay relative to housed persons, and face distinct challenges for complete medical recovery after an acute medical hospitalization (Levy et al., 2004). Neither shelters, which often require vacating the premises during daylight hours, nor the streets support adherence to post-hospital medical recommendations (e.g., elevating an infected leg, administering insulin, adhering to a diet, or seeing a doctor). Lack of appropriate post-hospital disposition options for homeless inpatients may lead to unexpected hospital readmissions, especially for homeless persons with no safe place to heal. The videotaped incident of a 63-year-old homeless woman transported via taxi from a suburban hospital to Los Angeles's skid row, and released to the street in gown and slippers, pricks the conscience, but national publicity regarding this common community challenge is rare (Winton & DiMassa, 2006).

In response to this challenge, 48 communities across the United States and Canada have adopted homeless medical respite programs; Boston, in particular, has offered respite care and 24-hour accommodations for homeless persons for nearly two decades. Nationally, respite services vary according to local needs and funding, but typically include a bed, meals, transportation to appointments, and care by a wide range of clinicians familiar with caring for homeless persons (Buchanan, Doblin, Sai, & Garcia, 2006). Respite programs exemplify a more general principle of customizing clinical practice and systems of care to respond to the unique needs and life circumstances of persons experiencing homelessness.

Although respite programs could be justified on the basis of pragmatic necessity alone, their continued operation and financial support remains tenuous, as medical respite care for homeless individuals, unlike hospice care, is not a recognized or reimbursed category of service among major health payers such as Medicare and Medicaid. Hospitals and health plan administrators considering proposals for respite programs would like to see evidence of efficacy in either reducing costs, improving health outcomes, or at the very least, reducing the demand for scarce acute hospital beds. To date, both financial and logistic barriers have precluded the use of a randomized controlled trial to study medical respite care, and observational data currently provide the only evidence relating to these questions.

Observational data from a Chicago respite program suggested that discharge from a county hospital to a local respite program was associated with significantly fewer days of hospital care during the subsequent 12 months, compared to persons referred to respite but not accepted due to lack of space (Buchanan et al., 2006). Chicago's respite accepted only patients of low medical acuity, and thus excluded individuals requiring 24-hour nursing supervision or onsite physician services. By contrast, Boston's respite program provides 24-hour nursing supervision, daily visits by nurse practitioners or physician assistants, onsite physician supervision, in-house dental and psychiatric care, and case management. Equipped for patients in more substantial need, the program has helped free up acute inpatient services in local hospitals since 1985.

Because of high local need, Boston's respite unit has typically run above 90% capacity. As a result, some hospitalized homeless patients have been discharged from the safety net hospital back to their customary living environments (e.g., streets and shelters), and to other care settings such as private nursing homes and a publicly funded recuperative hospital. The latter options are similar to approaches used in communities where no homeless-customized respite program exists, and therefore provide a natural comparison that could complement the study reported by Buchanan et al. (2006). In this report of data collected over 3 years in Boston, we compared 90-day hospital readmission among patients discharged to respite versus other settings, adjusting for differences in patient characteristics, including burden of illness.

METHOD

Participants

We used administrative data to retrospectively identify a cohort of homeless persons, 18 or older, surviving at least one non-maternity, medical or surgical hospital admission to Boston Medical Center during July 1, 1998–June 30, 2001. Each subject's first eligible admission in this period ("index admission")

was analyzed, thereby permitting statistical methods appropriate for independent and uncorrelated data.

Because housing status is not regularly documented in medical records, we identified individuals as homeless if they had at least one outpatient encounter at the Boston Health Care for the Homeless Program (BHCHP) within ±365 days of the index admission to Boston Medical Center (BMC), an approach used by others (Martell et al., 1992). BHCHP serves over 7,000 homeless individuals annually, identified from over 60 outreach sites. A convenience sample of 10 individuals' clinical records found 9 had explicit mentions of "homeless" or "living on the streets" in the hospital discharge summary; some participants could have been homeless before or after the hospitalization but not on the day of hospitalization.

Boston's respite program ("Respite" hereafter) included 90 beds for men and women at the time of these data, receiving 1,600 admissions yearly, with 30–35% from inpatient medical hospitals, the remainder from emergency departments, shelters, the streets, and outpatient clinics. Its services include daily medical care, 24-hour nursing, a psychiatrist, case management, in-house dental care, and medication administration. It serves a severely distressed population: among 306 randomly reviewed records of men admitted to respite, 90% had active substance abuse disorders (not including current tobacco or past drug/alcohol) and 53% had a DSM-IV non-addiction psychiatric diagnosis. Among 104 women, 75% had active substance abuse disorders and 85% had non-addiction psychiatric diagnoses. Seventy-seven percent of male and female admissions were homeless more than one year. Persons with four or more major medical illnesses, active substance abuse, *and* a non-substance abuse psychiatric diagnosis accounted for 40% of male and 55% of female respite admissions, respectively (O'Connell & Swain, 2001).

Procedure

In general, a decision to discharge a hospitalized patient to Respite involved the combined inputs of caregivers (residents, attending physicians, nurses, case managers, Boston Health Care for the Homeless visiting staff, shelter personnel), the patient, and potential receiving facilities (the Respite, shelters, and other potential settings, such as nursing homes). Typically hospital staff propose Respite for patients requiring additional service (e.g., dressing changes), observation, or a safe nonhomeless environment as a prerequisite to medical recovery outside of the hospital. Importantly, payment had little influence on disposition given the high rate of insurance among patients seen by BHCHP (85%, mostly Medicaid) and the availability of multiple public and private funding mechanisms for both Respite and Other Planned Care, including a state-funded secondary care hospital as well as an uncompensated care pool (Bovjberg & Ullman, 2002).

Preliminary extracts from BMC's Medical Information System identified 858 persons who had a BHCHP outpatient visit within ±365 days of an index hospitalization. We re-queried BMC's Medical Information System for all hospital and hospital-based ambulatory encounters from January 1, 1998 (6 months prior to July 1, 1998) to June 1, 2002 (11 months after June 30, 2001). This permitted us to:

Apply exclusions: Of 858 patients, 14 were hospitalized for childbirth (mother and infant care is not available through Respite), 35 did not survive to hospital discharge, 41 had unplanned medical discharges against advice, and 3 records could not be found (likely due to interval changes in identifiers). We also excluded 22 who had been readmitted within 24 hours of discharge, because it is fairly common for discharges to other facilities to be redirected to Respite during the first post-dicharge day.

Identify endpoints: We then identified BMC readmissions within 90 days of hospital discharge. Death within that 90-day period was compiled from BHCHP's Homeless Death Database and the Massachusetts' Registry of Vital Records and Statistics (1998–2001). The 8 persons who died (2 Respite, 3 Own Care, and 3 Other Care) were not included in the analysis of readmission outcome, leaving 735 (134, 171, and 430, respectively).

Obtain diagnostic information: We captured diagnoses from all BMC encounters for the index admission and the 6 prior months, including inpatient care, BHCHP's own primary care clinic at BMC, emergency, and outpatient specialty services.

Measures

One of three discharge dispositions was identified for each participant:

Respite. This category included persons referred to Respite up to one day *after* hospital discharge. Delayed referrals occurred when street/shelter clinicians encountered a newly discharged patient who appeared to require a place (i.e., Respite) in which to recover. Including such individuals in the Respite group reduced misclassification of disposition status.

Own Care. Homeless patients described in administrative hospital records as discharged "home" (the administrative system did not include a field for discharge to streets and shelters).

Other Planned Care. Non-Respite patients discharged to supervised recuperative care (e.g., skilled nursing facilities, chronic care hospitals, or home health care).

The primary study endpoint was inpatient hospital readmission ≤ 90 days from discharge, a timeframe appropriate for judging the adequacy of discharge planning. A key interest was to compare readmission for Respite versus discharge to streets or shelters (Own Care), the default in most

communities. However, other post-discharge settings including nursing homes (Other Planned Care) were considered, because these are often relied on in the absence of a Respite program (Gundlapalli et al., 2005).

Financial costs were estimated for all patients based on charges at the referring hospital (Boston Medical Center). For Respite patients, we estimated costs related to Respite care through reference to (a) average reimbursement to the Respite (per patient day) during the period studied and (b) the duration of each Respite stay. All figures were inflation-adjusted to 2002 dollars, and do not include costs for Other Planned Care facilities, or system-wide costs resulting from discharge of homeless, medically ill individuals to shelters or streets (e.g., ambulance, additional emergency room visits, additional shelter-based services, and jails), because the latter costs were not available. Although hospital charges tend to overstate hospital costs (thereby inflating the cost savings from reduced hospital days), this bias may be offset by the failure to count the money saved through likely reductions in these other publicly funded services.

Additional covariates, drawn from the hospital readmission literature (Corrigan & Martin, 1992) included: age, sex, race/ethnicity, length of the index hospital admission, presence in the record of drug and alcohol abuse diagnostic codes during the admission or the preceding 6 months, and medical illness burden. The latter was estimated using the Diagnostic Cost Groups (DCG) risk score (Ash et al., 2000), calculated from all medical and psychiatric diagnoses coded during the index admission and during the prior 6 months of inpatient and outpatient care at Boston Medical Center, including onsite primary care and mental health services from BHCHP. The DCG method, often used by health plans to predict high-cost patients, generates a numerical estimate for expected health service utilization, and has been shown to predict mortality, utilization, and health costs (Petersen, Pietz, Woodard, & Byrne, 2005). DxCG 6.1 for Windows software was used, applying a DCG model calibrated to Massachusetts Medicaid experience for 2000–2001.

Data Analysis

The primary unadjusted analysis compared 90-day readmissions among persons discharged to Respite, Own Care (i.e., streets and shelters), and Other Planned Care.

In the absence of a prospective randomized controlled trial, the adjusted analysis relied on a statistical technique (propensity scores) to match groups in regard to their likelihood of being discharged to Respite. Propensity adjustment reduces the bias affecting retrospective observational comparisons (Braitman & Rosenbaum, 2002; Rosenbaum & Rubin, 1983), and is simplest to apply to 2-group comparisons. Therefore, each subject's propensity to be discharged to Respite versus Own Care was calculated with multivariable logistic regression, using the covariates listed earlier. For the 90-day readmission outcome,

observations were weighted according to the propensity score so that the two groups being compared had the same overall propensity to be assigned to either discharge disposition. Specific weights were computed as: 1/(propensity to be discharged to Respite) for each Respite observation, and 1/(1-propensity to be discharged to Respite) for each Own Care observation (Hirano & Imbens, 2001). This method is similar to propensity score approaches that match individuals having similar propensities (but who received different treatments). Instead of dropping unmatched participants, however, it retains all subject data.

With propensity-weighted data, we computed the association between discharge disposition (Respite versus Own Care) and 90-day readmission using a logistic regression model that included the covariates of age, race, sex, index hospital length of stay, DCG score, alcohol abuse, and drug abuse. Secondarily, both the propensity score and logistic regression analysis were repeated to compare readmissions for Other Planned Care versus Respite.

We compared the 90-day total costs (combining inpatient hospital readmissions and, where applicable, Respite charges) for patients discharged to Respite versus Own Care, in both unadjusted (*t*-test) and adjusted analyses, the latter incorporating propensity-weighted data in a multiple linear regression adjusted for the same measured potential confounds. A comparison of costs for Respite versus Other Planned Care was not undertaken because Other Planned Care costs could not be obtained.

Because over 85% of patients had insurance, and Massachusetts provided back-up funding options for persons without insurance, we did not include this variable in the statistical model. All analyses were carried out with SAS System for Windows (Version 8.2).

RESULTS

Of the 743 individuals discharged from the hospital, 136 (17%) were discharged to Respite, 174 (22%) to Other Planned Care, and 433 (55%) to Own Care. Compared to Own Care, Respite patients were older, more likely to be White, less likely to be female, and somewhat more likely to have record of Alcohol Abuse, but less likely to have record of Drug Abuse (Table 1). The index hospital stay was roughly 3 days longer among those discharged to Respite and to Other Planned Care settings, and extremely short hospital stays (0–2 days) were less common among Respite compared to other patients (see Table 1). At 90 days, 8 patients had died (2 discharged to Respite, 3 to Own Care and 3 to Other Care), leaving 735 for readmission analysis.

Early Readmission

Readmission by 90 days occurred among 156 patients (21.2% of the sample). There was no difference in the proportion readmitted in comparisons not adjusted for patient characteristics (Table 1).

TABLE 1 Characteristics of 743 Homeless Individuals Discharged from Boston Medical Center (July 1, 1998–June 30, 2001) by Discharge Disposition[a]

n	ALL 743	Respite 136	Other Planned Care 174	Own Care 433	p^b
Age in years					
<40	26%	19%	21%	31%	*0.01*
40–55	52%	55%	52%	51%	
>55	27%	26%	27%	19%	
Mean Age (*SD*)	46.9 (11.0)	48.5 (11.8)	48.6 (10.6)	45.7 (10.8)	*0.003*
Sex					
Female	20%	13%	14%	24%	*<0.01*
Race/Ethnicity					
White	44%	56%	48%	39%	*0.02*
Black	41%	35%	40%	44%	
Hispanic	13%	8%	11%	16%	
Other	1%	1%	1%	1%	
Index Hospital LOS					*<0.001*
0–2 days	35%	18%	29%	43%	
3–5 days	41%	40%	34%	43%	
6+ days	24%	41%	37%	14%	
Mean LOS (SD)	4.6 (5.3)	6.4 (5.9)	6.1 (7.2)	3.5 (2.7)	*<0.001*
Illness burden (DCG)[c]					*0.002*
Low	13%	10%	5%	17%	
Medium	67%	69%	72%	65%	
High	20%	21%	23%	18%	
Mean Illness Burden score (SD)[c]	1.1 (1.0)	1.1 (0.9)	1.2 (1.0)	1.1 (1.0)	0.34
Alcohol abuse[d]	32%	34%	34%	31%	0.66
Drug abuse[d]	17%	8%	16%	19%	*0.009*
Readmitted within 90 days of discharge[e]	21%	21%	22%	21%	0.94

Note: Italicized comparisons are significant at $p < .05$.

[a]Percentages do not consistently add to 100% due to rounding.

[b]*p*-values reflect a 3-group comparison (Respite versus Own Care versus Other Planned Care) by Chi-squared test or analysis of variance ($df = 2$), with $\alpha = 0.05$, 2-tailed.

[c]Illness burden computed with the Diagnostic Cost Group (DCG) prospective relative risk score based on diagnoses recorded during 180 days previous to, and during, the index admission. Low, medium, and high risk indicate DCG relative risk scores of <0.5, 0.5–1.5, and >1.5, respectively.

[d]Alcohol and drug abuse are based on administratively coded (ICD-9) diagnoses from the index hospitalization and the prior 6 months of care at that hospital (Boston Medical Center).

[e]Computation of percentage readmitted excludes 8 of 743 patients who died during the 90-day follow-up interval (2 Respite, 3 Own Care, and 3 Other Care).

As expected, some potentially adverse characteristics were associated with readmission. For example, readmission was more common among persons discharged following index hospital stays lasting six or more days (31%), compared to shorter stays (21% for 3–5 days, and 15% for 0–2 days) and among those with higher versus mid-range or lower illness burden as measured by DCG score. Both characteristics were more common among

Respite and Other Planned Care patients, compared to patients discharged to Own Care (Table 1).

Propensity models were moderately robust in their capacity to predict each patient's likelihood to be discharged to Respite, compared to Own Care ($c = .76$, range 0–1, with 1 indicating perfect fit between the modeled propensity and the actual treatment assigned). To illustrate, when individuals were divided by quintiles based on propensity to be discharged to Respite, patients in the highest quintile had about 8 times greater likelihood of discharge to Respite (56 of 109 persons, 51%), compared to patients in the lowest quintile (7 of 109, 6.4%).

Table 2 shows that prior to balancing for propensity to be discharged to Respite, the Respite and Own Care groups differed substantially on several characteristics. Both illness burden and index hospitalization length of stay,

TABLE 2 Characteristics of Homeless Individuals Discharged to Respite ($n = 134$) Versus Own Care ($n = 430$) After Inpatient Hospitalization, Before and After Weighting by Propensity Scores

	Raw comparison (before propensity-score weighting)			Propensity-score weighted comparison[a]		
	Respite (%)	Own Care (%)	*p*	Respite (%)	Own Care	*p*
Age in years			*0.01*			*0.04*
<40	19	31		21	27	
40–55	55	51		55	52	
>55	26	18		24	21	
Sex			*0.005*			0.06
Male	87	76		83	79	
Female	13	24		17	21	
Race/Ethnicity						
Black	34	44	0.06	42	42	0.99
Hispanic	8	16	*0.03*	14	14	0.85
Other	1	1	0.94	1	1	0.40
White	55	38	*<0.001*	44	43	0.96
Index Hospital LOS			*<0.001*			0.89
0–2 days	17	43		36	37	
3–5 days	40	43		42	42	
6+ days	42	14		22	21	
Illness Burden (DCG)			0.09			0.18
Low	10	17		12	15	
Medium	69	65		70	66	
High	21	18		18	19	
Drug abuse	8	19	*0.002*	14	16	0.33
Alcohol abuse	33	30	0.60	34	31	0.34

Note: Propensity scores were developed by applying all displayed variables in a single logistic regression model predicting discharge location. Propensity score-weighted groups combine data available for all Respite and Own Care subjects, applying a weight of 1/(propensity score) for each Respite observation and 1/(1-propensity score) for each Own Care observation (Hirano et al., 2001). Italicized comparisons are significant at the $p < .05$ level, 2-tailed, applying Chi-squared and *t*-tests, as appropriate (all $df = 1$).

characteristics that predicted readmission, were greater for Respite patients. The right side of Table 2 also shows that these characteristics were more closely matched after reweighting the data with propensity scores.

In the final adjusted model comparing Respite to Own Care (Table 3), Respite patients had significantly reduced odds of hospital readmission by 90 days in comparison to Own Care patients. The estimate for Other Planned Care, compared to Own Care, also suggested reduced odds for readmission ($OR = 0.70$; 95% CI 0.46–1.06), but the association was not significant at the .05 level (full model not shown, but available from the authors).

Total Charges

The mean charges for a Respite stay were \$7,929 ($SD = \$8{,}649$) with mean length of stay 31.3 days ($SD = 32.6$, median = 20). The mean 90-day charges for individuals discharged to Respite, summing Respite and (where

TABLE 3 Predictors of Hospital Readmission Within 90 Days of Discharge Among Homeless Persons in Boston Discharged to Medical Respite Versus Discharge to Their Own Care (1998–2001)

	Respite versus own care
	Odds ratio (95% confidence interval)
Discharge Disposition	
Respite	*0.54 (0.34–0.85)*
Own Care	1.0 (Ref)
Age (years)	
<40	0.81 (0.43–1.52)
40–55	1.0 (Ref)
>55	0.85 (0.48–1.50)
Sex	
Female	1.03 (0.54–1.95)
Race/Ethnicity	
Black	*0.58 (0.36–0.94)*
Hispanic	0.46 (0.21–1.00)
White/Other	1.0 (Ref)
Index Hospital LOS	
0–2 days	*0.49 (0.28-0.85)*
3–5 days	1.0 (Ref)
6+ days	1.35 (0.79–2.30)
Illness Burden	
Low	0.44 (0.16–1.21)
Medium	1.0 (Ref)
High	*1.90 (1.10–3.28)*
Alcohol abuse	1.11 (0.68–1.82)
Drug abuse	0.90 (0.47–1.72)

Note: Results for a single multivariable logistic regression are shown, adjusted for all variables displayed, using propensity score-weighted data to minimize heterogeneity between the Respite versus Own Care disposition groups; italicized comparisons are significant at $p < .05$.

applicable) readmission charges was \$10,359 ($SD = \$10{,}523$). The 90-day total exceeded the mean readmission charges of \$2,819 ($SD = \$8{,}064$) among patients discharged to Own Care, $t(187) = 7.68$, $p < .001$. This comparison does not take into account the adverse characteristics associated with being a Respite candidate, or savings from reduced hospital readmissions at 90 days. In adjusted analysis, a Respite disposition was associated with +\$5994 (95% CI, \$4,210–\$7,779) in excess charges, relative to Own Care. The potentially higher Costs of Other Planned Care were not available to this study, and are discussed later.

DISCUSSION

In this sample, patients discharged to Boston's medical respite program had some characteristics associated with elevated risk for hospital readmission within 90 days, but in unadjusted analyses they were not readmitted more often than patients discharged to the streets and shelters, or to care facilities. In analyses controlling for individual characteristics, discharge to a homeless respite program was associated with an approximately 50% reduction in the odds of readmission at 90 days post-discharge, compared to discharge to streets and shelters (Own Care), similar to what was found in Chicago by Buchanan et al. (2006). Other Planned Care settings, such as nursing homes, did not achieve a similarly robust reduction in the likelihood of readmission when compared to those released to Own Care.

The Respite-associated reduction in readmission may reflect the program's customization for the complex problems of medically ill homeless individuals. Services included 24-hour nursing, as well as onsite physicians (including psychiatrists), nurse practitioners, physician assistants, caseworkers, and a dental team, all experienced in homeless health care. Recuperative care was accompanied by interventions for other illnesses, arrangements for (and transportation to) continuing outpatient care, establishment of a new primary care relationship, spiritual care, 12-step meetings, and identification of social and financial resources. Although some of these services may exist in other settings, few combine all these services for homeless individuals.

The present report should be compared to one prior study of respite, comparing post-discharge hospital utilization among 161 homeless patients discharged to a Chicago respite versus 64 patients referred to respite but not accepted due to lack of space (Buchanan et al., 2006). The authors reported a 49% reduction in hospital admissions in adjusted analyses. Our findings are not discordant, but reflect a program designed for patients with higher medical acuity, suggesting that a homeless respite program may sometimes take the place of skilled nursing facilities.

The analysis of measured costs, including hospital and respite care, suggest that a policy of discharging homeless patients to a respite program is

potentially more expensive than a policy of discharging them to the streets and shelters. This inference is tempered, however, by lack of data concerning the full range of costs associated with discharging people into homelessness. Where those costs have been measured, notably among chronically homeless persons in New York City, the combined judicial, medical, and mental system costs associated with homelessness exceeded $40,000 per year (Culhane, Metraux, & Hadley, 2002).

For policy makers the most relevant cost comparison may be the one this study could not formally accomplish, namely, between the Respite and Other Planned Care. A speculative estimate combining typical rehabilitative skilled nursing facility, professional fees, and the mean duration of post-hospital nursing home stays suggests that discharge to a non-respite nursing facility with professional services is likely to involve costs in the range of $4,512–$7,520 (Gundlapalli et al., 2005; Medicare Payment Advisory Commission, 2006). The mean cost of a discharge to Boston's Respite ($M = \$7{,}929$, falling to $5,994 after adjustment for hospital readmission savings) may be justifiable because: (a) Respite was associated with reduced 90-day readmission, while Other Planned Care settings were not, and (b) Respite offered a homeless-customized service model, as reviewed earlier.

The principal limitation to this study is reliance on observational data. Given the nearly universal prevalence of medical, mental, and substance abuse problems among the Respite patients, it is unlikely that selection of a particularly healthy subgroup of homeless individuals biased the results. Additionally, the analyses adjusted for measured confounds, some of which suggested that patients discharged to Respite were at higher readmission risk.

This study's strengths include the use of multiple data sources to identify a large cohort of hospitalized homeless patients, producing one of the largest comparative studies of a medical service for homeless persons to date. Comprehensive casemix adjustment and propensity scoring are important methodologic tools not previously applied to comparing interventions for the homeless. Given high hospital utilization by a growing homeless population, this study offers a methodological advance, and may lay the groundwork for a much-needed randomized trial of respite care in comparison to other care arrangements.

It should be emphasized that the design of this particular study was driven by our interest in an easily measured outcome, hospital readmission. However, Boston's respite program, like others, receives patients directly from emergency rooms, shelters, detoxification facilities, and the streets and may play a hospital diversion role unmeasured in the present study.

In March of 2004, a coalition of homeless persons in Birmingham, Alabama, pleaded "we need a surgical and hospital discharge shelter for the large number of us who are discharged from the hospital with no place to recuperate" (Letter of March 24, 2004 to City Council of Birmingham,

Alabama). This study suggests that offering a safe "place to recuperate" could meet patients' needs while reducing hospital readmissions. The findings should spur further research, and lend impetus to recognition of this service.

REFERENCES

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*(3), 7–25.

Bovjberg, R. R., & Ullman, F. C. (2002). *Recent Changes in Health Policy for Low-Income People in Massachusetts*. Washington, DC: The Urban Institute. Retrieved April 28, 2007, from http://www.urban.org/UploadedPDF/310431.pdf

Braitman, L. E., & Rosenbaum, P. R. (2002). Rare outcomes, common treatments: Analytic strategies using propensity scores. *Annals of Internal Medicine*, *137*(8), 693–695.

Buchanan, D., Doblin, B., Sai, T., & Garcia, P. (2006). The effects of respite care for homeless patients: A cohort study. *American Journal of Public Health*, *96*(7), 1278–1281.

Corrigan, J., & Martin, J. (1992). Identification of factors associated with hospital readmission and development of a predictive model. *Health Services Research*, *27*(1), 81–101.

Culhane, D., Metraux, S., & Hadley, T. (2002). Public service reductions associated with placement of homeless persons with severe mental illness in supportive housing. *Housing Policy Debate*, *13*(1), 107–162.

Gundlapalli, A., Hanks, M., Stevens, S. M., Geroso, A. M., Viavant, C. R., & McCall, Y., et al. (2005). It takes a village: A multidisciplinary model for the acute illness aftercare of individuals experiencing homelessness. *Journal of Health Care for the Poor & Underserved*, *16*(2), 257–272.

Hirano, K., & Imbens, G. (2001). Estimate of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services & Outcomes Research Methodology*, *2*, 259–278.

Kushel, M. B., Vittingoff, E., & Haas, J. S. (2001). Factors associated with the health care utilization of homeless persons. *Journal of the American Medical Association*, *285*(2), 200–206.

Levy, B. D., & O'Connell, J. J. (2004). Health care for homeless persons. *New England Journal of Medicine*, *350*(23), 2329–2332.

Martell, J. V., Seitz, R. S., Harada, J. K., Kobayashi, J., Sasaki, V. K., & Wong, C. (1992). Hospitalization in an urban homeless population: The Honolulu Urban Homeless Project. *Annals of Internal Medicine*, *116*(4), 299–303.

Medicare Payment Advisory Commission. (2006). *Report to the Congress: Medicare Payment Policy*. Washington, DC: MedPac. Retrieved November 10, 2006, from http://www.medpac.gov/publications/congressional_reports/Mar06_EntireReport.pdf.

National Alliance to End Homelessness. (2007). *Homelessness counts*. Washington, DC. Retrieved April 29, 2007, from http://www.naeh.org/content/article/detail/1441.

O'Connell, J. J., & Swain, S. (2001). *Medical respite care for homeless persons in Boston*. Medicaid review (document and slides): Boston Health Care for the Homeless Program.

Petersen, L. A., Pietz, K., Woodard, L., & Byrne, M. (2005). Comparison of the predictive validity of diagnosis-based risk adjusters for clinical outcomes. *Medical Care*, *43*(1), 61–67.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrica*, *70*, 41–55.

Winton, R., & DiMassa, C. M. (2006, November 16). L.A. files patient 'dumping' charges. *Los Angeles Times*, p. A1.

Low Vitamin D Status of Patients in Methadone Maintenance Treatment

Theresa W. Kim, MD, Daniel P. Alford, MD, MPH, Michael F. Holick, PhD, MD, Alan O. Malabanan, MD, and Jeffrey H. Samet, MD, MA, MPH

Aim: To examine the prevalence and risk factors of low vitamin D status (vitamin D deficiency or insufficiency) among patients in a methadone maintenance treatment (MMT) program.

Design: Cross-sectional study of subjects recruited from an MMT program in a higher latitude (Boston, MA).

Measurements: Standardized survey and medical record review were used to assess patient characteristics. Serum was tested to determine vitamin D deficiency (25-hydroxyvitamin D <20 ng/mL) and insufficiency (25-hydroxyvitamin D between 20 and 30 ng/mL). Multivariable analyses were used to assess risk factors associated with vitamin D deficiency.

Findings: Low vitamin D status was found in 52% of the subjects (48 of 93), deficiency in 36%, and insufficiency in an additional 16%. Older age (OR = 3.47; 95% CI 1.31–9.22) and black or Hispanic race/ethnicity (OR 3.34; 95% CI 1.30–8.58) were significantly associated with higher risk of vitamin D deficiency.

Conclusion: Low vitamin D status was present in a majority of patients recruited from an MMT program. This raises the question as to whether this is a generalizable phenomenon and whether these patients are at higher risk of complications of low vitamin D status including bone pain, periodontal disease, osteomalacia, and cardiovascular disease.

Key Words: methadone maintenance, vitamin D, drug dependence, medical complications

(*J Addict Med* 2009;3: 134–138)

The Institute of Medicine[1] recently stated that co-occurring medical conditions in drug-dependent populations are a reality that merit attention and action. Effective medical care is often underutilized among individuals with substance-use disorders, due in part to unawareness of treatable medical conditions. Attending to physical health problems may be of particular importance to individuals enrolled in methadone maintenance treatment (MMT); opioid-dependent individuals in the United Sates tend to be older than opioid-dependent individuals in other forms of addiction treatment.[2]

Low vitamin D status, which we use to refer to either vitamin D deficiency or insufficiency, has been recognized in a variety of populations with medical conditions; yet, it has received little attention in drug-dependent populations.[3–5] Examining vitamin D in individuals receiving MMT may be particularly important, because low vitamin D status can result in nonspecific musculoskeletal pain[6,7] as well as periodontal disease and tooth loss,[8] conditions common in MMT patients.[9,10] Low vitamin D status can also lead to a higher risk of fracture by exacerbating osteoporosis,[11] a painless bone disease of low bone mass. Low bone density has also been noted in opioid-dependent populations.[12,13]

Low vitamin D status is more common than previously thought, ranging from 21% in an elderly low-income African-American population in Boston[3] to 71% in patients with severe peripheral arterial disease.[14] The most important source of vitamin D is sunlight, the skin synthesis the vitamin from sunlight. Factors that affect vitamin D synthesis are the level of sunlight exposure (eg, season, latitude, and time of day), diet, skin pigmentation, sunscreen use, and age.

Better understanding of the contribution of vitamin D status to the bone health of MMT patients could lead to pragmatic interventions to identify patients in addiction treatment who are at risk for vitamin D deficiency and ameliorate the associated medical comorbidities. Thus, we sought to examine the prevalence and risk factors of low vitamin D status in a population enrolled in MMT.

METHODS

Study Design and Sampling

This was a cross-sectional study of participants recruited from the 350 patients of the Boston Public Health Commission's MMT program. Flyers were posted at the MMT program's dosing and counseling sites inviting patients to participate in a study about bone health. Research associates were present at the program to schedule research appointments at the Boston University School of Medicine General Clinical Research Center. All patients who had received methadone for at least 30 days from the MMT

From the Clinical Addiction Research and Education Unit, Section of General Internal Medicine (TWK, DPA, JHS), Section of Endocrinology, Diabetes, and Nutrition (MFH), Vitamin D, Skin, and Bone Research Laboratory (MFH), Department of Medicine, Boston University School of Medicine, Boston, MA; Department of Social and Behavioral Sciences (JHS), Boston University School of Public Health, Boston, MA; and Division of Endocrinology (AOM), Diabetes and Metabolism, Beth Israel Deaconess Medical Center, Boston, MA.

Received for publication July 1, 2008; accepted December 31, 2008.

Send correspondence and reprint requests to Theresa W. Kim, MD, 801 Massachusetts Avenue, Boston University Medical Center, Boston, MA 02118. e-mail: theresa.kim@bmc.org

ISSN: 1921–0629/09/0303-0134

program were eligible to participate. The study was conducted between August and December of 2003. The Boston University Institutional Review Board approved the study protocol.

Data Collection

After providing written informed consent, study participants met with trained research associates for a standardized interview assessing the following: demographics, vitamin D supplementation ("Have you taken vitamin D in the past week?"), medical conditions associated with low vitamin D status including chronic liver disease, chronic pancreatitis, and renal disease. Height and weight were measured to calculate body mass index (weight [kg]/height^2 [m]), because obesity (body mass index ≥30) is associated with low vitamin D status in some populations.[4] Data on alcohol use,[15] opioid use,[12,13] and HIV infection[16] were collected, because they have been associated with low bone density. Medical record review was performed to ascertain the presence of medications that are associated with low vitamin D status including corticosteroids and antiepileptics (eg, phenytoin, carbamazepine, and phenobarbital).

We used the widely accepted definition of vitamin D deficiency or serum 25-hydroxyvitamin D of <20 ng/mL.[17] We also report the prevalence of vitamin D insufficiency or 25-hydroxyvitamin D between 20 and 30 ng/mL. This is supported by the observation that elevated parathyroid hormone levels do not plateau until 25-hydroxyvitamin D levels are above 30 ng/mL. One source of confusion stems from the fact that older articles used different cutoffs for the terms vitamin D deficiency and insufficiency. Despite this source of confusion, we included the prevalence of insufficiency, because although there is no consensus on the optimal range of 25-hydroxyvitamin D, there is more agreement that levels between 20 and 30 ng/mL are still suboptimal.[18,19]

Statistical Analysis

Analysis included descriptive statistics for all variables, such as using *t* tests for continuous variables and χ^2 analysis or Fisher exact test for categorical variables. We used logistic regression models to examine the relationship between independent variables and vitamin D deficiency. We used vitamin D deficiency as the outcomes in these analyses, because it is more widely known than vitamin D insufficiency. The limited number of vitamin D deficiency cases precluded models that adjusted for all independent variables simultaneously. Therefore, separate adjusted models were created for each independent variable; all adjusted analyses included age, gender, and race/ethnicity. All statistical analyses were performed using SAS version 8.2 (SAS System for Windows 2001; SAS Institute, NC).

RESULTS

Sample Characteristics

Of the 350 patients enrolled in the MMT program, 106 volunteered to be included in the study. Thirteen (12%) were excluded due to inadequate venous access, which precluded serum collection. Of the remaining 93 subjects (Table 1), the majority were women (63%), black or Hispanic (58%), and overweight or obese (62%). Most of the subjects entered MMT after a median of 11 years of heroin use. About half had been heavy drinkers for at least a year, but only 14% reported recent (past month) heavy alcohol use. Many subjects reported past month medical conditions associated with low vitamin D deficiency status including liver disease (54%), kidney disease (7%), and chronic pancreatitis (1%). In addition, 12% of the sample was taking medications associated with vitamin D deficiency. Despite multiple risk factors for vitamin D deficiency, only 2 participants were taking any form of vitamin D supplementation. No differences were found between the entire MMT program population and the study sample in terms of age ($P = 0.48$), gender ($P = 0.65$), or race/ethnicity ($P = 0.5$).

TABLE 1. Characteristics of Study Participants Recruited From a Methadone Maintenance Treatment Clinic (n = 93)

Variable	n (%)
Gender	
Female	59 (63)
Race/ethnicity	
Black	42 (45)
Hispanic	12 (13)
White	39 (42)
Body mass index (kg/m^2)	
<18.5	1 (1)
18.5–24.9	35 (38)
25–29.9	23 (25)
≥30	34 (37)
HIV infection	25 (27)
Heavy alcohol use, ≥1 yr	48 (53)
Heavy alcohol use, past month*	13 (14)
Heroin use, lifetime years, median (range)	11 (0–38)
Heroin use, past month	21 (23)
Methadone maintenance treatment, lifetime years, median (range)	3.2 (0.08–25)
Age, years, median (range)	42 (21–66)
Condition associated with low vitamin D status	
Chronic liver disease†	50 (54)
Chronic kidney disease	6 (7)
Chronic pancreatitis	1 (1)
Medications (including prednisone, carbamazepine, phenytoin, and phenobarbital)	11 (12)

*Defined as more than 3 drinks/occasion, more than 3 occasions/wk.

†"Has a doctor ever told you that you had ongoing liver disease (eg, cirrhosis or chronic hepatitis B or C)?"

The prevalence of vitamin D deficiency was 36% (95% CI 26%–45%). The prevalence of vitamin D insufficiency was found in an additional 16% (95% CI 9%–23%). Only 2 participants reported having been informed by a doctor of having vitamin D deficiency before this study. Participants were encouraged to give a letter with their 25-hydroxyvitamin D levels to his/her doctor.

Individuals older than 40 years were more than 3 times likely to have vitamin D deficiency (OR 3.47; 95% CI 1.31–9.22). In addition, those who were either black or Hispanic were more likely to be vitamin D deficient (OR

3.34; 95% CI 1.30–8.58). Recent heavy alcohol consumption, recent heroin use, and longer years in MMT were associated with a nonstatistically significantly higher OR of vitamin D deficiency (Table 2).

DISCUSSION

More than half of the opioids-dependent population recruited from an MMT program had vitamin D status that is considered insufficient for optimal bone health. Although medical comorbidities of substance-use disorders have received greater recognition recently,[20] previous studies have not addressed vitamin D deficiency or insufficiency as a medical comorbidity. The findings in this study suggest that low vitamin D status is common among individuals with opioid dependence and deserves greater attention as a treatable medical issue in this population.

TABLE 2. Predictors of Vitamin D Deficiency in Adjusted Analyses*

Variable	% Vitamin D Deficiency	OR (95% CI)
Age† (yr)		
≥40	46	3.47 (1.31–9.22)
<40	19	1
Gender		
Female	37	1.24 (0.51–3.03)
Male	32	1
Race/ethnicity		
Non-white	46	3.34 (1.30–8.58)
White	21	1
Body mass index (kg/m^2)		
≥30	44	1.42 (0.54–3.73)
<30	31	1
HIV infection		
Yes	32	0.69 (0.24–2.01)
No	37	1
Liver disease		
Yes	34	0.83 (0.33–2.13)
No	37	1
Season		
November–April	35	1.15 (0.37–3.58)
May–October	36	1
Heavy alcohol use, past month		
Yes	46	1.47 (0.39–5.50)
No	33	1
Heavy alcohol use, lifetime (yr)		
≥1 yr	38	1.08 (0.42–2.82)
<1 yr	33	1
Heroin use, past month		
Yes	43	2.02 (0.67–6.07)
No	33	1
Heroin use, lifetime (yr)‡		
≥11	40	1.01 (0.37–2.76)
<11	30	1
MMT, lifetime (yr)‡		
≥3.2	43	1.60 (0.61–4.17)
<3.2	28	1

*Results of separate logistic regression models for each independent variable, adjusting for age, gender, and race/ethnicity.

How does this estimate of low vitamin D status of patients on chronic methadone treatment compare with other populations in the literature? Methodological differences can often explain vitamin D differences between studies; the following similarities made it possible to compare healthy volunteers in a study by Tangpricha et al[5]: (1) study location in the Northeast, specifically, the city of Boston, (2) identical assay used to measure serum 25-hydroxyvitamin D levels,[21] and (3) primary outcome or 25-hydroxyvitamin D deficiency <20 mg/dL. Our study, which was conducted predominantly at the end of summer (after maximal sun exposure), found more vitamin D deficiency (36%) compared with the study of Tangpricha et al (11%). Other studies of more debilitated populations have reported higher prevalence of vitamin D deficiency including adults with HIV infection (79%)[22] elderly African-Americans in Boston (73%),[3] and uninsured women in Michigan (67%).[23] In summary, our estimate of vitamin D deficiency is higher than a similar study of healthy volunteers but lower than potentially more debilitated populations.

Very few individuals knew of their vitamin D status despite a high proportion of patients with medical conditions or medications associated with low vitamin D. Our results underscore the pressing need to treat co-occurring medical conditions as well as the addiction needs of patients in substance use treatment so as to improve overall health. Primary care physicians should be aware that low vitamin D status could contribute to the chronic pain of patients with opioid dependence.

The finding that black and Hispanic individuals are at higher risk for vitamin D deficiency is consistent with other studies.[3,11,23] One reason for this may be reduced skin synthesis of vitamin D in individuals with darker skin pigmentation. Among women of reproductive age in the third National Health and Nutrition Examination Survey, the prevalence of 25-hydroxyvitamin D ≤15 ng/mL was 42% in African-American women versus 4% among white women.[4] This biologic phenomenon could give rise to ethnicity-related health disparities. We also found that older patients were more likely to have vitamin D deficiency in our sample, which may be due to the skin's decreased ability to synthesize vitamin D with aging.[24] Older age may also be accompanied by lower physical functioning, which may result in less sun exposure.

Low vitamin D status in individuals in an MMT program is worthy of concern in light of the relatively high prevalence of chronic pain in drug-dependent populations.[25–27] Patients in MMT experiencing chronic limb or joint pain are likely to attribute this pain to the effects of methadone.[28] Vitamin D deficiency causes osteomalacia, a bone disease of ineffective bone matrix mineralization and a syndrome of diffuse skeletal pain and muscle aches.[29] In a study of patients with chronic pain, 93% of them were of vitamin D deficiency.[6] Long-term therapy with some medications such as corticosteroids, antiepileptic drugs, and some

HIV antiretroviral therapies can lower vitamin D status by induction of the steroid and xenobiotic receptor.[30,31] Whether methadone (or all opioids) activates steroid and xenobiotic receptor in a similar manner is not known. However, the high prevalence of low vitamin D status in this study raises the question of whether patients in MMT with chronic pain may have low vitamin D status.

Low vitamin D status for any population is important because of its role in a number of other medical conditions. Low vitamin D status is associated with muscle weakness that can lead to lower pulmonary function[32] and higher risk of falls.[33] Vitamin D is vital for bone health. Low vitamin D status is also associated with cardiovascular disease[34,35] and metabolic syndrome.[36] Moreover, vitamin D receptors are found in a variety of tissues in the body including colon, breast, ovarian, and prostate; sufficient vitamin D may be protective against cancers in these areas.[37–39]

To treat vitamin D deficiency, 50,000 IU of vitamin D_2 once a week for 8 weeks is often effective. Reoccurrence of vitamin D deficiency can safely be prevented with 50,000 IU every other week. Another approach to prevention is daily supplementation of 1000 IU of vitamin D to maintain serum levels above 30 ng/mL.[40]

The results of this study should be interpreted with the following limitations. First, because the study was conducted in the summer and fall when 25-hydroxyvitamin D levels are generally highest, a greater proportion of the study sample is likely to have low vitamin D at the end of winter when sunlight exposure is at it lowest. Second, although we did not find differences between the study sample and the clinic sample in terms of age, gender, and race/ethnicity, patients with bone pain may have been more likely to participate in a study on "bone health" leading to an overestimation of the incidence of low vitamin D. Third, the study did not include a comparison group of patients not receiving methadone, but rather notes a comparable historical control group. Thus, this study does not provide definitive evidence but rather suggests that the vitamin D status of MMT patients is worse than those not in MMT. Finally, it is possible that these findings may not generalize to other opioid-dependent populations, because most MMT patients are required to leave their house to attend clinic for methadone dosing on a daily basis where they are potentially exposed to sunlight. Further study should examine low vitamin D status in other opioid-dependent populations including those in other forms of opioid treatment (ie, buprenorphine) and those not in addiction treatment.

In conclusion, low vitamin D status was common in patients receiving MMT and unawareness of vitamin D status was almost universal. Low vitamin D status is another largely unrecognized medical comorbidity of patients with addictions, a group with complex medical problems. These findings merit further investigation into the need and benefits of vitamin D supplementation in MMT patients.

REFERENCES

1. Institute of Medicine. Improving the Quality of Health Care for Mental and Substance-Use Conditions. Washington, DC: The National Academies Press; 2006.
2. Brands B, Blake J, Marsh D. Changing patient characteristics with increased methadone maintenance availability. *Drug Alcohol Depend.* 2002;66:11–20.
3. Harris S, Soteriades E, Coolidge J, et al. Vitamin D insufficiency and hyperparathyroidism in a low income, multiracial, elderly population. *J Clin Endocrinol Metab.* 2000;85:4125–4130.
4. Nesby-O'Dell S, Scanlon KS, Cogswell ME, et al. Hypovitaminosis D prevalence and determinants among African American and white women of reproductive age: Third National Health and Nutrition Examination Survey, 1988–1994. *Am J Clin Nutr.* 2002;76:187–192.
5. Tangpricha V, Pearce EN, Chen TC, et al. Vitamin D insufficiency among free-living healthy young adults. *Am J Med.* 2002;112:659–662.
6. Plotnikoff GA, Quigley JM. Prevalence of severe hypovitaminosis D in patients with persistent, nonspecific musculoskeletal pain. *Mayo Clin Proc.* 2003;78:1463–1470.
7. Venning G. Recent developments in vitamin D deficiency and muscle weakness among elderly people. *BMJ.* 2005;330:524–526.
8. Dietrich T, Joshipura KJ, Dawson-Hughes B, et al. Association between serum concentrations of 25-hydroxyvitamin D3 and periodontal disease in the US population. *Am J Clin Nutr.* 2004;80:108–113.
9. Novick DM, Richman BL, Friedman JM, et al. The medical status of methadone maintenance patients in treatment for 11–18 years. *Drug Alcohol Depend.* 1993;33:235–245.
10. Titsas A, Ferguson MM. Impact of opioid use on dentistry. *Aust Dent J.* 2002;47:94–98.
11. Hannan MT, Litman HJ, Araujo AB, et al. Serum 25-hydroxyvitamin D and bone mineral density in a racially and ethnically diverse group of men. *J Clin Endocrinol Metab.* 2008;93:40–46.
12. Kim TW, Alford DP, Malabanan A, et al. Low bone density in patients receiving methadone maintenance treatment. *Drug Alcohol Depend.* 2006;85:258–262.
13. Pedrazzoni M, Vescovi PP, Maninetti L, et al. Effects of chronic heroin abuse on bone and mineral metabolism. *Acta Endocrinol (Copenh).* 1993;129:42–45.
14. Fahrleitner A, Dobnig H, Obernosterer A, et al. Vitamin D deficiency and secondary hyperparathyroidism are common complications in patients with peripheral arterial disease. *J Gen Intern Med.* 2002;17:663–669.
15. Peris P, Pares A, Guanabens N, et al. Bone mass improves in alcoholics after 2 years of abstinence. *J Bone Miner Res.* 1994;9:1607–1612.
16. Arnsten JH, Freeman R, Howard AA, et al. HIV infection and bone mineral density in middle-aged women. *Clin Infect Dis.* 2006;42:1014–1020.
17. Holick MF. Vitamin D deficiency. *N Engl J Med.* 2007;357:266–281.
18. Adams JS, Kantorovich V, Wu C, et al. Resolution of vitamin D insufficiency in osteopenic patients results in rapid recovery of bone mineral density. *J Clin Endocrinol Metab.* 1999;84:2729–2730.
19. Dawson-Hughes B, Heaney RP, Holick MF, et al. Estimates of optimal vitamin D status. *Osteoporos Int.* 2005;16:713–716.
20. Mertens J, Lu YW, Parthasarathy S, et al. Medical and psychiatric conditions of alcohol and drug treatment patients in an HMO: Comparison with matched controls. *Arch Int Med.* 2003;163, 2511–2517.
21. Chen TC, Turner AK, Holick MF. Methods for the determination of the circulating concentration of 25-hydroxyvitamin D. *J Nutr Biochem.* 1990;1:315–319.
22. Curtis JR, Smith B, Weaver M, et al. Ethnic variations in the prevalence of metabolic bone disease among HIV-positive patients with lipodystrophy. *AIDS Res Hum Retroviruses* 2006; 22:125–131.
23. Kakarala RR, Chandana SR, Harris SS, et al. Prevalence of vitamin D deficiency in uninsured women. *J Gen Intern Med.* 2007;22:1180–1183.
24. Holick MF, Matsuoka LY, Wortsman J. Age, vitamin D, and solar ultraviolet. *Lancet.* 1989;2:1104–1105.
25. Langrod J, Lowinson J, Ruiz P. Methadone treatment and physical complaints: A clinical analysis. *Int J Addict.* 1981;16:947–952.
26. Larson MJ, Paasche-Orlow M, Cheng DM, et al. Persistent pain is associated with substance use after detoxification: A prospective cohort analysis. *Addiction.* 2007;102:752–760.
27. Rosenblum A, Joseph H, Fong C, et al. Prevalence and characteristics of chronic pain among chemically dependent patients in methadone maintenance and residential treatment facilities. *JAMA.* 2003;289:2370–2378.
28. Karasz A, Zallman L, Berg K, et al. The experience of chronic severe

pain in patients undergoing methadone maintenance treatment. *J Pain Symptom Manage.* 2004;28:517–525.
29. Reginato AJ, Falasca GF, Pappu R, et al. Musculoskeletal manifestations of osteomalacia: Report of 26 cases and literature review. *Semin Arthritis Rheum.* 1999;28:287–304.
30. Pascussi JM, Robert A, Nguyen M, et al. Possible involvement of pregnane X receptor-enhanced CYP24 expression in drug-induced osteomalacia. *J Clin Invest.* 2005;115:177–186.
31. Zhou C, Assem M, Tay JC, et al. Steroid and xenobiotic receptor and vitamin D receptor crosstalk mediates CYP24 expression and drug-induced osteomalacia. *J Clin Invest.* 2006;116:1703–1712.
32. Black PN, Scragg R. Relationship between serum 25-hydroxyvitamin D and pulmonary function in the Third National Health and Nutrition Examination Survey. *Chest.* 2005;128:3792–3798.
33. Broe KE, Chen TC, Weinberg J, et al. A higher dose of vitamin D reduces the risk of falls in nursing home residents: A randomized, multiple-dose study. *J Am Geriatr Soc.* 2007;55:234–239.
34. Holick MF. Vitamin D: Important for prevention of osteoporosis, cardiovascular heart disease, type 1 diabetes, autoimmune diseases, and some cancers. *South Med J.* 2005;98:1024–1027.
35. Wang TJ, Pencina MJ, Booth SL, et al. Vitamin D deficiency and risk of cardiovascular disease. *Circulation.* 2008;117:503–511.
36. Ford ES, Ajani UA, McGuire LC, et al. Concentrations of serum vitamin D and the metabolic syndrome among U.S. adults. *Diabetes Care.* 2005;28:1228–1230.
37. Bertone-Johnson ER, Chen WY, Holick MF, et al. Plasma 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2005;14:1991–1997.
38. Flanagan JN, Young MV, Persons KS, et al. Vitamin D metabolism in human prostate cells: Implications for prostate cancer chemoprevention by vitamin D. *Anticancer Res.* 2006;26:2567–2572.
39. Gorham ED, Garland CF, Garland FC, et al. Vitamin D and prevention of colorectal cancer. *J Steroid Biochem Mol Biol.* 2005;97:179–194.
40. Holick MF. Vitamin D2 is as effective as vitamin D3 in maintaining circulating concentrations of 25-hydroxyvitamin D. *J Clin Endocrinol Metab.* 2008;93:677–681.

DIABETES TECHNOLOGY & THERAPEUTICS
Volume 11, Number 4, 2009

DOI: 10.1089/dia.2008.0060

Self-Monitoring of Blood Glucose with Finger Tip Versus Alternative Site Sampling: Effect on Glycemic Control in Insulin-Using Patients with Type 2 Diabetes

Philip E. Knapp, M.D., M.S.,[1] Kara M. Showers, M.P.H.,[1] Jenna C. Phipps, M.P.H.,[1] Jeanne L. Speckman, M.Sc.,[2] Elliot Sternthal, M.D.,[1] Karen M. Freund, M.D., M.P.H.,[3] Arlene S. Ash, Ph.D.,[2] and Caroline M. Apovian, M.D.[1]

Abstract

Objective: This study compared glycemic control in finger tip versus forearm sampling methods of self-monitoring of blood glucose (SMBG).

Research Design and Methods: One hundred seventy-four insulin-using patients with type 2 diabetes were randomized to SMBG using either finger-tip testing (FT) or forearm alternative site testing (AST) and followed up for 7 months. Hemoglobin A1C (HbA1C) was measured at baseline, month 4, and month 7. The study was designed to test the noninferiority of the AST method for the primary end point of change in HbA1C from baseline to month 7. Adherence with the testing schedule and frequency of hypoglycemic episodes were also measured.

Results: The FT ($n = 85$) and AST ($n = 89$) groups each had significant decreases in mean HbA1C from baseline to month 7 (FT, $-0.4 \pm 1.4\%$, $P = 0.008$; AST, $-0.3 \pm 1.2\%$, $P = 0.045$), and noninferiority between groups was demonstrated with a margin of equivalence of 0.5 ($P = 0.043$). There was no observable difference in HbA1C change between the groups ($P = 0.442$). Adherence was better in the FT (87%) than the AST (78%) group ($P = 0.003$), which may have been because of the difficulty some subjects had in obtaining blood samples for AST. The number of hypoglycemic episodes was too small to assess for a difference between groups.

Conclusions: SMBG by the AST, rather than FT, method did not have a detrimental effect on long-term glycemic control in insulin-using patients with type 2 diabetes. Although adherence with testing was expected to be better in the AST group, it was actually better in the FT group.

Introduction

Individuals with diabetes have traditionally obtained samples for self-monitoring of blood glucose (SMBG) using lancets to prick their fingertips, but this process can be painful and may be a barrier to testing with adequate frequency.[1] In the last several years, a new generation of SMBG devices has been developed that allows the use of far smaller volumes of blood than previous devices ($<1\ \mu$L). These newer meters allow patients to perform SMBG using samples obtained from skin sites other than the fingertip, such as the forearm or thigh, which are less vascular than the fingertips and yield less blood after skin puncture. They are also less densely innervated with pain receptors than the fingertips, which allows for less painful sampling.[2,3] Alternative site testing (AST) has been widely marketed by the manufacturers of SMBG meters as a less painful alternative to finger-tip testing (FT). It has also been suggested that AST might improve patients' adherence to SMBG regimens, potentially improving their glycemic control.[2,4–7] However, whether the option to use AST actually improves adherence has not been adequately answered. One small nonrandomized study suggested that it does,[6] while a larger crossover study showed no such difference.[8]

Thus far in the evaluation of AST it has also been noted that this method can yield results that are significantly different from both FT and reference blood glucose measures at times that blood glucose is changing rapidly, such as after a glucose load,[5,9,10] after meals,[4,11,12] after the administration of

[1]Center for Endocrinology, Diabetes and Nutrition and Weight Management, Section of Endocrinology, Evans Department of Medicine; [2]Health Care Research Unit and Section of General Internal Medicine; and [3]Women's Health Unit and Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts.

This study is identified through http://clinicaltrials.gov as NCT00207207.

insulin,[9,10] and after exercise.[11] These findings have been attributed to a delayed equilibration of arm and thigh skin with arterial blood glucose resulting from lesser blood flow.[10] Manufacturers of SMBG devices have recommended preparatory rubbing or heating of AST sites to increase blood flow and mitigate the difference due to equilibration time, but this may not be effective.[11] No significant difference has been found between AST and FT measurements in the fasting or steady state.[4,5,11–14]

The delay in equilibration in arm and thigh testing appears to be greatest between 60 and 90 min after a meal or glucose load[4,5,9–12] but may persist for as long as 240 min after a combination of glucose and insulin.[10] These findings have led to serious concerns about the potential failure of AST to identify hypoglycemia,[7,9,10,14,15] and the phenomenon has led the Food and Drug Administration to recommend clear labeling and instructions for patients not to use AST when hypoglycemia is suspected or at times when blood glucose may be changing rapidly.[16] Another potential consequence of the lag in AST equilibration is that postprandial blood glucose levels could be systematically underestimated. Postprandial measurements are being increasingly used in clinical practice because they contribute significantly to long-term control[17–19] and can guide later mealtime insulin dosing. Underestimation in measuring postprandial blood glucose with AST could hamper efforts at glycemic control.

When AST is used in clinical care, it is generally introduced as an option for patients to use for some but not all of their SMBG tests in order to reduce the cumulative discomfort from multiple finger sticks. Our goal was to explore the question of whether changing from exclusive finger testing, the standard of care in patients with diabetes using insulin, to introduce an option to use AST in addition to FT would lead to worsened control as a result of differences between AST and FT readings. The only previous study to have compared long-term control between these two methods found no difference between them, but its study population was in such good initial control that a meaningful difference would have been difficult to detect.[8] We hypothesized that the option to use AST would not lead to poorer glycemic control as assessed by hemoglobin A1C (HbA1C) in a diverse population of insulin-using patients with type 2 diabetes with varying levels of initial glycemic control. As a secondary consideration, we planned to track adherence to testing to determine whether use AST might improve adherence to testing and thereby have the potential to positively affect long-term glycemic control. The results were previously presented in abstract form at the 2006 Endocrine Society Annual Meeting.[20]

Research Design and Methods

Population

The study was conducted between July 2003 and December 2005 at Boston University Medical Center, Boston, MA, with subjects recruited from the Center for Endocrinology, Diabetes and Nutrition and Weight Management and affiliated primary care practices. Newspaper advertising was also employed to recruit subjects from outside the medical center. Potential subjects were eligible for participation if they were 18–70 years of age with type 2 diabetes, using insulin, and performing SMBG measurements. Exclusion criteria included type 1 diabetes, prior use of AST, pregnancy, and serious comorbid illness (unstable cardiovascular disease or metastatic cancer). Potential subjects were also excluded if, within the past year, they had a hypoglycemic episode requiring urgent medical attention, resulting in cognitive impairment, or a lack of symptoms during a hypoglycemic episode. The Institutional Review Board at Boston University and the Food and Drug Administration's Research Involving Human Subjects Committee each approved the protocol. All subjects gave verbal and written informed consent prior to study entry.

Study design

This was a randomized, parallel group trial consisting of nine visits. At the screening visit, information was collected on baseline characteristics as listed in Table 1, and subjects were asked to return approximately 2 weeks later for the randomization/training visit. Those who arrived for this visit were randomized to either the FT or the arm AST group using a block randomization stratified by six strata of initial HbA1C to ensure similar mean initial HbA1C in both groups. The strata of HbA1C (%) were <7.0%, 7.0–7.5%, 7.5–8.0%, 8.0–8.5%, 8.5–9.5%, and >9.5%. All SMBG monitoring in the study was performed using the OneTouch® Ultra® device (Lifescan®, Milpitas, CA). Subjects not already using this device were provided with a new one. When subjects were not able to obtain an adequate number of test strips as part of their usual health coverage, they were provided with free strips donated by the manufacturer.

Once randomized, each subject received a 30-min training session from a qualified diabetes nurse educator in the use of the SMBG device, including device calibration and settings adjustment. For subjects randomized to the AST group, the training also included instruction on obtaining samples from the forearm. Subjects in the AST group were asked to use AST as much as possible; however, because of the potential limitations of AST in detecting hypoglycemia, they were instructed to use FT when experiencing symptoms of hypoglycemia and to repeat any AST reading <5.55 mmol/L using FT. They were also told that if they had difficulties obtaining a blood sample from arm puncture on any particular occasion, it was acceptable to substitute a finger test. As intended, this resulted in a mixture of arm and finger testing in the AST group, which is consistent with actual clinical use of AST.

Subjects were given standardized SMBG log sheets that prompted them to test a minimum of three times per day: before breakfast, before dinner, and 2 h after dinner. The log sheets were designed to allow some flexibility such that the subject or their diabetes provider could alter the timing of tests if the individual subject's meal or insulin dosing schedule so required. Space was also provided to record test results for episodes of suspected hypoglycemia. Adherence was measured by counting the number of tests recorded, regardless of timing or method, as well as the number of tests requested for each log sheet (21 for a full week). When a test was repeated at an individual time point because of suspicion of hypoglycemia, only one test was counted.

Subjects were asked to return for study visits at 1, 2, 3, 4, 5, and 7 months after the randomization/training visit. At each follow-up visit the study coordinator collected completed

Table 1. Baseline Characteristics in the FT and Forearm AST Groups

	FT group (n = *85*)	*AST group* (n = *89*)	P *value*
Age (years)	53.2 ± 9.5	53.1 ± 10.2	0.959
Body mass index (kg/m^2)	35.9 ± 9.6	35.9 ± 9.2	0.975
Waist circumference (inches)	44.4 ± 7.2	45.3 ± 6.8	0.412
Female	52 (61%)	42 (47%)	0.064
Race/ethnicity			
African American	39 (46%)	48 (54%)	
Black Caribbean	7 (8%)	9 (10%)	
White/Caucasian	21 (25%)	24 (27%)	0.162[a]
Hispanic/Latino	12 (14%)	8 (9%)	
Asian/Pacific	2 (2%)	0 (0%)	
Highest educational level achieved			
Completed high school/GED	30 (35%)	35 (39%)	
Some post-secondary	34 (40%)	33 (37%)	0.972
Baseline HbA1C (%)	8.8 ± 2.2	8.7 ± 2.1	0.649
Years with diabetes	12.0 ± 9.8	12.7 ± 9.1	0.643
SMBG tests prior to study			
<1 per day	8 (9%)	10 (11%)	
1–2 per day	40 (47%)	51 (57%)	0.258
≥3 or more per day	37 (44%)	28 (32%)	
Frequency of insulin injections			
1 per day	12 (14%)	12 (14%)	
2 per day	37 (44%)	34 (38%)	0.633
≥3 per day	33 (39%)	37 (42%)	
Using oral diabetes agent (plus insulin)	52 (63%)	47 (57%)	0.374

Continuous variables were expressed as mean ± SD values, and categorical variables as *n* (%). No difference between groups was considered significant. Baseline characteristics were also compared between the total number of completers ($n = 135$) versus non-completers ($n = 39$). Non-completers were slightly younger, on average, than completers (50.0 vs. 54.1 years old, $P = 0.02$) and had a trend towards fewer insulin injections per day (2.3 ± 0.8 vs. 2.6 ± 1.1, $P = 0.10$). There were no other significant differences noted in baseline characteristics between completers and non-completers, including initial mean HbA1C (8.76 ± 2.10% and 8.72 ± 2.22%, respectively).

[a]By Fisher's Exact Test.

SMBG log sheets and gave out new sheets for use until the next visit. At each visit the study coordinator also recorded the 30-day glucose average from the SMBG device. At the 1-, 3-, and 5-month visits, subjects were also seen by one of nine diabetes care providers (physician or nurse practitioner). Subjects were asked to show their SMBG log sheets from the preceding month to the provider who used this information to modify the treatment regimen as part of their routine diabetes care. Because of the nature of the intervention, neither subjects nor providers could be blinded to group assignment. At months 4 and 7 subjects had blood drawn for HbA1C measurement. These HbA1C values were therefore obtained 3 and 6 months after the first provider visit in which providers had the opportunity to make management decisions based on study SMBG values. HbA1C measurements were made by gas chromatography using the Bio-Rad (Hercules, CA) Variant V-II instrument. Subjects were compensated with a total of $170 for participation.

Predetermined withdrawal criteria included a serious hypoglycemic episode during the course of the study (defined as hypoglycemia requiring urgent medical attention or causing seizure or loss of consciousness) or inability to comply with the intended timeline.

Sample size and statistical analysis

The primary hypothesis was that glycemic control in the AST group would not be meaningfully worse than in the FT group. In planning for the noninferiority test we defined a clinically significant difference in HbA1C, the margin of equivalence, to be 0.5%. We powered our study to test the hypothesis that the change from baseline to month 7 HbA1C measurement in the AST group would be no more than 0.5 units higher (worse) than in the FT group. Assuming an SD in HbA1C change of 1.3 (which was an estimate obtained from review of multiple studies using HbA1C as an end point) and using a one-sided rejection rule with alpha = 0.05, we calculated that a sample of 66 subjects in each group completing the study would result in 71% power to detect noninferiority, with margin of equivalence = 0.5.

The goals for diabetes management depend on the level of control already achieved. For patients already in good control, the goal is to maintain the same level of control, and for those in poor control, clinicians strive to improve the HbA1C. Therefore, post hoc analyses were performed to compare the difference in HbA1C change between the two groups stratified by initial level of glycemic control. Statistical analyses were performed using SAS software (PC-SAS version 8, SAS Institute, Cary, NC). Continuous variables were assessed for normality of distribution and compared using two-tailed *t* tests or analysis of variance unless otherwise noted. The paired-sample *t* test was used for within-group comparisons, and independent sample *t* test was used to compare between groups. Categorical variables were compared using the χ^2 test unless otherwise noted. All significance tests were performed at the alpha = 0.05 level.

Results

Population characteristics

Two hundred thirty-three potential subjects were screened in person. Of these, 26 were ineligible (Fig. 1). Of the 233 patients screened, 207 expressed initial interest. One hundred seventy-four returned for the randomization visit and were enrolled at that time; 85 were randomized to the FT group and 89 to the arm AST group. Of those randomized, 71 (83%) in the FT group and 64 (72%) in the AST group completed the study. The higher dropout/withdrawal rate in the AST group approached statistical significance ($P = 0.07$). Baseline characteristics were similar for the AST and FT groups (Table 1).

Correlation between SMBG and HbA1C

HbA1C reflects the average glucose concentration over a 3-month period with weighting toward the most recent blood glucose levels, such that levels within the 1-month prior to the HbA1C analysis account for approximately 50% of the measured value.[21] To assess the correlation between month 4 HbA1C and SMBG readings by each testing method, we calculated a weighted average from the monthly 30-day averages generated by the SMBG meters. Thirty-day average SMBG were recorded from each subject's meter at each of the monthly visits for the 3 months prior to the month 4 HbA1C measurement. The weighted average was then calculated as (0.5×month 3 average) + (0.35×month 2 average) + (0.15×month 1 average). HbA1C was highly correlated with the SMBG weighted average in both groups. The Pearson correlation coefficient (r) for the FT group was 0.84 ($n = 66$, $P < 0.0001$) and for the AST group was 0.73 ($n = 64$, $P < 0.0001$).

Long-term glycemic control

The central hypothesis was that glycemic control, as assessed by HbA1C, would not be meaningfully worse in the AST than in the FT group. We tested this with an intention-to-treat, noninferiority analysis. Missing values for month 4 and month 7 HbA1C in subjects who did not complete the study were taken from actual measurements when available in the medical records. When later HbA1C values were not available, the initial HbA1C value was carried forward. For the FT group HbA1C values were based on actual measurements in 94% of subjects at month 4 and 91% at month 7. For the AST group HbA1C values were based on actual measurements in 90% at month 4 and 83% at month 7. The intention-to-treat analysis demonstrated a statistically significant improvement in glycemic control between baseline and month 7 in both groups (Table 2). The noninferiority test was conducted with pooled variance after assessing for equality of variances. We

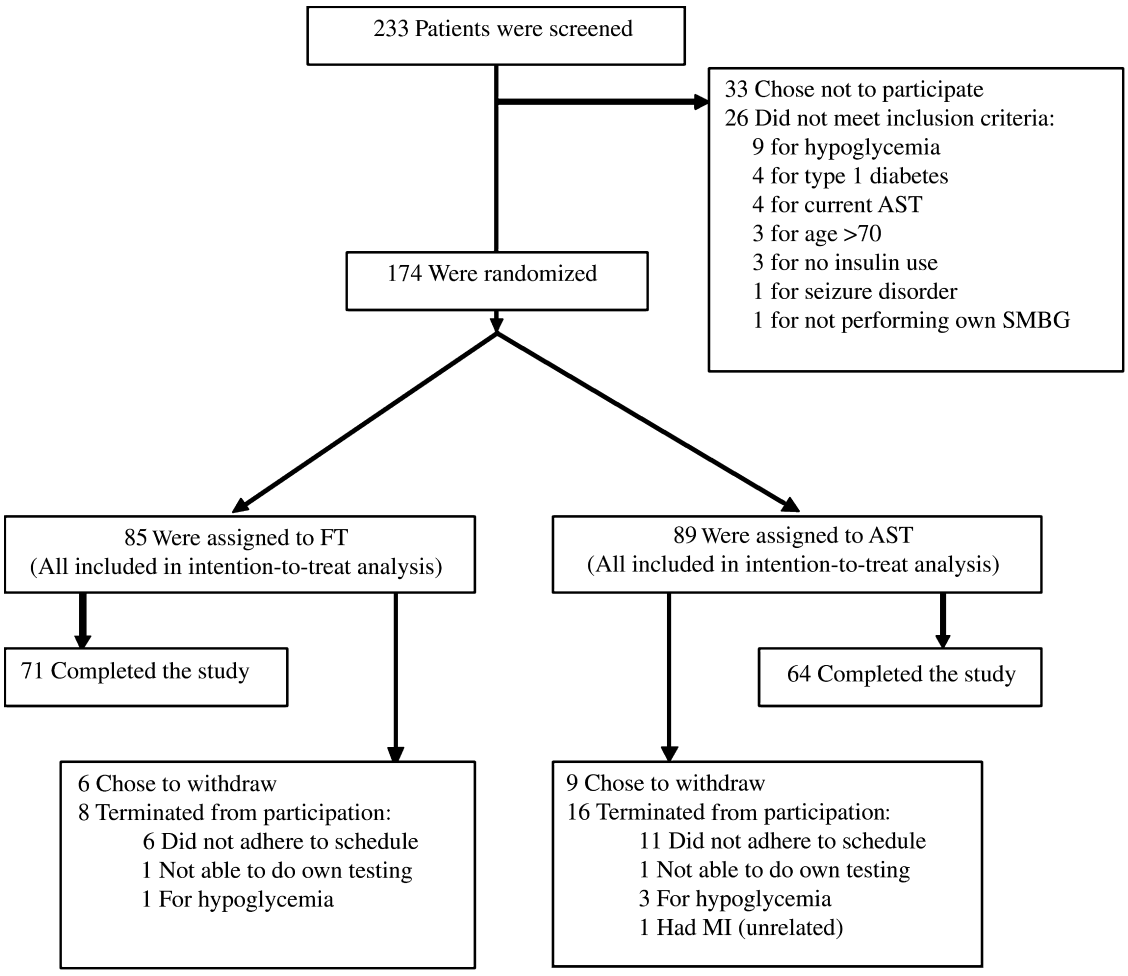


FIG. 1. Subject enrollment and follow-up. MI, myocardial infarction.

Table 2. Percentage HbA1C over Time by FT Versus AST Groups Stratified by Level of Initial Glycemic Control

	n	HbA1C (%) (mean ± SD)				
		Baseline (T0)	*Month 4 (M4)*	*Month 7 (M7)*	*Difference (M7–T0)*	P *value (M7–T0)*
All subjects						
FT	85	8.8 ± 2.2	8.4 ± 1.9	8.4 ± 1.7	−0.4 ± 1.4	0.008
AST	89	8.7 ± 2.1	8.3 ± 1.8	8.4 ± 1.8	−0.3 ± 1.2	0.045
Good initial control (baseline HbA1C ≤7.0%)						
FT	18	6.4 ± 0.4	6.7 ± 0.5	6.8 ± 0.5	+0.4 ± 0.7	0.024
AST	21	6.3 ± 0.6	6.5 ± 0.9	6.7 ± 0.7	+0.4 ± 0.6	0.040
Intermediate initial control (baseline HbA1C 7.0–8.5%)						
FT	26	7.8 ± 0.4	7.6 ± 1.0	7.9 ± 1.1	+0.1 ± 1.0	0.705
AST	26	7.8 ± 0.4	7.8 ± 0.8	7.9 ± 0.9	+0.1 ± 0.9	0.493
Poor initial control (baseline HbA1C >8.5%)						
FT	41	10.5 ± 1.9	9.7 ± 1.8	9.5 ± 1.6	−1.0 ± 1.6	<0.001
AST	42	10.4 ± 1.6	9.5 ± 1.6	9.6 ± 1.6	−0.8 ± 1.4	<0.001

There was a statistically significant improvement in HbA1C from T0 to M7 in both the FT and the AST groups but no significant difference between the groups in the magnitude of this change ($P = 0.442$). Stratification into three levels of initial glycemic control demonstrated that only the stratum in poor control at baseline actually showed improvement during the course of the study. There was no significant difference between randomization groups in the good, intermediate, or poor initial control strata ($P = 0.876$, 0.839, and 0.413, respectively).

rejected the hypothesis that the change from baseline to month 7 HbA1C for the AST group was at least 0.5 units higher than for the FT group ($P = 0.043$). There was no significant difference in the degree of improvement between FT and AST groups ($P = 0.442$) (Table 2). Findings were similar when the analysis was repeated excluding the imputed values for month 7 HbA1C.

Because the goals of care for patients depend on initial level of glycemic control, we provide an analysis of HbA1C change by group stratified for starting HbA1C (Table 2). For subjects in good initial control (HbA1C <7.0%) there was a slight increase in HbA1C from baseline to month 7 in both the FT and AST groups, but the mean HbA1C remained below 7.0% in both groups. For subjects in intermediate glycemic control (7.0–8.5%) there was no change in HbA1C from baseline in either group, and for subjects in poor control (HbA1C>8.5%) there was a significant improvement in both groups. There was no detectable difference in mean HbA1C or change in HbA1C between the FT and AST groups in any of the strata of initial glycemic control (sign test).

To confirm the lack of difference between FT and AST groups, we categorized subjects into improved, unchanged, or worsened glycemic control by change in HbA1C from baseline to month 7 of 0.25% or more. In the FT group 39 (46%) improved, 19 (22%) were unchanged, and 27 (32%) worsened. In the AST group 39 (44%) improved, 21 (24%) were unchanged, and 29 (33%) worsened. There was no difference in frequencies between groups ($P = 0.929$).

Adherence to SMBG schedule

In the AST group subjects were encouraged to use AST as much as possible but were asked to use FT for suspected or possible hypoglycemia or if they had difficulty obtaining blood from the arm. Tests were counted as completed regardless of whether they were performed on the arm or finger. Adherence overall was better in the FT compared to the AST group (Fig. 2).

Hypoglycemic episodes

Information on hypoglycemic episodes was collected using a strict definition. Events were only considered to be hypoglycemic episodes if they consisted of hypoglycemic symptoms followed by an SMBG test that confirmed blood glucose <4.44 mmol/L. As a result, the number of such events was

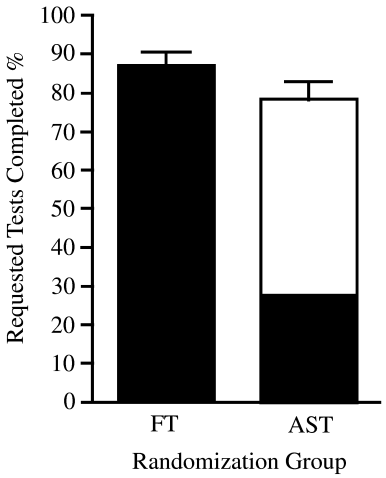


FIG. 2. Adherence to testing regimen by randomization group. Mean individual adherence with SMBG regimen in the FT and forearm AST groups was expressed as percentage of requested tests completed. Finger tests are shown in black, arm tests in white. Percentages were obtained by dividing the total number of SMBG tests performed by the total number of tests requested (three tests per day for each day the subject was in the study). The denominator reflected returned log sheets only. Mean percentage of arm tests within the AST group was 65%. Mean overall adherence was higher in the FT group than in the AST group: 87% for FT (95% CI 83.2%, 90.4%) and 78% for AST (95% CI 73.5%, 82.9%) ($P = 0.003$). Error bars represent 95% confidence intervals.

very small and concentrated in a few individuals. The mean number of hypoglycemic episodes per month was 0.183 in the FT group and 0.176 in the AST group with no significant difference between the two ($P = 0.16$, Wilcoxon rank-sum test). Only three subjects in each group reported an average of more than one hypoglycemic episode per month. Because of the small number of recorded hypoglycemic events we did not have adequate power to detect a difference in the number between groups.

Four subjects were withdrawn from the study because of a severe hypoglycemic episode requiring urgent medical attention: one in the FT group and three (two including seizures) in the AST group. While the occurrence of three out of four severe hypoglycemic events in the AST group is concerning, this is not sufficient to suggest a relationship between AST and such episodes.

Discussion

Glycemic control improved in both groups over the course of their participation in the study, primarily because of substantial improvements in subjects who began the study in poor control (HbA1C>8.5%). Presumably, this was caused by an increased attention to their diabetes and increased frequency of SMBG testing through their participation in the study. We did not observe a difference in the degree of improvement between the FT and AST groups.

Despite the frequently postulated but rarely tested idea that the option for AST might improve adherence with SMBG, AST actually reduced the degree of adherence in our study population (Fig. 1). The fairly regimented nature of our study procedure in which subjects were asked to return their SMBG log sheets to the study coordinator on a monthly basis appeared to lead both groups to have relatively high adherence rates, but the rate in the AST group of 78% tests completed/tests requested was lower than the 87% rate in the FT group ($P = 0.003$). While this difference is statistically significant, it is small in real terms, and it conflicts with a prevailing opinion that AST can be expected to improve adherence with SMBG. One possible explanation is that subjects in the AST group frequently reported difficulty in obtaining blood from the arm despite having had specific training in this method with a qualified nurse educator. Many AST subjects reported having to make multiple attempts to obtain an adequate sample, and several reported frustration in this regard.

Limitations and other considerations

Our sample size was designed to give our study sufficient power to detect a difference between the FT and AST groups as small as 0.5 units on the HbA1C percentage scale. A larger study would be required to conclusively rule out the possibility of a smaller change in mean HbA1C caused by the use of AST. The number of recorded hypoglycemic events in our study was small enough that we can neither prove nor exclude an effect of AST. A larger study is also needed to better characterize the risk for hypoglycemia or the failure to detect it.

In choosing a SMBG testing schedule we attempted to strike a balance between the unique requirements for the timing of tests for each individual with diabetes based on his or her insulin regimen and meal schedule, on one hand, and standardization between groups, on the other. We chose a standard regimen including two premeal tests and one postprandial test per day while allowing individual patients and providers to deviate from this schedule if they chose. It is possible that a study using an SMBG regimen more heavily weighted toward postprandial measurements would have different results regarding the impact of AST.

In designing the study, we also attempted to strike a balance between instructing the AST group to exclusively perform arm tests and giving them the flexibility to choose the method on a test-by-test basis. AST is generally suggested for use in clinical practice as an option for patients rather than an exclusive testing modality, but complete flexibility may have led to few actual arm tests. We therefore chose to ask subjects in the AST group to utilize arm testing "as much as possible," but to use FT under the particular circumstances described. This was intended to maximize our ability to detect a difference between testing methods while maintaining the generalizability of our findings to real clinical practice.

Acknowledgments

Financial support for the study was provided by Food and Drug Administration contract number 223-03-8720 to C.M.A. Some SMBG meters and test strips were provided by Lifescan.

Author Disclosure Statement

No competing financial interests exist.

References

1. Vincze G, Barner JC, Lopez D: Factors associated with adherence to self-monitoring of blood glucose among persons with diabetes. Diabetes Educ 2004;30:112–125.
2. Fineberg SE, Bergenstal RM, Bernstein RM, Laffel LM, Schwartz SL: Use of an automated device for alternative site blood glucose monitoring. Diabetes Care 2001;24:1217–1220.
3. Tieszen KL, New JP: Alternate site blood glucose testing: do patients prefer it? Diabet Med 2003;20:325–328.
4. Ellison JM, Stegman JM, Colner SL, Michael RH, Sharma MK, Ervin KR, Horwitz DL: Rapid changes in postprandial blood glucose produce concentration differences at finger, forearm, and thigh sampling sites. Diabetes Care 2002;25: 961–964.
5. Peled N, Wong D, Gwalani S: Comparison of glucose levels in capillary blood samples obtained from a variety of body sites. Diabetes Technol Ther 2002;4:35–44.
6. Suzuki Y, Atsumi Y, Matusoka K: Alternative site testing increases compliance of SMBG (preliminary study of 3 years cohort trials). Diabetes Res Clin Pract 2003;59:233–234.
7. Fedele D, Corsi A, Noacco C, Prisco F, Squatrito S, Torre E, Iafusco D, Errico MK, Toniato R, Nicolucci A, Franciosi M, De Berardis G, Neri L: Alternative site blood glucose testing: a multicenter study. Diabetes Technol Ther 2003;5:983–989.
8. Bennion N, Christensen NK, McGarraugh G: Alternate site glucose testing: a crossover design. Diabetes Technol Ther 2002;4:25–33.
9. Jungheim K, Koschinsky T: Risky delay of hypoglycemia detection by glucose monitoring at the arm. Diabetes Care 2001;24:1303–1306.
10. Jungheim K, Koschinsky T: Glucose monitoring at the arm: risky delays of hypoglycemia and hyperglycemia detection. Diabetes Care 2002;25:956–960.
11. Bina DM, Anderson RL, Johnson ML, Bergenstal RM, Kendall DM: Clinical impact of prandial state, exercise, and site

preparation on the equivalence of alternative-site blood glucose testing. Diabetes Care 2003;26:981–985.
12. Lee DM, Weinert SE, Miller EE: A study of forearm versus finger stick glucose monitoring. Diabetes Technol Ther 2002;4:13–23.
13. Szuts EZ, Lock P, Malomo KJ, Anagnostopoulos A: Blood glucose concentrations of arm and finger during dynamic glucose conditions. Diabetes Technol Ther 2002;4:3–11.
14. Lucidarme N, Alberti C, Zaccaria I, Claude E, Tubiana-Rufi N: Alternate-site testing is reliable in children and adolescents with type 1 diabetes except at the forearm for hypoglycemia detection. Diabetes Care 2005;28:710–711.
15. Cembrowski G: Alternative site testing: first do no harm. Diabetes Technol Ther 2002;4:45–47.
16. Ginsberg B: The FDA reevaluates alternative site testing for blood glucose. Diabetes Technol Ther 2002;4:347–349.
17. De Veciana M, Major CA, Morgan MA, Asrat T, Toohey JS, Lien JM, Evans AT: Postprandial versus preprandial blood glucose monitoring in women with gestational diabetes mellitus requiring insulin therapy. N Engl J Med 1995;333: 1237–1241.
18. Avignon A, Radauceanu A, Monnier L: Nonfasting plasma glucose is a better marker of diabetic control than fasting plasma glucose in type 2 diabetes. Diabetes Care 1997;20: 1822–1826.
19. Monnier L, Lapinski H, Colette C: Contributions of fasting and postprandial plasma glucose increments to the overall diurnal hyperglycemia of type 2 diabetic patients: variations with increasing levels of HbA1C. Diabetes Care 2003;26:881–885.
20. Knapp PE, Clemente KM, Phipps JC, Sternthal ES, Speckman JL, Ash AS, Freund KM, Apovian CM: Self-monitoring of blood glucose by finger-tip vs. alternate site sampling: effect on long term glycemic control [abstract]. In: Program and Abstracts of the Endocrine Society Annual Meeting, June 24–27, 2006, Boston, MA. Chevy Chase, MD: Endocrine Society, 2006.
21. Rohlfing CL, Wiedmeyer HM, Little R, England JD, Tennill A, Goldstein DE: Defining the relationship between plasma glucose and HbA_{1c}: analysis of glucose profiles and HbA_{1c} in the Diabetes Control and Complications Trial. Diabetes Care 2002;25:275–278.

Address reprint requests to:
Philip E. Knapp, M.D.
Section of Endocrinology
Boston University Medical Center
88 East Newton Street, E201
Boston, MA 02118

E-mail: Philip.Knapp@bmc.org

Pediatric Clinicians Can Help Reduce Rates of Early Childhood Caries

Effects of a Practice Based Intervention

Nancy R. Kressin, PhD,†‡§ Martha E. Nunn, DDS, PhD,*¶ Harpreet Singh, RDH, MS,*¶ Michelle B. Orner, MPH,*§ Lori Pbert, PhD,‖ Catherine Hayes, DMD, DMSc,* ** Corinna Culler, RDH, MPH,*¶ Stephan R. Glicken, MD,†† Sean Palfrey, MD,‡‡ Paul L. Geltman, MD, MPH,*‡‡ Cynthia Cadoret, RDH, MPH,*¶ and Michelle M. Henshaw, DDS, MPH*¶*

Objective: Early childhood caries (ECC) is a serious and preventable disease which pediatric clinicians can help address by counseling to reduce risk.

Research Design: We implemented a multifaceted practice-based intervention in a pediatric outpatient clinic treating children vulnerable to ECC (N = 635), comparing results to those from a similar nearby clinic providing usual care (N = 452).

Intervention: We provided communication skills training using the approach of patient centered counseling, edited the electronic medical record to prompt counseling, and provided parents/caregivers with an educational brochure.

Outcome Measures: We assessed changes in provider knowledge about ECC after the intervention, and examined providers' counseling practices and incidence of ECC over time by site, controlling for baseline ECC, patient sociodemographics and parents'/caregivers' practice of risk factors (diet, oral hygiene, tooth-monitoring), among 1045 children with complete data.

Results: Provider knowledge about ECC increased after the intervention training (percentage correct answers improved from 66% to 79%). Providers at the intervention site used more counseling strategies, which persisted after adjustment for sociodemographic characteristics. Children at the intervention site had a 77% reduction in risk for developing ECC at follow up, after controlling for age and race/ethnicity, sociodemographics and ECC risk factors; $P \leq 0.004$.

Conclusions: The multifaceted intervention was associated with increased provider knowledge and counseling, and significantly attenuated incidence of ECC. If validated by additional studies, similar interventions could have the potential to make a significant public health impact on reducing ECC among young children.

Key Words: early childhood caries, physician-patient relations, physicians,' practice patterns, intervention studies

(*Med Care* 2009;47: 1121–1128)

Early childhood caries (ECC) is a serious, but preventable, form of tooth decay that affects children's primary dentition. Untreated, it can lead to serious illness, including abscesses, necessitating costly therapeutic interventions such as surgery with general anesthesia. The infection and pain caused by ECC can impair growth and weight gain,[1] cause speech, learning and eating problems, and increase school absenteeism,[2] negatively affecting children's quality of life.[3] Thus, a national public health goal is to reduce the prevalence of children with dental caries in primary teeth (Healthy People 2010 goal 21–1a).[4]

ECC disproportionately affects poor, racial/ethnic minority children.[5–8] Several risk factors including sugary diet, excessive bottle use, and poor oral hygiene have been identified.[9,10] ECC risk can be reduced by minimizing children's exposure to caries-promoting food and drinks, weaning from the bottle as early as possible, and regularly cleaning chil-

From the *Northeast Center for Research to Evaluate and Eliminate Dental Disparities, Boston, MA; †HealthCare Disparities Research Program, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA; ‡VA Boston Healthcare System, Boston, MA; §Center for Health Quality, Outcomes and Economic Research, VAMC, Bedford, MA; ¶Boston University School of Dental Medicine, Boston, MA; ‖Department of Medicine, Division of Preventive and Behavioral Medicine, University of Massachusetts Medical School, Worcester, MA; **Tufts University School of Dental Medicine, Boston, MA; ††Department of Pediatrics, Allegheny General Hospital, Pittsburgh, PA; and ‡‡Department of Pediatrics, Boston University School of Medicine, Boston, MA.

Supported by NIH/NIDCR by U54 DE14264, K24 DE00419, K23 DE00454, U54 DE019275, and U54 RR024381. Supported by Department of Veterans Affairs Health Services Research and Development Research Career Scientist Award (N.R.K.).

The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. No authors have any affiliation, financial agreement, or other involvement with any company whose product figures prominently in the submitted manuscript.

Reprints: Nancy R. Kressin, PhD, Section of General Internal Medicine, Crosstown Center, 2nd Floor, 801 Massachusetts Ave, Boston, MA 02118. E-mail: nkressin@bu.edu.

ISSN: 0025-7079/09/4711-1121

dren's teeth. It is vital that at-risk children and their caregivers be educated, advised, and counseled about prevention strategies. Thus, "counseling, reinforcement of health promoting behaviors with care givers of children, and intervention by dental and other professionals to improve parenting practices" have been called the best available means to prevent or mitigate ECC.[11]

To monitor children's teeth for ECC, the American Academy of Pediatric Dentistry recommends that all children begin regular dental visits by 1 year. Unfortunately, children most at risk for ECC face the greatest barriers to accessing health care in general,[12] and dental care in particular.[13] However, because most (88%) American children see a pediatrician annually,[14] pediatric clinicians could potentially counsel parents about decreasing children's ECC risk. Thus, the American Academy of Pediatrics has adopted a policy supporting the use of caries risk assessment and referral to a dental home.[15] However, pediatricians are not well trained to do so,[16,17] despite their belief that they have an important role in identifying dental problems and in counseling parents about caries prevention.[18,19]

Others have addressed the problem of ECC by teaching pediatricians to apply fluoride varnish,[20] or increasing their knowledge regarding the disease and ability to screen children for ECC risk.[21–26] Evidence is clear that increasing clinician knowledge does not necessarily lead to behavior change,[27] but notably absent from prior approaches are strategies to provide pediatric clinicians with the communication tools needed to effectively convey such information, or the skills needed to enlist parents/caregivers in behaviors to reduce children's ECC risk.

To capitalize on pediatricians' commitment to children's oral health, and to address the deficit in pediatricians' preparation to help prevent ECC[19,28] we developed and implemented a multifaceted pediatric practice-based intervention where children especially vulnerable to ECC receive care, building on the proven educational methodology of patient centered counseling (which has been successfully applied in other clinical settings[29–32]). Patient centered counseling has successfully changed provider behavior, effected changes in patients' risk-related behaviors, and, ultimately, improved clinical outcomes,[29–34] although the evidence for the effectiveness of such programs applied specifically to the oral health of young children is lacking.[28] Our goal was to assess the effects of this intervention on provider ECC counseling practices, and on children's subsequent development of ECC.

METHODS

Study Locations

Parents and children were recruited from the pediatric outpatient practices of 2 academic medical centers in Boston. The intervention site was selected because it serves underserved populations (primarily African-American and Latino children). The comparison site was another nearby urban hospital-based primary care pediatric clinical practice, serving a similar patient population but with more Asian Americans.

Study Participants

Patients

We asked parents/caregivers of children aged 6 months to no older than 5 years attending well-child visits to participate in the study (during the same time period for both sites), to ensure the children likely had at least some erupted teeth and thus were "at risk," and to limit our sample to younger children. Children were to be excluded from the study if they self-reported or had congenital oral anomalies, ectodermal dysplasias, or other diseases (other than ECC) upon examination affecting the dentition or oral mucosa, although no children were excluded for these criteria. At the intervention site, a total of 635 children were recruited after their providers had received the study training intervention, and 452 children were recruited at the comparison site.

Providers

We invited all attending pediatricians (not residents or medical students) who care for regular panels of patients ($N = 19$), as well as clinic nurses (RNs and NPs; $N = 14$), to participate in the 1-hour study training intervention, as each has the opportunity to provide anticipatory guidance counseling in the clinic. After offering multiple early morning and lunchtime sessions, and a $100 gift certificate incentive, 68% of the eligible physicians (13/19), and 100% of the eligible nurses (14/14) were trained. Although we did not conceal the purpose of the study, we did not explicitly describe it during the training.

Questionnaire Interview and Clinical Examination–Procedures and Measures

After the educational program was given to providers at the intervention site, and with the simultaneous initiation of recruitment at the comparison site, parents/caregivers of young children were approached in the clinic waiting room before a regular well-child visit, the study was described (respondents were not blinded to the study purpose but were blinded as to which group they were in), and those expressing interest were asked to provide informed consent. After the visit, participants completed an interview assessing demographic information including the child's date of birth, gender, race and ethnicity (Hispanic or not, following the US Census conventions for assessing ethnicity), the education and employment status of the parent, and language spoken at home. Study participants were again asked to participate in a similar interview and clinical examination, approximately 1 year later, at another well-child visit. In our analyses, we used baseline data on all variables except ECC, positing that follow up ECC would be a function of physician counseling regarding ECC at the baseline visit.

One component of the questionnaire was a "Patient Exit Interview" (PEI), a series of questions inquiring about the parent's discussions with the child's doctor or nurse. These assessed the degree to which the clinician covered the topics on which they had been trained to counsel regarding ECC risk reduction. PEIs have been demonstrated to accurately measure the actual content of clinical discussions through comparisons of audiotapes of such interactions to patient reports[35]

TABLE 1. Sociodemographic and Risk Factor Characteristics of the Sample, by Study Site

	Provider Training Intervention Site (n = 635)	Comparison Site (n = 452)	*P*
Age (%)			0.0962
<1 yr	1	3	
1 to <2 yr	55	55	
2 to <3 yr	25	26	
3–4 yr	19	16	
% male	51	53	0.4810
Caregiver employed (% yes)	57	69	<0.0001
Caregiver education (≥12 yr)	84	93	<0.0001
Race			<0.0001
% white	17	45	
% black	76	35	
% Asian	6	19	
Hispanic (%)	13	15	0.3811
Language spoken at home (% English)	40	74	<0.0001
Diet summary score*			<0.0001
Mean ± SD (Median)	3.2 ± 1.0 (3.6)	3.5 ± 1.1 (4.0)	
Range	0–6	0–6	
Hygiene summary score*			<0.0001
Mean ± SD (Median)	4.9 ± 1.0 (5.0)	4.5 ± 1.1 (5.0)	
Range	2–6	2–6	
Tooth-monitoring summary score*			0.0197
Mean ± SD (Median)	0.7 ± 0.7 (1.0)	0.9 ± 0.9 (1.0)	
Range	0–3	0–3	
ECC (baseline) %	5.8	6.4	0.664
ECC (follow up) %	17.7	31.7	0.086

*Baseline score; higher scores indicate better diet, hygiene and more tooth monitoring.

(see Table 1 for all PEI questions); they do not assess knowledge gained. The PEI score was a sum of the questions to which there was a positive response (range: 0–22).

The interview also included questions about the 3 dimensions important to preventing ECC (eg, "risk factors"). We assessed the child's diet and feeding behaviors (eg, "Does your child usually drink from a bottle?" and "How often does the child get a bottle in bed with something besides water?," 6 questions total; score range: 0–6), oral hygiene (eg, "Do you help your child brush his or her teeth?," "When brushing, do you use toothpaste with fluoride?": 6 questions total: score range: 0–6), and tooth monitoring (eg, "Do you ever check your child's teeth for spots or cavities?": 3 questions total; score range: 0–3). We created summary scores from the items assessing each dimension, by dichotomizing answers and summing the affirmative responses, provided that at least 75% of the questions were answered.

The research hygienist examined each child's dentition to identify ECC, recording both white and cavitated lesions, by tooth (because this was a research-focused interaction, no counseling regarding hygiene or diet was provided by the study hygienist). In our analyses, we define ECC as the more serious, irreversible cavitated lesions in any tooth, which require restorative treatment (vs. the reversible white lesions which do not).

Content of the Intervention

The intervention had 3 components: communication skills training, edits to the electronic medical record's (EMR's) anticipatory guidance section, and provision of an educational brochure. The communication skills training educational program was designed to enhance clinicians' ability to advise and counsel patients' parents or caregivers about decreasing risks for ECC. In a 1 hour training session led by experts in dentistry and patient centered counseling, pediatric clinicians (nurses and physicians) were taught, using a 1-page counseling algorithm handout, to address 3 primary dimensions with parents/caregivers: consuming foods and drinks that strengthen teeth and limit sugars (diet), toothbrushing/keeping teeth clean (hygiene), and monitoring teeth to detect the development of caries (tooth monitoring). Through didactic presentations and role play exercises, providers were asked to implement the 4A's: Assess the parent/caregiver's status on each of the dimensions, identifying barriers to each, Assist with addressing barriers to each behavior, Advise (or educate) about ECC and its etiology, and finally, Arrange for follow up (eg, making a dentist appointment; see the Fig. 1 for additional detail).

Immediately before and after the study intervention training sessions, we administered a 5 item multiple choice pre- and post-test to the participants to assess the effects on ECC knowledge. We also edited the anticipatory guidance section (a listing of topics that parents should be counseled about at each well child visit, which providers must document they followed) of the EMR to include age-appropriate information for each of the dimensions we trained the providers to address. Our edits were not the pop-up prompts typically used in EMR reminders but rather additions to the topics on which providers are to counsel. We also prepared an educational brochure that summarized these same areas for the parent/caregiver to address in caring for the child. These brochures were available in both clinics, but intervention providers were asked to distribute them, as part of their counseling.

Institutional Review Board

The Institutional Review Boards at both study sites approved this study's protocol.

Data Analysis

We compared sociodemographics, and summary scores for the ECC risk factors of diet, oral hygiene, and tooth monitoring practices by site using χ^2 tests of independence for the former and Mann-Whitney *U* tests for the latter 3 variables. We examined PEI items and overall scores by site, using the Mann-Whitney *U* test.

Then, we conducted 2 random effects least squares regressions to examine the independent effects of site on provider counseling, adjusting for sociodemographic characteristics, with and without controlling for ECC risk factors, which we anticipated might affect rates of clinician counseling. These analyses accounted for clustering of patients-within-provider.

Tooth brushing/keeping teeth clean

Assess	*Assist*
• Does child have own toothbrush?	• Offer toothbrush from dental study
• Brush teeth nightly? *Goal: establish habit!*	• How do you think you can get him to do that?
• Do you help child brush? (up to age 6)	• What do you think would work?
• Use toothpaste with fluoride? (pea sized amount)	• Find a tasty, colorful fluoridated toothpaste

Consume foods and drinks that strengthen teeth and limit sugars

Assess	*Assist*
• Child getting fluoride?	• Prescribe Fluoride drops/tablets
• Limiting sugary drinks and foods?	• How do you think you could do that?
• Limit bottle/sippy cup use to meal or snacktime? *(goal: decrease frequency of exposure to sugars – sippy cups aren't any better than bottles)*	• How could you limit these?
• Eliminating bedtime bottles by one year old (except water)?	• How could you eliminate bedtime bottles?

Monitor teeth (parent, doctor and dentist)

Assess	*Assist*
• Check teeth for spots?	• Here's how you can do it…(also can do while brushing)
• Have a dentist?	• Here's a referral list
• Seen the dentist yet? (first visit at 2 years)	• What would help you get to the dentist?

- **Summarize and develop plan**
  - You're doing a great job with _____, keep up the good work
  - [SHOW BROCHURE]. In the brochure, I've checked the things we've agreed you'll work on. Your plan is to…(e.g. put child to bed with only water; substitute fruit for snacks instead of candy)
  - **Arrange follow-up:** I'll check in with you at your next visit to see how you're doing.

Important Education Points About Cavities (Advise):
- A disease which can be seen as white, yellow, brown or gray spots on teeth
- Can lead to pain, trouble eating and talking, or to lost teeth, and infected baby teeth can lead to having infected permanent teeth
- Caused by germs that feed on sugars
- Can be prevented by keeping teeth clean, using fluoride, and limiting sugary/starchy foods and drinks

FIGURE 1. Patient Centered Counseling to Decrease ECC Risk: 4As: Assess, Assist, Advise, Arrange.

To determine the effect of the intervention on the development of ECC over time (defined as the presence/absence of an irreversible cavitated lesion in any tooth), survival analysis was used. A multiple mixed model (frailty) proportional hazards regression model was fit including only children free of ECC at baseline with physician treated as a random effect, adjusting for age group, race/ethnicity, caregiver employment status and educational level, whether English was spoken at home, and the ECC risk factors.

RESULTS

Only parents/caregivers who could be interviewed in English were included (study staff made this determination; 9% of the parents screened at the comparison site and 7% of those screened at the intervention site were excluded for this reason). Refusal rates were 3% at the intervention site, and 14% at the comparison site. Of the children recruited, the study hygienist was unable to conduct a clinical examination with 42 children, 28 (67%) of these patients were from the intervention clinic and 14 (33%) were from the control clinic. We excluded these 42 children from analyses using clinical data, leaving an analysis sample of 1045. The number of patients seen by intervention providers ranged from 1 to 134 for each provider; mean: 32, median: 25. The average (mean) number of patients seen by all providers was 18 (out of 62). The median was 7.

Provider ECC Knowledge

Prior to the training, participants answered 66% of the questions correctly on a 5 item ECC knowledge test. Subsequent to the intervention, this rate increased to 79%. Improvements in knowledge were observed on 3 of the items–those focused on knowledge about dental caries being the most common infectious disease of childhood, understanding of risk factors for dental caries in infants and toddlers, and the recommended age for a first dental visit, but many clinicians still did not realize that ECC risk is related to the oral health of the caregiver.

Characteristics of the Patients

Sociodemographic Characteristics

The mean initial age of the sample was just under 2 years, with no differences by site (not shown; 1.93 years vs. 1.87, $P = 0.20$). There was no difference in age group or gender distribution across sites (Table 2). More parents from the comparison site were employed (69% vs. 56.7%, $P < 0.0001$). The racial group distribution, but not the proportion of Hispanics, differed significantly by site ($P < 0.0001$)–about half of the comparison site sample were white, while over three-quarter of the intervention site were black. More parents from the comparison site reported that English was the primary language spoken at home (73.6% vs. 39.5%; $P <$

TABLE 2. Patient Exit Interview Items and Overall Score, by Study Site

Did the Doctor or Nurse . . .	Intervention Site (% Yes)	Comparison Site (% Yes)	*P*
Ask for your ideas about how best to keep your child's teeth clean?	41	30	0.001
Ask for your ideas about how to limit the sugars your child gets in foods and drinks?	41	25	<0.0001
Ask for your ideas about how to get your child to the dentist?	30	22	0.003
Explain what cavities are?	22	9	<0.0001
Explain what causes cavities?	19	13	0.015
Explain how cavities can be prevented?	19	16	0.387
Describe how cavities can affect other aspects of your child's health?	13	7	0.007
Discuss limiting sugary foods and drinks to reduce the chance of cavities?	52	29	<0.0001
Discuss limiting your child's bottle or sippy cup use to meals or snack time only?	33	28	0.086
Discuss stopping bed- or naptime bottles by 1 year of age, except for those containing water?	25	29	0.277
Discuss whether your child is getting enough fluoride?	24	16	0.005
Discuss cleaning teeth every night to prevent cavities?	60	32	<0.0001
Discuss using toothpaste with fluoride?	32	15	<0.0001
Discuss having a separate toothbrush for each child?	17	13	0.098
Discuss helping your child brush his or her teeth up until age 6?	52	18	<0.0001
Discuss monitoring kids' teeth for spots?	24	13	<0.0001
Discuss whether your child has a dentist?	37	28	0.002
Discuss whether your child has been to the dentist yet?	33	28	0.074
Recommend that your child see a dentist now?	23	24	0.934
Mention that they would talk about preventing cavities in your child at the next visit?	13	5	0.0001
Give you written information about cavities and how to prevent them?	10	3	<0.0001
Go over things for you to work on to help prevent your child from getting cavities?	24	10	<0.0001
Overall PEI summary score mean ± SD (Median) range	6.4 ± 5.6 (5.0) 0–21	4.1 ± 4.8 (3.0) 0–22	<0.0001

0.0001), and they were more likely to have a high school education or greater (92.9% vs. 84.3%, $P < 0.0001$).

Risk Factor Profiles

Patients at the comparison site had better baseline diets and more tooth-monitoring than those at the intervention site ($P < 0.0001$ and $P = 0.001$, respectively; Table 2); however, hygiene practices were better among patients at the intervention site ($P < 0.0001$).

Patient Exit Interview

Clinicians at the intervention site were more likely to ask the parents for ideas about how best to keep his/her child's teeth clean, limit dietary sugars, and how to get the child to the dentist (Table 1). They were also significantly more likely to explain what cavities are, their causes and their effects on other aspects of the child's health. More clinicians from the intervention site discussed limiting sugary foods and drinks, the child's fluoride intake, and cleaning teeth nightly. Intervention providers more frequently discussed using toothpaste with fluoride, helping the child brush his or her teeth up until age 6, and monitoring the child's teeth for spots, or whether the child had a dentist. Also, a greater percentage of intervention providers mentioned that they would discuss caries prevention at the next visit, gave parents written information about cavities, and discussed caries prevention strategies. Intervention providers asked or counseled an average of 2 more issues than comparison site providers (means: 6.4 vs. 4.1; medians: 5 vs. 3; $P < 0.0001$).

Factors Associated With Provider Counseling

Random effects least squares regression analyses were performed to examine the independent effects of site on provider counseling (PEI), after adjusting for sociodemographic factors and before and after controlling for ECC risk factors (Table 3). In the first model, the PEI score was significantly higher (better) at the intervention site, indicating that intervention providers counseled on approximately 2 more issues than comparison site providers. After adjusting for ECC risk factors, however, the PEI score was not significantly different between the 2 sites.

Factors Associated With ECC

A person-level multiple proportional hazards frailty regression model with physician included as a random effect was fit for ECC incidence for all children free of ECC (cavitated lesions) at baseline, adjusting for demographic variables that varied significantly between sites without a large number of missing observations (Table 4). At baseline, there was no difference in ECC prevalence at the 2 sites (5.8% at the intervention site vs. 6.4%; $P = 0.664$). After adjustment for age and race/ethnicity (2 known influences on ECC), children at both sites were similarly likely to have ECC (OR = 1.13, 95% CI: 0.64–2.00), $P = 0.672$). At the

TABLE 3. Multiple Linear Regression Results: Site Differences in Rates of Provider Counseling*

	Model 1† PEI Score§		Model 2‡ PEI Score§	
	Parameter Estimate	Std Error	Parameter Estimate	Std Error
Intervention Site (ref = comparison site)	1.24¶	0.53	1.02	0.55
Age group (ref = 3 yr or older)				
<2 yr old	−1.70¶	0.46	−1.40¶	0.54
2 to <3 yr old	−0.87‖	0.53	−0.69	0.54
Non-Hispanic white (ref = Hispanic or Non-white)	−1.18¶	0.49	−1.02¶	0.49
Caregiver employment (ref = not employed)	0.58	0.36	0.60	0.36
Caregiver education (ref = <12 yr)	−0.41	0.52	−0.40	0.52
English spoken at home (ref = other)	−0.28	0.38	−0.38	0.38
Diet score§	—	—	−0.30	0.18
Hygiene score§	—	—	0.39¶	0.18
Tooth monitoring score§	—	—	0.36	0.23

*Scores from Patient Exit Interview (PEI) reflect the number of counseling behaviors performed by physicians, based on exit interviews with parents and caregivers. For instance, in Model 1, physicians at the intervention site performed, on average, 1.24 more counseling behaviors than physicians at the control site.

†Controlling for sociodemographic factors and accounting for clustering of patients-within- provider.

‡Controlling for sociodemographic factors as well as diet, hygiene, and tooth monitoring scores and accounting for clustering of patients-within- provider.

§Higher scores are better on all dimensions (eg more counseling, better dietary habits, more tooth monitoring).

¶$P < = 0.05$.

‖$P > 0.05$ and $P < = 0.10$.

TABLE 4. Multiple Proportional Hazards Regression Model for Development of Early Childhood Caries (Cavitated Lesions) in Any Primary Tooth (n = 375 Subjects)

Variable	HR*	95% CI†	P
Site			
Comparison site	1.00	—	0.004
Intervention site	0.23	(0.09, 0.62)	
Age group			
<2 yr old	1.00	—	0.001
2 to <3 yr old	3.23	(1.57, 6.57)	
3 yr or older	3.18	(1.52, 6.68)	
Race/ethnicity			
Non-Hispanic white	1.00	—	0.117
Hispanic or Non-white	2.56	(0.79, 8.30)	
Caregiver employment			
Yes	1.00	—	0.465
No	0.81	(0.45, 1.43)	
Caregiver education			
Less than high school	1.00	—	0.115
High school or greater	0.57	(0.29, 1.15)	
Language spoken at home			
English	1.00	—	0.435
Other	1.27	(0.70, 2.32)	
Diet score	1.09	(0.82, 1.45)	0.551
Hygiene score	1.08	(0.80, 1.46)	0.599
Tooth monitoring score	1.40	(1.00, 1.98)	0.049

*HR indicates hazard ratio; adjusted for all other factors listed in the table.

†CI indicates confidence interval.

last follow-up visit, unadjusted ECC prevalence at the intervention site was 17.7%, compared with 31.7% at the comparison site ($P = 0.086$). Children at the intervention site were 77% less likely to develop ECC over time compared with children at the comparison site (HR = 0.23, 95% CI: 0.09–0.62, $P = 0.004$).

DISCUSSION

Recognition of the profound deleterious effects of ECC on children's health and well being has led to a growing interest in and commitment to the role of primary care clinicians in children's oral health.[19] To address the paucity of interventions oriented towards enhancing providers' knowledge and skills in counseling parents to reduce risk for ECC,[28] we developed, implemented, and evaluated the effects of a multifaceted practice-based intervention for pediatric providers. Providers at the intervention site provided significantly more counseling regarding reducing ECC risk than comparison site providers, with the differences in rates spanning between 7 and 34 percentage points, depending on the item. Further, more provider counseling for ECC at the intervention site was associated with 77% lower incidence of ECC over time. While the rates of ECC increased over time at both sites, the progression was markedly attenuated at the intervention site, suggesting that such multifaceted interventions in the pediatric setting might have significant public health impact.

While intervention providers provided more counseling, their rates of counseling were often only around 20% to 40% (the maximum rate of counseling observed, on only 1 item, was 60%, Table 1). Thus, although our intervention was associated with more counseling, there is still room for improvement, so further training during medical school or residency might help foster such adoption. However, competing demands for time during clinic visits presents a challenge for increasing such rates.

Providers of patients who are perceived to be at greater risk might exert greater effort at counseling to prevent ECC, as we have found in other work on hypertension care.[36] Intervention providers conducted significantly more counseling than at the comparison site after controlling for patients' sociodemographic factors, but not after adjusting for ECC risk factors. However, such adjustment conservatively assumes that providers conducted an accurate risk assessment, which may not be the case.

Interest by pediatricians regarding children's oral health has led to a variety of efforts, many focusing on increasing pediatricians' knowledge base regarding the etiology of dental caries.[17,24,26] Our study is one of the first attempts to increase providers' skills in translating knowledge about the risk factors for ECC into useable information which

can be conveyed to parents to change behavior to reduce ECC risk. Relatively low cost interventions (brief provider training, educational materials, and changing an existing EMR to include cues to counsel patients) were associated with significantly greater amounts of provider counseling, which suggests that it may be valuable to implement such interventions on a more widespread basis. While the limited length of clinic visits may be a barrier to counseling, our results suggest that providers were able to incorporate ECC counseling into their visits, although we were not able to assess the added time needed to do so. The tools we developed for our intervention are easily accessible and transportable for use by other providers (Available at http://www.creedd.org/affiliate.html) including family physicians, midlevel providers, and for pediatric residency training and continuing medical education.

These results are consistent with prior studies of patient centered counseling, which have consistently shown that providers are willing to incorporate such methods into their practices, and that they successfully do so after training.[29,31–34] Our findings indicate that this method of changing provider behavior is useful in the pediatric setting as well.

This study was limited by the absence of baseline data on providers' counseling habits, and by the fact that the samples were different in several respects. However, the findings from the training pretest/post-test indicated an increase in provider knowledge (albeit with a short recall period, and without comparative data from clinicians at the control site). Another potential limitation of the study was that we were unable to disentangle the specific effects of each element of the intervention, although we viewed them as a package which should be implemented together in the future because each piece addresses a different aspect of ECC prevention.

Further, we still observed differential rates of counseling after adjusting for the sociodemographic differences between sites. Also, while our focus on very young children may be viewed as a limitation in that our findings are not generalizable to older children, the dearth of information about ECC and its risk factors among children in this age group warrants such a focus.

Ideally, a cluster-randomized controlled design (randomized by site) would be used to test the effectiveness in training pediatric clinicians to conduct oral health counseling with parents/caregivers of very young children. To obtain a true random sample, one would need to recruit clinicians from the private sector to participate, which might introduce selection bias since such clinicians are less likely to see the economic and racial diversity of children whom we studied. Although our quasi-experimental design offered benefits for investigating the effect of training pediatric clinicians to deliver such counseling, these findings should be replicated elsewhere before widespread use of this intervention.

In summary, a relatively brief intervention was associated with increased provider counseling and reduced subsequent ECC. Such interventions are feasible to implement on a more widespread basis, and if validated by additional studies, may have a significant public health impact, by reducing rates of ECC in young children.

ACKNOWLEDGMENTS

The authors thank the technical assistance of Kristal Raymond, MPH, in accomplishing this work who was supported by a NIH/NIDCR Research Supplement to Promote Diversity in Health-Related Research (U54 DE14264-S2), as well as the support of Janis Johnson, MPH. We also appreciate the thoughtful comments and many helpful discussions with Drs. Raul Garcia, DMD, MMSc, and Ruth Nowjack-Raymer, PhD, in conducting this research. We also appreciate the helpful comments from Drs. Susan Reisine, PhD, and the late Larry Meskin, DDS, PhD, on earlier drafts of this manuscript. Nancy Kressin, PhD, 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.

REFERENCES

1. Acs G, Lodolini G, Kaminsky S, et al. Effect of nursing caries on body weight in a pediatric population. *Pediatr Dent*. 1992;14:302–305.
2. Gift HC, Reisine ST, Larach DC. The social impact of dental problems and visits. *Am J Public Health*. 1992;82:1663–1668.
3. Edelstein B. Children's experience of dental pain: a public policy and clinical policy issue. Paper presented at: Background paper prepared for the Mayday Fund and the Milbank Memorial Fund meeting on Preventing and Treating Children's Oral Pain. Washington, DC; 1999.
4. Office of Disease Prevention and Health Promotion, Office of Public Health and Science, Office of the Secretary, et al. Healthy people 2010. 2006. Available at http://www.healthypeople.gov/document/html/objectives/21-01.htm. Accessed September 12, 2009.
5. Weinstein P. Public health issues in early childhood caries. *Community Dent Oral Epidemiol*. 1998;26:84–90.
6. Milnes AR. Description and epidemiology of nursing caries. *J Public Health Dent*. 1996;56:38–50.
7. Tang JM, Altman DS, Robertson DC, et al. Dental Caries Prevalence and treatment levels in Arizona preschool children. *Public Health Rep*. 1997;112:319–331.
8. Barnes G, Parker W, Lyon T Jr, et al. Ethnicity, location, age, and fluoridation factors in baby bottle tooth decay and caries prevalence of Head Start children. *Public Health Rep*. 1992;107:167–173.
9. Selwitz R, Ismail A, Pitts N. Dental caries. *Lancet*. 2007;369:51–59.
10. Harris R, Nicoll A, Adair P, et al. Risk factors for dental caries in young children: a systematic review of the literature. *Community Dent Health*. 2004;21:71–85.
11. Office of the Surgeon General. 1995. Available at: www.cdc.gov/nccdphp/oh/fl-surgeon.htm. Accessed April 6, 2001.
12. Flores G, Bauchner H, Feinstein AR, et al. The impact of ethnicity, family income, and parental education on children's health and use of health services. *Am J Public Health*. 1999;89:1066–1071.
13. Casmassimo P. Oral health in maternal and child health. In: Wallace HM, Green G, Jaros KJ, et al, eds. *Health and Welfare for Families in the 21st Century*. Sudbury, MA: Jones and Bartlett; 1999:441–453.
14. National Center for Health Statistics. Health, United States, 2006: with chartbook on trends in the health of Americans. 2006. Available at: http://www.cdc.gov/nchs/data/hus/hus06.pdf#079. Accessed September 27, 2007.
15. American Academy of Pediatrics. Oral health risk assessment timing and establishment of the dental home. *Pediatrics*. 2003;111:1113–1116.
16. Sanchez OM, Childers NK, Fox L, et al. Physicians' views on pediatric preventive dental care. *Pediatr Dent*. 1997;19:377–383.
17. Krol D. Educating pediatricians on children's oral health: past, present and future. *Pediatrics*. 2004;113:e487–e492.
18. Lewis C, Grossman D, Domoto P, et al. The role of the pediatrician in the oral health of children: a national survey. *Pediatrics*. 2000;106:e84.
19. Lewis CW, Grossman DC, Domoto PK, et al. The role of the pediatrician in the oral health of children: a national survey. Pediatric Academic Societies and American Academy of Pediatrics Joint Meeting. Seattle, WA; 2000.
20. Rozier R, Sutton B, Bawden J, et al. Prevention of early childhood caries in North Carolina medical practices: implications for research and practice. *J Dent Educ*. 2003;67:876–885.

21. Schaff-Blass E, Rozier R, Chattopadhyay A, et al. Effectiveness of an educational intervention in oral health for pediatric residents. *Ambul Pediatr*. 2006;6:157–164.
22. Maternal and Child Health Bureau. National Maternal and Child Health Resource Center. Bright futures toolbox. Bright futures in practice: oral health. 2007. Available at: http://www.mchoralhealth.org/Toolbox/professionals.html. Accessed September 27, 2007.
23. American Academy of Pediatrics. PCCPP. Oral health risk assessment training for pediatricians and other child health professionals. 2007. Available at: http://www.aap.org/commpeds/dochs/oralhealth/ohra/Final%20Version.pdf. Accessed September 27, 2007.
24. Mouradian W, Reeves A, Kim S, et al. An oral health curriculum for medical students at the University of Washington. *Acad Med*. 2005;80:434–442.
25. dela Cruz G, Rozier R, Slade G. Dental screening and referral of young children by pediatric primary care providers. *Pediatrics*. 2004;114:e642–e652.
26. Slade G, Rozier R, Zeldin L, et al. Training pediatric health care providers in prevention of dental decay: results from a randomized controlled trial. *BMC Health Serv Res*. 2007;7:176.
27. O'Brien MT, Oxman A, Davis D. Audit and feedback versus alternative strategies: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2000:CD000260.
28. Bader J, Rozier R, Lohr K, et al. Physicians' roles in preventing dental caries in preschool children: a summary of the evidence for the U.S. Preventive Services Task Force. *Am J Prev Med*. 2004;26:315–325.
29. Ockene IS, Hebert JR, Ockene JK. Effect of training and a structured office practice on physician-delivered nutrition counseling: the Worcester-Area Trial for Counseling in Hyperlipidemia (WATCH). *Am J Prev Med*. 1996;12:252–258.
30. Ockene JK, Quirk ME, Goldberg RJ, et al. A residents' training program for the development of smoking intervention skills. *Arch Intern Med*. 1988;148:1039–1045.
31. Ockene IS, Hebert JR, Ockene JK, et al. Effect of physician-delivered nutrition counseling training and an office-support program on saturated fat intake, weight, and serum lipid measurements in a hyperlipidemic population. *Arch Intern Med*. 1999;159:725–731.
32. Adams A, Ockene JK, Wheeler EV, et al. Alcohol counseling: physicians will do it. *J Gen Intern Med*. 1998;13:692–698.
33. Ockene J, Kristeller J, Kalan K, et al. *Physician Delivered Smoking Intervention Program*. Worcester, MA: University of Massachusetts; 1990.
34. Ockene JK, Kristeller J, Goldberg R, et al. Increasing the efficacy of physician-delivered smoking interventions: a randomized clinical trial. *J Gen Intern Med*. 1991;6:1–8.
35. Pbert L, Adams A, Quirk M, et al. The patient exit interview as an assessment of physician-delivered smoking intervention: a validation study. *Health Psychol*. 1999;18:183–188.
36. Kressin N, Wang F, Long J, et al. Hypertensive patients' race, health beliefs, process of care, and medication adherence. *J Gen Intern Med*. 2007;22:768–774.

Journal of Dual Diagnosis, 5:191–196, 2009

ISSN: 1550-4263 print / 1550-4271 online
DOI: 10.1080/15504260902869121

Routledge
Taylor & Francis Group

Update in Smoking and Mental Illness: A Primary Care Perspective

KAREN E. LASSER, MD, MPH

Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, Massachusetts; and Department of Medicine, Cambridge Health Alliance and Harvard Medical School, Cambridge, Massachusetts, USA

This article will review population-based nationally representative data on rates of smoking and tobacco cessation in adults with and without mental illness. We begin with a review of the methods and findings from the 1991–1992 National Comorbidity Survey. This study found that 41% of persons who had a mental illness in the past month were current smokers, that persons with mental illness are twice as likely to smoke as persons without mental illness, and that heavy smoking is rare in persons without mental illness. Persons with a current mental illness smoked 44% of all cigarettes in the United States. We then explore the reasons that persons with mental illness smoke at such high rates and examine the directions of causality between smoking and mental illness. We review tobacco companies' marketing activities that have targeted mentally ill smokers. The health consequences of smoking in this vulnerable group are dire. An estimated 200,000 smokers with mental illness or addiction die each year from smoking. Despite their high rates of smoking, a substantial proportion of persons with psychiatric disorders are able to quit.

KEYWORDS *Smoking, mental disorders, primary care*

EPIDEMIOLOGY OF SMOKING AND MENTAL ILLNESS

Many seasoned clinicians know that persons with mental illness smoke at higher rates than persons without mental illness. When I was a primary

The author would like to acknowledge Maxim D. Shrayer, PhD, for his constructive comments on earlier drafts of this paper.

Address correspondence to Karen E. Lasser, MD, MPH, General Internal Medicine, 801 Massachusetts Ave., 2nd Floor, Crosstown Building, Boston, MA 02118. E-mail: karen.lasser@bmc.org

care resident, one of my supervising attendings advised me that one can "practically diagnose schizophrenia from the nicotine stains on a patient's fingers." I remember wondering what this clinical observation meant on a public health level. How heavily did people with mental illness smoke? And, furthermore, how representative were they of the average American smoker (if such a person exists)?

Several years later, when I was a general internal medicine research fellow, I returned to these questions. I reviewed the literature, where I found that a number of studies documented high smoking rates among selected populations with mental illness. These included psychiatric outpatients (Hughes, Hatsukami, Mitchell, & Dahlgren, 1986), patients in state mental hospitals (de Leon et al., 1995), and patients with individual diagnoses, such as bipolar disorder (Gonzalez-Pinto et al., 1998), depression (Glassman et al., 1990), schizophrenia (Kelly & McCreadie, 1999; Goff, Henderson, & Amico, 1992), and panic disorder (Breslau, 1995; Breslau & Klein, 1999). However, none of these studies seemed to reflect a broader primary care population, and few took an even broader public health perspective.

Only the analysis by Glassman et al. (1990) of data from the Epidemiologic Catchment Area Study was population-based. This analysis of data from the early 1980s showed that persons with major depression, dysthymia, agoraphobia, and alcoholism were 1.6 to 4.7 times more likely to have ever smoked than persons without mental illness. Yet, this study was not conclusive, for it did not administer a structured psychiatric interview to a nationally representative sample. It was not until the 1990–1991 National Comorbidity Survey (NCS) that such data were collected (Lasser et al., 2000).

The NCS was a congressionally mandated, nationally representative study of the prevalence of psychiatric disorders in the United States from 1990 to 1992. Trained lay interviewers administered a structured psychiatric interview to the non-institutionalized civilian population aged 15 to 54 years. While the survey excluded persons dwelling in psychiatric facilities and prisons, the study design accounted for such exclusions. The interviewers asked 4,411 respondents detailed questions about tobacco use, thus providing the data to explore further the relationship between smoking and mental illness.

In analyses of these data, my coauthors and I found that persons with mental illness were about twice as likely to smoke as persons without mental illness. Twenty-two percent of people without mental illness smoked as compared to 41% of people with a current mental illness. Heavy smoking, defined as the consumption of 25 or more cigarettes per day, was rare in persons without mental illness. Only 10% of people without mental illness were heavy smokers. Persons with a past-month mental illness represented 40.6% of all current smokers in the United States and smoked 44% of all cigarettes in the United States.

WHY DO THE MENTALLY ILL SMOKE MORE?

Scholars have proposed numerous theories to explain the high smoking rates observed among persons with mental illness. First, there is the theory of self-medication of psychiatric symptoms. Smoking reduces the negative symptoms and cognitive deficits in patients with schizophrenia. In patients with depression, smoking affects noradrenergic proteins in the locus ceruleus much in the way that antidepressants do (Klimek et al., 2001). These theories implicitly assume that mental illness precedes the onset of smoking. However, a number of studies have shown that smoking preceded the onset of certain mental disorders, such as adolescent depression (Wu & Anthony, 1999), first-time panic attacks (Breslau & Klein, 1999; Isensee, Wittchen, Stein, Hofler, & Lieb, 2003), anxiety disorders (Johnson et al., 2000), and schizophrenia (Kelly, 1999). It is also possible that smokers with mental illness have a genetic predisposition to both conditions. Analyses of cross-sectional data cannot determine causality; thus, we are not likely to know for sure whether mental illness causes smoking or whether smoking causes mental illness.

Internal documents from the tobacco companies suggest that the tobacco industry may be partially responsible for high smoking rates among persons with mental illness (Lasser et al., 2000). In 1981, one company conducted a study of different segments of the tobacco market. They identified psychologically vulnerable persons as a key market segment and may have targeted their marketing and advertising accordingly. In their study, the company identified the 3 following positive aspects of smoking: "mood enhancement," "positive stimulation," and "anxiety relief." They described how smoking "helps perk you up," "helps you think out problems," and for anxious individuals, helps them "gain self control," "calm down," and "cope with stress."

SEQUELAE OF SMOKING IN THE MENTALLY ILL

Smoking is the leading preventable cause of death in the United States, with 440,000 deaths annually. Twenty times as many people suffer from smoking-related disability. Extrapolating from the smoking rates found in the NCS, 200,000 smokers with mental illness or addiction die each year from smoking (Williams & Ziedonis, 2004). Yet, there is a common conception that people with mental illness often die of violent causes—suicide or homicide. Mortality follow-up data from the New Haven node of the Epidemiologic Catchment Area Study refute this contention (Bruce, Leaf, Rozal, Florio, & Hoff, 1994). Among persons with mental illness, the investigators found that the leading causes of death were circulatory diseases (55%), cancer-related (23%), other natural causes (21%), and 1% unnatural causes (accident, homicide, and suicide). Smoking could certainly be implicated as

a cause of many of the circulatory and cancer-related deaths. More recent data from Massachusetts also show a high prevalence of smoking-related mortality among persons with mental illness. A 2001 mortality report by Sudders (Commonwealth of Massachusetts Executive Office of Health and Human Services Department of Mental Health, unpublished data, May 2001) found that age-specific mortality from cardiovascular disease for individuals aged 25 to 44 was more than 6 times as high among Department of Mental Health clients than among the general Massachusetts population. Smoking is only one of many other potential causes of high cardiovascular mortality rates—obesity, sedentary lifestyle, medications, social deprivation, and poor quality of medical care—and further research is needed to elucidate the role of smoking in these deaths.

ARE SMOKERS WITH MENTAL ILLNESS ABLE TO QUIT?

Fortunately, smokers with mental illness *are* able to quit. In our analyses of the NCS, quit rates of persons with a lifetime history of mental illness (37.1%) or a mental illness in the past month (30.5%), while lower that those of people without a history of mental illness (42.5%), were substantial. El-Guebaly, Cathcart, Currie, Brown, and Gloster (2002), in their review of 24 studies of smoking cessation in persons with mental illness or addictive disorders, found that quit rates of patients with psychiatric disorders were similar to those in the general population. Yet, other studies have demonstrated low rates of cessation in smokers with schizophrenia, anxiety or depression (past or present), and current alcohol use. The latter studies focused on selected populations of mentally ill smokers, psychiatric outpatients, and more severely ill patients, which may explain their lower cessation rates.

VARENICLINE (CHANTIX): THE LATEST WONDER DRUG?

At our weekly research meeting at Cambridge Hospital, a psychiatrist colleague and I were exchanging stories about adverse reactions our mentally ill smoking patients had had from the newest smoking cessation medication, varenicline. Pharmaceutical companies frequently test medications in healthy populations who are often not representative of patients who ultimately take the drug once it is introduced onto the market. We wondered whether the same was true with varenicline. A quick literature search revealed that, indeed, persons with mental illness were systematically excluded from the premarketing trials of varenicline. The trials excluded patients with the following diagnoses: major depression within the prior year; history of or current panic disorder, psychosis, or bipolar disorder; and drug or alcohol abuse or dependence within the past year (Jorenby et al., 2006). Analyses of psychiatric epidemiologic studies suggest that such persons represent 27%

of the general population (Kessler et al., 1994) and surely an even higher proportion of all smokers.

Despite this major limitation in the premarketing trials of varenicline, the drug has been heavily marketed to and widely consumed by all smokers, with and without mental illness. Since its approval in May 2006, at least 5 million people have taken the drug, and its sales in 2007 totaled $681 million, which represents 90% of the smoking cessation market. Varenicline has also acquired a series of warnings regarding the risk of serious neuropsychiatric symptoms, including agitation, depression, suicidal behavior, and suicidal ideation (U.S. Food and Drug Administration, 2007). My coauthor and I wrote in a recent correspondence to *The Lancet*, "we question the wisdom of excluding these patients [the mentally ill], given the fact that those suffering from mental illness smoke at very high levels and also might be especially vulnerable to post-approval pharmaceutical marketing efforts (just as they have been shown to be vulnerable to tobacco industry marketing)" (Lasser & Boyd, 2008).

CONCLUSIONS

Those who continue to smoke most frequently and most heavily in the United States are the mentally ill. What are some interventions that might impact smoking rates among the mentally ill? First, we (both primary care and mental health providers) need to document smoking status in the patient's medical record. If we cannot identify our smokers, how will we be able to assist them in quitting? Second, we need to support health policy initiatives such as smoke-free psychiatric facilities, clean indoor air laws, and tobacco cessation. Finally, smokers with mental illness must be included in future smoking cessation drug trials. We cannot continue to ignore those in our society who are increasingly marginalized by both their smoking and their mental illness.

REFERENCES

Breslau, N. (1995). Psychiatric comorbidity of smoking and nicotine dependence. *Behavior Genetics*, *25*(2), 95–101.

Breslau, N., & Klein, D. F. (1999). Smoking and panic attacks. *Archives of General Psychiatry*, *56*, 1141–1147.

Bruce, M. L., Leaf, P. J., Rozal, G. P. M., Florio, L., & Hoff, R. A. (1994). Psychiatric status and 9-year mortality data in the New Haven Epidemiologic Catchment Area Study. *American Journal of Psychiatry*, *151*, 716–721.

de Leon, J., Dadvand, M., Canuso, C., White, A. O., Stanilla, J. K., & Simpson, G. M. (1995). Schizophrenia and smoking: An epidemiological survey in a state hospital. *American Journal of Psychiatry*, *152*, 453–455.

El-Guebaly, N., Cathcart, J., Currie, S., Brown, D., & Gloster, S. (2002). Smoking cessation approaches for persons with mental illness or addictive disorders. *Psychiatric Services*, *53*, 1166–1170.

Glassman, A. H., Helzer, J. E., Covey, L. S., Cottler, L. B., Stetner, F., Tipp, J. E., et al. (1990). Smoking, smoking cessation, and major depression. *Journal of the American Medical Association*, *264*, 1546–1549.

Goff, D. C., Henderson, D. C., & Amico, E. (1992). Cigarette smoking in schizophrenia: Relationship to psychopathology and medication side effects. *American Journal of Psychiatry*, *149*, 1189–1194.

Gonzalez-Pinto, A., Gutierrez, M., Ezcurra, J., Aizpuru, F., Mosquera, F., Lopez, P., et al. (1998). Tobacco smoking and bipolar disorder. *Journal of Clinical Psychiatry*, *59*, 225–228.

Hughes, J. R., Hatsukami, D. K., Mitchell, J. E., & Dahlgren, L. A. (1986). Prevalence of smoking among psychiatric outpatients. *American Journal of Psychiatry*, *143*, 993–997.

Isensee, B., Wittchen, H. U., Stein, M. B., Hofler, M., & Lieb, R. (2003). Smoking increases the risk of panic: findings from a prospective community study. *Archives of General Psychiatry*, *60*, 692–700.

Johnson, J. G., Cohen, P., Pine, D. S., Klein, D. F., Kasen, S., & Brook, J. S. (2000). Association between cigarette smoking and anxiety disorders during adolescence and early adulthood. *Journal of the American Medical Association*, *284*, 2348–2351.

Jorenby, D. E., Hays, J. T., Rigotti, N. A., Azoulay, S., Watsky, E. J., Williams, K. E., et al. (2006). Efficacy of varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs placebo or sustained-release bupropion for smoking cessation: A randomized controlled trial. *Journal of the American Medical Association*, *296*, 56–63.

Kelly, C., & McCreadie, R. G. (1999). Smoking habits, current symptoms, and premorbid characteristics of schizophrenic patients in Nithsdale, Scotland. *American Journal of Psychiatry*, *156*, 1751–1757.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry*, *51*, 8–19.

Klimek, V., Zhu, M. Y., Dilley, G., Konick, L., Overholser, J. C., Meltzer, H. Y., et al. (2001). Effects of long-term cigarette smoking on the human locus coeruleus. *Archives of General Psychiatry*, *58*, 821–827.

Lasser, K., & Boyd, J. W. (2008). Varenicline and smokers with mental illness. *Lancet, 372*, 1218–1219.

Lasser, K., Boyd, J. W., Woolhandler, S., Himmelstein, D. U., McCormick, D., & Bor, D. H. (2000). Smoking and mental illness: A population-based prevalence study. *Journal of the American Medical Association*, *284*, 2606–2610.

U.S. Food and Drug Administration. (2007). *2007 safety alerts for drugs, biologics, medical devices, and dietary supplements*. Retrieved July 2, 2008, from http://www.fda.gov/medwatch/safety/2007/safety07.htm#Chantix

Williams, J. M., & Ziedonis, D. (2004). Addressing tobacco among individuals with a mental illness or an addiction. *Addictive Behaviors*, *29*, 1067–1083.

Wu, L. T., & Anthony, J. C. (1999). Tobacco smoking and depressed mood in late childhood and early adolescence. *American Journal of Public Health*, *89*, 1837–1840.

BMC Family Practice

BioMed Central

Research article

A multilevel intervention to promote colorectal cancer screening among community health center patients: results of a pilot study

Karen E Lasser*†1, Jennifer Murillo†2, Elizabeth Medlin†3, Sandra Lisboa†4, Lisa Valley-Shah†5, Robert H Fletcher†6, Karen M Emmons†7 and John Z Ayanian†8

Address: [1]Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA, [2]Department of Medicine, Cambridge Health Alliance and Harvard Medical School, Cambridge, MA, USA, [3]Centers for Disease Control and Prevention, Atlanta, GA, USA, [4]Department of Community Affairs, Cambridge Health Alliance and Harvard Medical School, Cambridge, MA, USA, [5]Department of Gastroenterology, Cambridge Health Alliance and Harvard Medical School, Cambridge, MA, USA, [6]Department of Ambulatory Care and Prevention/Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA, USA, [7]Dana Farber Cancer Institute/Harvard School of Public Health, Boston, MA, USA and [8]Division of General Medicine, Brigham and Women's Hospital and Department of Health Care Policy, Harvard Medical School, Boston, MA, USA

Email: Karen E Lasser* - karen.lasser@bmc.org; Jennifer Murillo - jmurillo@challiance.org; Elizabeth Medlin - bnd0@cdc.gov; Sandra Lisboa - slisboa@challiance.org; Lisa Valley-Shah - lvalley@challiance.org; Robert H Fletcher - robert_fletcher@hms.harvard.edu; Karen M Emmons - Karen_M_Emmons@dfci.harvard.edu; John Z Ayanian - ayanian@hcp.med.harvard.edu

* Corresponding author †Equal contributors

Published: 29 May 2009

BMC Family Practice 2009, **10**:37 doi:10.1186/1471-2296-10-37

Received: 13 November 2008
Accepted: 29 May 2009

This article is available from: http://www.biomedcentral.com/1471-2296/10/37

Abstract

Background: Colorectal cancer screening rates are low among poor and disadvantaged patients. Patient navigation has been shown to increase breast and cervical cancer screening rates, but few studies have looked at the potential of patient navigation to increase colorectal cancer screening rates.

Methods: The objective was to determine the feasibility and effectiveness of a patient navigator-based intervention to increase colorectal cancer screening rates in community health centers. Patients at the intervention health center who had not been screened for colorectal cancer and were designated as "appropriate for outreach" by their primary care providers received a letter from their provider about the need to be screened and a brochure about colorectal cancer screening. Patient navigators then called patients to discuss screening and to assist patients in obtaining screening. Patients at a demographically similar control health center received usual care.

Results: Thirty-one percent of intervention patients were screened at six months, versus nine percent of control patients ($p < .001$).

Conclusion: A patient navigator-based intervention, in combination with a letter from the patient's primary care provider, was associated with an increased rate of colorectal cancer screening at one health center as compared to a demographically similar control health center. Our study adds to an emerging literature supporting the use of patient navigators to increase colorectal cancer screening in diverse populations served by urban health centers.

Background

Colorectal cancer is the second leading cause of cancer death in the United States (US). In 2008, an estimated 148,810 people will be diagnosed with colorectal cancer, and it is estimated that 49,960 will die of the disease. [1] Current guidelines from the U.S. Preventive Services Task Force [2] recommend screening individuals age 50 until age 75 with one of the following tests: flexible sigmoidoscopy every 5 years, colonoscopy every 10 years, or fecal occult blood test (FOBT) every year. Despite the availability of these effective screening tests [3-7] a large proportion of Americans are still not being screened. [8-10] Patients at greatest risk of not being screened include racial and ethnic minorities,[10,11] patients with Medicaid or no health insurance,[8,12,13] those who are foreign born,[12,14] and patients with low socioeconomic status [15] – groups that are commonly served by community health centers.[8,16]

In a prior qualitative study of community health centers at Cambridge Health Alliance that included patients from Brazil, Portugal, the Azores, Cape Verde and Haiti, [17] large immigrant groups in Massachusetts and elsewhere in the US, we found that the following factors prevented patients from being screened for colorectal cancer: 1) lack of trust in doctors; 2) lack of symptoms; 3) lack of a doctor's recommendation for screening and 4) fatalistic views about cancer. Few physicians were aware that lack of trust and fatalistic beliefs about cancer were barriers to screening for their patients. Physicians typically cited comorbid medical conditions and numerous psychosocial stressors as the main reasons why patients did not receive colorectal cancer screening.

We used these findings to inform the development of a patient navigator-based intervention. Patient navigators are people selected from the community who are trained to guide patients through the health care system to receive appropriate services.[18] A type of care management, patient navigation encompasses a wide range of advocacy and coordination activities.[19] Most published research on patient navigators has focused on breast and cervical cancer screening, showing that navigation increases the rate of patient completion of screening and follow-up evaluation.[20,21]

Several studies, all conducted in New York City, have shown that patient navigation can increase rates of colorectal cancer screening among urban minority patients.[18,21-24] Our study adds to the existing literature by including Haitian Creole and Portuguese-speaking patients, and patients in a geographic area other than New York. We report the results of a pilot study to assess the feasibility of using patient navigators to increase rates of colorectal cancer screening among community health center patients in Massachusetts.

Methods

Study setting and sample

Cambridge Health Alliance (CHA) is a Primary Care Practice-Based Research Network (PBRN)[25] including 15 community health centers. The health centers predominantly serve a multi-cultural, low-income population in Cambridge, Somerville, and Everett, MA. We selected one health center to pilot-test the intervention, and a demographically similar health center to serve as the control health center. The CHA institutional review board approved the study protocol. The institutional review board provided a waiver of informed consent, since the study was promoting an established screening standard and primary care providers (PCPs) were able to identify patients who were not appropriate to contact.

Using an electronic clinical data system (Meditech), we identified patients aged 52–80 who appeared to be unscreened for colorectal cancer. We included patients age 75–80 because at the time of the study, age 80 was considered to be the upper age limit of screening by the U.S. Preventive Services Task Force. We chose to begin at age 52 instead of age 50 (the age at which guidelines suggest that screening begin), because we sought consistency with the Healthcare Effectiveness Data and Information Set (HEDIS) measure on colorectal cancer screening. [26,27] The unscreened patient report used in our study also served as the basis for our ambulatory quality improvement colorectal cancer screening measure. We based eligibility for colorectal cancer screening on a modified version of the most recent HEDIS measure. US health plans utilize HEDIS measures to assess performance on important dimensions of care, including cancer screening. We modified the denominator of the measure to include any patient aged 52–80 who had one visit to a primary care physician in a community health center in each of the two previous years. The numerator included any patient who received colonoscopy in the past 10 years, sigmoidoscopy or barium enema in the past five years, or fecal occult blood testing (FOBT) during the prior year. Using this definition, 47% of eligible patients in our network of community health centers received colorectal cancer screening in the year 2006. Since the data report did not capture tests performed outside of Cambridge Health Alliance, or FOBT cards that were not billed, we suspect that the true screening rate was higher than 47%.

We limited our intervention group to patients who spoke English, Portuguese, Spanish or Haitian Creole and who received care at one center in Somerville, MA. We excluded patients of two primary care providers (PCPs) at the intervention center: one PCP who was a study investi-

gator (KEL), and one PCP who was leaving the health center at the time of the study. The control group consisted of a random sample of similarly defined patients (speaking the same languages and unscreened for colorectal cancer based on the abovementioned definition) at another health center in Somerville.

Study Procedures

Because the electronic data system did not capture diagnostic tests performed outside of the health center network, one investigator (KEL) reviewed the medical records of all patients at both the intervention and control health centers who appeared unscreened in the data report to confirm that they were, in fact, unscreened. After reviewing 196 medical records at the intervention center and 191 medical records at the intervention center, we identified 93 intervention patients and 90 control patients who had not received colorectal cancer screening according to the criteria specified above.

We asked each of the eight PCPs at the intervention center to review their list of unscreened patients and to identify any patient who they deemed inappropriate for telephone outreach, based on the following criteria: 1) patient has a medical contraindication to screening or a short life expectancy so that they do not warrant screening[25] 2) the patient will be out of the country continuously for at least three months during the period of navigation 3) the patient had severe cognitive or mental impairment, and no one who can be identified as a caretaker or proxy and 4) other reason as designated by the PCP.

Of the 93 unscreened patients, PCPs deemed 38 (41%) to be inappropriate for outreach for the following reasons: patient has a long history of refusing screening (n = 16), patient with medical comorbidity (n = 7), gastrointestinal symptoms or gastrointestinal workup in progress (n = 6), mental illness or substance abuse (n = 5), other reasons (n = 4; patient uninsured, out of the country, or moving). Fourteen of the 38 patients deemed ineligible for outreach were uninsured.

Intervention

The remaining 55 patients were eligible to receive the intervention. We sent letters by first-class mail, signed by each PCP, notifying patients that they were overdue for colorectal cancer screening, and that a patient navigator would be calling them. The mailing also included a colorectal cancer screening brochure designed by the Harvard Center for Cancer Prevention and the Massachusetts Colorectal Cancer Working Group ("Take Control: Get Tested for Colorectal Cancer"). The brochure, written at a sixth-grade reading level, offered patient-oriented information about the reasons for screening, the different screening modalities, and lifestyle changes to lower risk of colorectal cancer. We sent brochures to patients in English, Portuguese, Spanish, or French (for Haitian Creole-speaking patients).

The study patients were also eligible to receive navigation from navigators speaking English and Spanish, Portuguese, and Haitian Creole, respectively. The navigators were based in the hospital's Department of Community Affairs; they did not have a presence at the intervention health center. The navigator who worked with English and Spanish-speaking patients was originally from Nicaragua, had completed college, and had extensive experience doing community health outreach. She was also a trained certified nurse's assistant (CNA). The Portuguese-speaking navigator had been a masters-level clinical psychologist in Brazil, and was an experienced community health worker. The Haitian navigator was also an experienced community health worker, and worked as a medical assistant in a local community health center. All of the navigators were women, and were age 47, 42, and 37, respectively.

The navigators attended a two day training program in October 2007. The training program included lectures and interactive role plays about the following subjects: 1) the principles of motivational interviewing [28] 2) colorectal cancer and how patients can be screened for it; 3) logistics ("how-to," pros, and cons) of FOBT cards and colonoscopy 4) prevention of colorectal cancer (including prevention by removal of adenomas) 5) use of open vs. closed questions, reflective listening, and summarizing; 6) assessment of patient's readiness for screening and 7) approaches for patients who refuse screening (pre-contemplation), are willing to think about it (contemplation), or are ready to act (action).[28] We chose to frame the intervention around a "stages of change" model as other cancer prevention studies have successfully employed this model.[29]

During the study implementation, the project manager (who also attended the training sessions) audited between one and five patient calls by each navigator for adherence to a calling script and for motivational interviewing techniques. The patient navigators and the project manager also met on a weekly basis to discuss challenges arising during the outreach calls and to review the use of motivational interviewing techniques.

Over a three week period in October 2007, the patient navigators made between 8 and 11 attempts to call each patient on different days (weekdays and weekends) and at different times (morning, afternoon, and evening) until they reached a patient. The navigators also left at least two messages for the patient, either on the answering machine or with a family member.

Once the navigator reached a patient, the navigator discussed the need for colorectal cancer screening with the patient, the screening options of colonoscopy vs. FOBT cards, and the advantages and disadvantages of each test. The navigators did not discuss other screening test options, such as flexible sigmoidoscopy and barium enema, since such options were not routinely offered to patients by their PCPs.

If a patient was interested in completing FOBT cards, the navigator reviewed the FOBT instructions with the patient and mailed FOBT cards and illustrated instructions to patients by first-class mail. The navigator also offered to review the FOBT instructions with the patient over the phone as soon as the patient received the FOBT cards. If a patient did not return the FOBT cards within four weeks, the navigator called the patient to provide support and to address barriers to completion.

For patients who were interested in pursuing colonoscopy, the navigators described the test in detail and the project manager contacted the patient's PCP to arrange a colonoscopy referral. Based on the patient's comorbid medical conditions, the PCP either referred the patient directly for colonoscopy or for a routine appointment with a gastroenterologist to discuss colonoscopy. Patients with any of the following conditions were not eligible for direct referral: sleep apnea, obesity (BMI > 30), previous history of anesthesia problems, congestive heart failure, presence of an automatic implanted cardiac defibrillator, renal failure (as defined by the PCP), and warfarin use for any reason. For patients referred directly to colonoscopy, a registered nurse (LV) called the patient, educated him/her about the procedure and the bowel preparation, and mailed instructions for the bowel preparation to the patient. The patient did not require a medical visit prior to the colonoscopy procedure. The gastroenterology office placed reminder calls to all patients one day prior to their procedure. Due to medico-legal concerns, the navigators did not escort patients home after the colonoscopy. In the event that a patient did not have someone to escort them home, the navigators advised them to complete FOBT cards instead.

At the control health center, patients eligible for colorectal cancer screening received usual care. PCPs offered patients screening on an ad-hoc basis during primary care visits. Unlike the PCPs at the intervention center, the PCPs at the control center did not review their lists of unscreened patients. At both health centers, PCPs had some decision support to promote colorectal cancer screening in the Epic electronic medical record. The electronic record includes a health maintenance grid which flags age-appropriate patients who have not received colorectal cancer screening. The PCPs at the control health center could also refer patients directly for colonoscopy at the time of the study, but they did not have access to patient navigators to advise patients on screening options or to assist them in completing the test.

Measures

The primary outcome of the study was completion of colorectal cancer screening at six months. While the intervention focused on the completion of colonoscopy or a set of three FOBT cards from home, patients who completed any of the following during the study period were considered to have been screened: colonoscopy, sigmoidoscopy or barium enema, or FOBT cards. One of the investigators (KEL) conducted a non-blinded chart review to determine completion of colorectal cancer screening tests.

Process Evaluation

During the study, the navigators maintained paper records in which they documented details of their interactions with patients, including the patients' readiness to be screened, barriers to screening, and actions that were taken to promote screening. The project manager entered these data into a Microsoft Access database.

Statistical Methods

We included all patients at the intervention center in an intention-to-treat analysis, regardless of whether they were designated by their PCP to receive navigation, or whether a navigator successfully reached the patient. Using the χ^2 test, we compared screening rates at six months among intervention patients and control patients. We chose to analyze the data at six months because the wait for a screening colonoscopy at the time of the study was on the order of weeks, and we assumed that patients would have had sufficient time to complete their colonoscopy during the six-month period.

Results

Table 1 shows the demographic characteristics of the intervention and control center patients. The patients at both centers were of similar age, race (note that race data were missing for 7 persons), and insurance status. Of those patients who had insurance, the majority at both sites had Medicaid or free care (66% at the intervention center and 51% at the control center). At the time of the study, after being determined ineligible for other payment options, Massachusetts residents were able to apply for help paying for health center bills from the Massachusetts uncompensated (free) care pool. The non-English speaking patients at both sites were mostly Portuguese speaking, with small numbers of Spanish and Haitian Creole speaking patients.

Table 1: Community Health Center Patient Characteristics

Variable	Intervention n = 93	Control N = 90	Chi-square p-value
Female (%)	63.4	75.6	.08
Mean Age (SD)	60.6 (6.6)	60.9 (7.1)	.54
Race (%)			
White	67.0	71.8	.50
Non-White	33.0	28.2	
Insurance coverage (%)			
No coverage	24.7	17.8	.25
Coverage	75.3	82.2	
Language used in visit (%)			
English	51.6	53.3	.82
Non-English	48.4	46.7	

Screening Outcomes

Table 2 shows the main study results. Patients in the intervention center were much more likely to be screened within six months than patients in the control group (31% vs. 9%, χ^2 $p < .001$). Due to small numbers we did not present P values for comparisons between the different types of screening (FOBT and colonoscopy). Three of the 38 patients (8%) whom PCPs at the intervention site deemed ineligible for outreach were screened at six months (Figure 1).

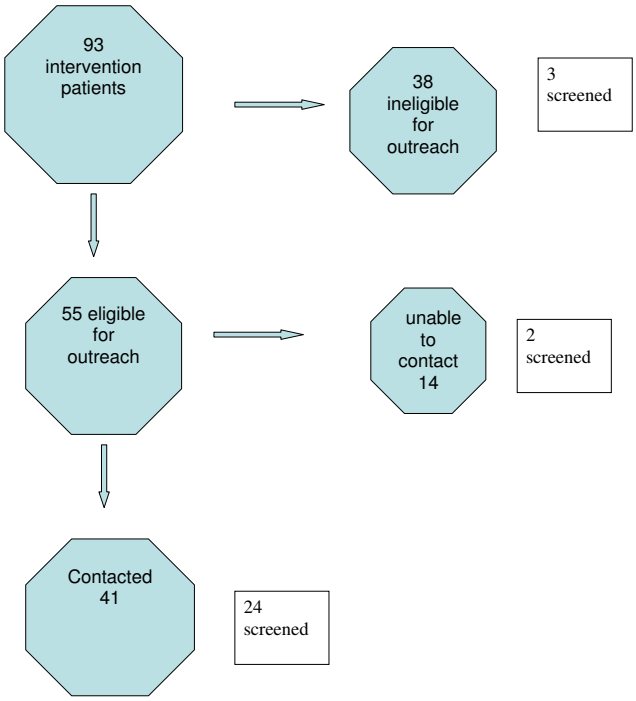


Figure 1
Flow diagram of intervention patients.

Of the 29 patients screened at the intervention site, 16 completed FOBT cards, and 13 completed colonoscopy. Among patients who completed FOBT cards, all tests were negative. Of those patients who completed colonoscopy, three had high-risk lesions; one patient had a tubulovillous adenoma, one patient had four tubular adenomas, and another had a 35 mm. tubular adenoma. A fourth patient had two small tubular adenomas. Among the eight patients screened at the control site, seven completed FOBT cards and one completed colonoscopy. Three patients at the control site had positive FOBT results, but only one of these patients completed a follow-up colonoscopy within six months.

Process Outcomes

Of the 55 patients who were offered navigation, the patient navigators were unable to contact 14 (25%) after between eight and eleven attempted telephone calls. Two of these 14 patients (14%) were screened at six months, while 24 of the 41 patients (59%) whom the navigators were able to contact were screened at six months. For patients reached by the navigator, the median number of contacts was five (range 1–16). Patients received, on average, about four hours of telephone outreach. Patients who received more contact (eight or more calls) were no more likely to be screened than those who received less contact (fewer than eight calls).

In their discussions with patients, the navigators learned that many patients had not been screened because their PCP had not taken enough time to educate them about colorectal cancer screening. For example, one patient stated, "my doctor asked me if I wanted to have it (colon-

Table 2: Colorectal Cancer Screening Results

Variable	Intervention n = 93	Control n = 90	Chi-square p-value
	(%)	(%)	
Screened for colorectal cancer at 6 months	31.2	8.9	.0002
Screened by FOBT	17.2	7.8	*
Positive tests	0	3.3	*
Screened by colonoscopy	14.0	1.1	*
Adenomas	4.3	0	*

oscopy) done, and I said no and that was it." The patient noted that the PCP did not explore her reasons for declining screening. In addition, patients related not being able to take time off from work to undergo colonoscopy.

Discussion

We found that a patient navigator-based intervention was associated with an increased rate of colorectal cancer screening at one health center as compared to a demographically similar control health center. Almost one-third of intervention patients were screened at six months versus nine percent of control patients. Our study adds to an emerging literature supporting the use of patient navigators to increase colorectal cancer screening in diverse populations served by urban health centers.[19,22-24]

While our intervention was effective, it did not achieve the screening rates observed in other studies. For example, studies by Chen et al[18] and Christie et al[24] found that over 50% of navigated patients completed colonoscopy. These studies offered patient navigation only after a patient had been referred for screening colonoscopy by their PCP, which may explain their higher screening rates. In addition, these studies excluded patients who required a gastrointestinal clinic visit for pre-screening evaluation. Jandorf et al[22] also achieved higher screening rates, in both the intervention and control groups. It is possible that the higher screening rates observed in all of these navigation studies could partially be attributed to secular trends. In New York City, where all three of these studies were conducted, 1.25 million people were screened in 2007, up from 826,000 in 2003, with the biggest rates of increase in minority communities.[30]

Our study was limited by the fact that only 41 (44%) of 93 unscreened patients at the intervention health center were actually contacted by a patient navigator. The PCPs at the intervention site identified 38 patients (41%) as inappropriate for outreach. While some of the PCPs reasons for excluding patients were legitimate, such as medical comorbidity, gastrointestinal symptoms or gastrointestinal workup in progress, and mental illness or substance abuse (our navigators were not trained to deal with these special populations), some of the patients who were excluded may have been good candidates for patient navigation services. Such patients included those with a long history of refusing screening and the uninsured. By excluding these patients, we may have underestimated the potential impact of patient navigation. Our study is also limited by small sample size, which precluded us from examining the individual effects of different components of the intervention (letter versus navigation) and from performing exploratory subgroup analyses.

Unlike prior studies of patient navigation, which included mostly Hispanic and African American patients, our study included immigrants from Brazil, Portugal, the Azores, and Haiti. Our inability to contact a substantial proportion (25%) of patients, which decreased the effectiveness of the intervention, may be due to the fact that many patients travel back and forth to their country of origin. The PCPs at the intervention site were often unaware of their patients' migratory patterns, and hence did not exclude such patients from being outreached. These patients also experience housing instability.

Unmeasured differences between the two health centers could account for the differential screening rates. In addition, the PCPs at the control center did not have an opportunity to identify patients whom they deemed inappropriate for screening. We attempted to account for this difference by including all of the intervention center patients in an intent-to-treat analysis. A further potential source of bias is the fact that our qualitative study of barriers to colorectal cancer screening [17] included one PCP from the intervention site, and no PCPs from the control site. We doubt that a one-hour interview conducted with a PCP in 2005 would have significantly affected his colorectal cancer screening practices.

Conclusion

This study supports the feasibility and effectiveness of a patient navigator intervention to increase colorectal cancer screening rates in a community health center serving ethnically and linguistically diverse patients. Future studies will need to examine the cost-effectiveness of such an intervention, and a randomized trial would confirm the effectiveness of patient navigation for immigrant groups who have not been previously studied.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KL, principal investigator, led clinic recruitment, led analysis, and wrote and edited drafts of the manuscript. EM, project manager, oversaw training and supervised the patient navigators. JM and SL navigated the patients, and LV talked to patients about the open access colonoscopy procedure. RF, JA, and KE participated in the initial study design and interpretation of findings. All authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge the patients and PCPs who participated in the study. This study was supported by Mentored Research Scholar Grant MRSGT-05-007-01-CPPB from the American Cancer Society. (Dr. Lasser)

References

1. **Cancer.org** [http://www.cancer.org/docroot/CRI/content/CRI_2_4_1X_What_are_the_key_statistics_for_colon_and_rectum_cancer.asp?sitearea=]
2. **Agency for Healthcare Research and Quality** [http://www.ahrq.gov/clinic/uspstf/uspscolo.htm]
3. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, Ederer F: **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.
4. Newcomb PA, Norfleet RG, Storer BE, Surawicz TS, Marcus PM: **Screening sigmoidoscopy and colorectal cancer mortality.** *J Natl Cancer Inst* 1992, **84(20):**1572-1575.
5. Selby JV, Friedman GD, Quesenberry CP Jr, Weiss NS: **A case-control study of screening sigmoidoscopy and mortality from colorectal cancer.** *N Engl J Med* 1992, **326(10):**653-657.
6. Winawer SJ, St John J, Bond J, Hardcastle JD, Kronborg O, Flehinger B, Schottenfeld D, Blinov NN: **Screening of average-risk individuals for colorectal cancer. WHO Collaborating Centre for the Prevention of Colorectal Cancer.** *Bull World Health Organ* 1990, **68(4):**505-513.
7. Grazzini G, Castiglione G, Ciabattoni C, Franceschini F, Giorgi D, Gozzi S, Mantellini P, Lopane P, Perco M, Rubeca T, *et al.*: **Colorectal cancer screening programme by faecal occult blood test in Tuscany: first round results.** *Eur J Cancer Prev* 2004, **13(1):**19-26.
8. Cokkinides VE, Chao A, Smith RA, Vernon SW, Thun MJ: **Correlates of Underutilization of Colorectal Cancer Screening among U.S. Adults, Age 50 Years and Older.** *Preventive Medicine* 2002, **36:**85-91.
9. Centers for Disease Control and Prevention: **Trends in screening for colorectal cancer-United States, 1997 and 1999.** *Morbidity and Mortality Weekly Report* 2001, **50:**162-166.
10. Seeff LC, Nadel MR, Klabunde CN, Thompson T, Shapiro JA, Vernon SW, Coates RJ: **Patterns and predictors of colorectal cancer test use in the adult U.S. population.** *Cancer* 2004, **100(10):**2093-2103.
11. Walsh JM, Kaplan CP, Nguyen B, Gildengorin G, McPhee SJ, Perez-Stable EJ: **Barriers to colorectal cancer screening in Latino and Vietnamese Americans. Compared with non-Latino white Americans.** *J Gen Intern Med* 2004, **19(2):**156-166.
12. Swan J, Breen N, Coates RJ, Rimer BK, Lee NC: **Progress in cancer screening practices in the United States: results from the 2000 National Health Interview Survey.** *Cancer* 2003, **97(6):**1528-1540.
13. Walsh JME, Posner SF, Perez-Stable EJ: **Colon Cancer Screening in the Ambulatory Setting.** *Preventive Medicine* 2002, **35:**209-218.
14. Goel MS, Wee CC, McCarthy EP, Davis RB, Ngo-Metzger Q, Phillips RS: **Racial and ethnic disparities in cancer screening: the importance of foreign birth as a barrier to care.** *J Gen Intern Med* 2003, **18(12):**1028-1035.
15. Weller D, Coleman D, Robertson R, Butler P, Melia J, Campbell C, Parker R, Patnick J, Moss S: **The UK colorectal cancer screening pilot: results of the second round of screening in England.** *Br J Cancer* 2007, **97(12):**1601-1605.
16. Wee CC, McCarthy EP, Phillips RS: **Factors associated with colon cancer screening: the role of patient factors and physician counseling.** *Prev Med* 2005, **41(1):**23-29.
17. 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.
18. Chen LA, Santos S, Jandorf L, Christie J, Castillo A, Winkel G, Itzkowitz S: **A program to enhance completion of screening colonoscopy among urban minorities.** *Clin Gastroenterol Hepatol* 2008, **6(4):**443-450.
19. Freeman HP: **Patient navigation: a community based strategy to reduce cancer disparities.** *J Urban Health* 2006, **83(2):**139-141.
20. 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):**359-367.
21. Dohan D, Schrag D: **Using navigators to improve care of underserved patients: current practices and approaches.** *Cancer* 2005, **104(4):**848-855.
22. Jandorf L, Gutierrez Y, Lopez J, Christie J, Itzkowitz SH: **Use of a patient navigator to increase colorectal cancer screening in an urban neighborhood health clinic.** *J Urban Health* 2005, **82(2):**216-224.
23. Nash D, Azeez S, Vlahov D, Schori M: **Evaluation of an intervention to increase screening colonoscopy in an urban public hospital setting.** *J Urban Health* 2006, **83(2):**231-243.
24. Christie J, Itzkowitz S, Lihau-Nkanza I, Castillo A, Redd W, Jandorf L: **A randomized controlled trial using patient navigation to increase colonoscopy screening among low-income minorities.** *J Natl Med Assoc* 2008, **100(3):**278-284.
25. **AHRQ Support for Primary Care Practice-Based Research Networks (PBRNs)** [http://www.ahrq.gov/research/pbrn/pbrnfact.htm]
26. **National Quality Measures Clearinghouse** |http://www.qualitymeasures.ahrq.gov/summary/summary.aspx?doc_id=10028||
27. Gross CP, McAvay GJ, Krumholz HM, Paltiel AD, Bhasin D, Tinetti ME: **The effect of age and chronic illness on life expectancy after a diagnosis of colorectal cancer: implications for screening.** *Ann Intern Med* 2006, **145(9):**646-653.
28. Prochaska J, C D: **Toward a comprehensive model of change.** In *Treating Addictive Behaviours: Process of Change* Edited by: Miller W, Healther N. New York: Plenum Pr; 1986:3-27.
29. Dietrich AJ, Tobin JN, Cassells A, Robinson CM, Greene MA, Sox CH, Beach ML, DuHamel KN, Younge RG: **Telephone care management to improve cancer screening among low-income women: a randomized, controlled trial.** *Ann Intern Med* 2006, **144(8):**563-571.
30. **More People Undergoing Colonoscopy in New York** 2008:B6 [http://www.nytimes.com/2008/06/06/nyregion/06colon.html]. New York Times

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2296/10/37/prepub

Age at Natural Menopause and Risk of Ischemic Stroke
The Framingham Heart Study

Lynda D. Lisabeth, PhD; Alexa S. Beiser, PhD; Devin L. Brown, MD; Joanne M. Murabito, MD, ScM; Margaret Kelly-Hayes, EdD, RN; Philip A. Wolf, MD

Background and Purpose—Women have increased lifetime stroke risk and more disabling strokes compared with men. Insights into the association between menopause and stroke could lead to new prevention strategies for women. The objective of this study was to examine the association of age at natural menopause with ischemic stroke risk in the Framingham Heart Study.

Methods—Participants included women who survived stroke-free until age 60, experienced natural menopause, did not use estrogen before menopause, and who had complete data (n=1430). Participants were followed until first ischemic stroke, death, or end of follow-up (2006). Age at natural menopause was self-reported. Cox proportional hazards models were used to examine the association between age at natural menopause (<42, 42 to 54, ≥55) and ischemic stroke risk adjusted for age, systolic blood pressure, atrial fibrillation, diabetes, current smoking, cardiovascular disease and estrogen use.

Results—There were 234 ischemic strokes identified. Average age at menopause was 49 years (SD=4). Women with menopause at ages 42 to 54 (hazard ratio=0.50; 95% CI: 0.29 to 0.89) and at ages ≥55 (hazard ratio=0.31; 95% CI: 0.13 to 0.76) had lower stroke risk compared with those with menopause <42 years adjusted for covariates. Women with menopause before age 42 had twice the stroke risk compared to all other women (hazard ratio=2.03; 95% CI: 1.16 to 3.56).

Conclusion—In this prospective study, age at natural menopause before age 42 was associated with increased ischemic stroke risk. Future stroke studies with measures of endogenous hormones are needed to inform the underlying mechanisms so that novel prevention strategies for midlife women can be considered. **(*Stroke*. 2009;40:1044-1049.)**

Key Words: stroke ■ cerebrovascular disease ■ women ■ menopause ■ bone mineral density

Average life expectancy for women in the United States is 80 years, 5 years longer than that of men.[1] Although men have an increased stroke risk, more women than men will experience a stroke during their lifetime because of their increased life span.[2] Studies consistently show that women are more functionally impaired after stroke and are less likely to receive tissue plasminogen activator compared with men.[3] Given the increased stroke burden and barriers to acute stroke therapy in women, it is critical to understand risk factors unique to women so that new strategies for stroke prevention can be considered.

Results from a meta-analysis demonstrated that menopause before age 50 was associated with a 25% increased risk of cardiovascular disease.[4] Three of the 12 studies in the meta-analysis included stroke,[5–7] with only 1 focused on incident stroke versus stroke mortality.[7] This investigation from the Nurse's Health Study failed to find an association between age at natural menopause and stroke risk; however, a protective effect of older age at menopause and ischemic stroke risk was suggested.[8] With the exception of this study, prospective data on the association of age at natural menopause and stroke risk among US women are lacking.

Beyond age at natural menopause, duration of ovarian activity may be a marker of stroke risk. A recent case-control study found that a longer lifetime estrogen exposure, defined as the difference between age at menopause and age at menarche, was associated with decreased stroke risk.[8] An alternative measure of cumulative endogenous estrogen exposure is bone mineral density (BMD). BMD is associated with age at menarche,[9,10] age at menopause,[11,12] and endogenous estrogen levels among peri- and postmenopausal women.[13,14] Data from one prospective US study of elderly women demonstrated a strong association between low BMD and risk of incident stroke,[15] whereas data from Third National Health and Nutrition Examination Survey I–Epidemiological Follow-up Study (NHANES I) failed to find an association.[16] Further investigation of the relationship between BMD and stroke incidence is warranted.

Received November 14, 2008; final revision received February 4, 2009; accepted February 5, 2009.

From the Department of Epidemiology (L.D.L.), University of Michigan School of Public Health, Ann Arbor, Mich; the Stroke Program (L.D.L., D.L.B.), University of Michigan Medical School, Ann Arbor, Mich; the Department of Neurology (A.S.B., M.K.-H., P.A.W.), Boston University School of Medicine, Boston, Mass; the Department of Biostatistics (A.S.B.), Boston University School of Public Health, Boston, Mass; the Boston University School of Medicine (J.M.M.), Department of Medicine, Section of General Internal Medicine, Boston, Mass; and the National Heart, Lung and Blood Institute's Framingham Heart Study (A.S.B., J.M.M., M.K.-H., P.A.W.), Bethesda, Md.

Correspondence to Lynda Lisabeth, PhD, 109 S. Observatory St., Room 4641, Ann Arbor, MI 48109. E-mail llisabet@umich.edu

Stroke **is available at http://stroke.ahajournals.org** **DOI: 10.1161/STROKEAHA.108.542993**

The primary objective of this study was to prospectively examine the association of age at natural menopause with risk of ischemic stroke in the Framingham Heart Study (FHS). A secondary objective was to examine the association of BMD and risk of ischemic stroke. Analyses were limited to ischemic stroke given the purported role of estrogen deficiency in promoting atherosclerosis.

Methods

FHS is an ongoing prospective cohort study of 5209 participants (2873 women, ages 28 to 62 at the time of enrollment) that began in 1948 in the town of Framingham, Mass. Participants undergo biennial examinations including medical histories, physical examinations, laboratory tests for vascular risk factors, and, at some examinations, brain imaging studies. Details of the study methods have been published.[17,18]

This study investigated the association of age at natural menopause and incident ischemic stroke after age 60. Among the 2873 women in the original cohort, there were 2461 who attended an examination within 3 years of age 60; this examination was designated the participant's baseline. Participants were excluded if they had no information on age at menopause (n=7), surgical menopause or menopause of unknown cause (n=702), prevalent ischemic stroke at entry (n=18), no follow-up after entry (n=11), estrogen use before menopause (n=26), or missing risk factor data (n=267). The remaining 1430 participants comprised the study sample for the analysis of the association of age at natural menopause with risk of ischemic stroke (primary objective).

Of the 2873 women in the original cohort, 866 were alive and attended examination 20 (1986–1990), when BMD was measured; this examination was designated the participant's baseline for the analysis of BMD and risk of ischemic stroke (secondary objective). Participants were excluded from this analysis if they had prevalent ischemic stroke at entry (n=43) or no follow-up after entry (n=7). Of the remaining 816, BMD was measured in 654 women. These women comprised the study sample for the secondary objective. All participants provided written informed consent, and the study was approved by the Boston Medical Center Institutional Review Board.

Baseline Covariates

Baseline covariates were assessed at age 60 (±3 years) for the primary objective and at the time of BMD measurement for the secondary objective. The following covariates were considered: systolic blood pressure, diabetes, atrial fibrillation, cardiovascular disease, current smoking status, body mass index, and estrogen use. Systolic blood pressure was recorded as the average of 2 physician-recorded measurements. Diabetes was defined as a random blood glucose >200 mg/dL, previous diagnosis or treatment with diabetes medication (insulin or oral hypoglycemia agent). Prior cardiovascular disease included coronary heart disease, congestive heart failure and intermittent claudication. Atrial fibrillation was obtained from a standard 12-lead ECG completed at or before the baseline examination. Estrogen use was defined as someone taking estrogen at their baseline assessment. Women taking estrogen before menopause were excluded from the analysis so this covariate measured estrogen started after menopause. Analyses were limited to those with complete covariate data.

Stroke Ascertainment

The primary outcome was incident ischemic stroke. Stroke was defined clinically as a focal neurological deficit of sudden or rapid onset that persisted for more than 24 hours. Continuous surveillance for cerebrovascular events included daily hospital monitoring, tracking of medical encounters, and examination of those with possible stroke symptoms identified at routine biennial examinations. Events were adjudicated by at least 2 neurologists, and with verification of stroke by imaging when available. Stroke occurrence and characteristics, including subtypes, were determined at the end of the acute stroke phase according to uniform criteria and a standardized protocol.[18,19]

Age at Natural Menopause

At each biennial examination, women were queried as to whether periods had stopped for 1 year or more, the age at which periods ceased, the cause of stopped periods (natural, surgical, other), whether a hysterectomy was performed, and number of ovaries removed. Natural menopause occurred if a woman had ceased menstruating naturally for at least 1 year. Age at natural menopause was retrospectively assigned as the self-reported age at last menstrual period.

Bone Mineral Density

BMDs of the femur (neck and trochanter) and distal third of the radius were measured in members of the cohort who came for their 20th biennial examination in 1986 to 1990. Measurements were done using dual-photon absorptiometry for the hip (DP3; Lunar Corp, Madison, Wisc) and single-photon absorptiometry for the distal third of the radius (LUNAR SP2; Lunar Corp).

Statistical Analysis

Baseline characteristics were calculated using frequencies and percents or means and standard deviations (SD). Cox proportional hazards models were used to examine the association between age at natural menopause and risk of ischemic stroke. Individuals were censored at death, hemorrhagic stroke, last examination or contact date, or end of follow-up (December 2006). Survival age was used as the outcome in all models, with entry age used as the left truncation limit. Given an observed nonlinear relationship, age at natural menopause was modeled categorically (<42 [referent]), 42 to 54, ≥55). Models were run age-adjusted and adjusted for age plus baseline covariates (systolic blood pressure, atrial fibrillation, diabetes, current smoking, cardiovascular disease, estrogen use). All covariates were modeled dichotomously with the exception of systolic blood pressure and age which were modeled continuously. A Wald χ^2 test was used to test the overall association between age at natural menopause and risk of stroke in the adjusted model. Models were also run limited to never smokers and never estrogen users given the potential confounding effects of these covariates.

Cox proportional hazards models were used to examine the association between BMD and risk of ischemic stroke, with individuals censored as described above. Survival age was again used as the outcome in all models, with age at the 20th examination used as the left truncation limit. Models were run separately for each BMD site. BMD was modeled categorically based on quintiles of the distribution of BMD at each site with the middle quintile as the referent. BMD quintiles were determined within age groups (67 to 69, 70 to 74, 75 to 79, ≥80). Models were run age-adjusted and adjusted for the covariates described above with additional adjustment for body mass index. Using the adjusted models, Wald χ^2 tests were used to test the overall associations of BMD at each site and risk of stroke. Models were also run limited to those not taking antihypertensives given the potential confounding effects of this covariate.

Results

For the primary objective, there were 1430 women with complete data. Baseline covariate data for these women is included in Table 1. Average age at menopause was 49 years (SD=4). Women were followed for an average of 22 years (SD=9). There were 234 incident ischemic strokes occurring at an average age of 80 years (SD=9). The Figure displays cumulative incidence of ischemic stroke by age and age at natural menopause, and Table 2 displays the model results. In the age-adjusted model, women with menopause at ages 42 to 54 (hazard ratio [HR]=0.57; 95% CI: 0.33 to 1.01) and at ages ≥55 (HR=0.33; 95% CI: 0.14 to 0.79) had lower stroke risk compared with those with menopause <42 years. These associations were relatively unchanged with adjustment for baseline covariates. In the adjusted model, there was a

Table 1. Baseline Characteristics (percent or mean±SD) at Age 60 Among Women in the Framingham Heart Study (n=1430)

Baseline covariate	
Age, years	60.0±0.8
Systolic blood pressure, mm Hg	141±24
History of diabetes	4%
History of CVD	7%
History of AF	1%
Current smoking	32%
Hormone replacement therapy	19%
BMI, kg/m^2	27±5

CVD indicates cardiovascular disease; AF, atrial fibrillation; BMI, body mass index.

significant overall association between age at natural menopause and ischemic stroke risk (P=0.02). Women with menopause before age 42 had twice the risk of ischemic stroke compared to all other women (HR=2.03; 95% CI: 1.16 to 3.56). Limiting to never smokers or to never estrogen users, results were similar (Table 2).

Six hundred fifty-four women had at least one BMD measurement with an average age at measurement of 76 years (SD=5). Women were followed an average of 12 years (SD=5). In this subset, there were 92 ischemic strokes. Table 3 displays the model results. Cut-points for defining quintiles of BMD are provided in supplemental Table I, available online at http://stroke.ahajournals.org. In adjusted models, there were borderline significant associations with BMD at the trochanter (P=0.07) and radius (P=0.07) and ischemic stroke risk. For BMD measured at the trochanter, a U-shaped pattern in risk was observed with women in the lowest (HR=2.36, 95% CI: 1.15 to 4.83) and highest (HR=1.94, 95% CI: 0.94 to 3.95) quintiles of BMD having elevated stroke risk compared with women in the middle quintile. A similar pattern was observed at the radius (Q1–HR=2.92, 95% CI: 1.55 to 6.34; Q5–HR=2.34, 95% CI: 1.12 to 5.30). Limiting the analysis to those not using antihypertensives, results were similar (data not shown).

Discussion

In this prospective study, we observed a significant association between age at natural menopause and ischemic stroke risk in a cohort of women followed from age 60. This association was nonlinear and reflected an increased risk of ischemic stroke in those with natural menopause before 42. Menopause at ≤40 years is termed premature ovarian failure (POF). The etiology of POF is unknown, although POF is thought to arise from different processes than those leading to natural menopause around age 50. Prevalence of POF is 1% to 2% among women, with an additional 3% to 10% of women experiencing "early" menopause defined as natural menopause before age 45.[20,21] Although women with menopause before 42 years represent a small subgroup of the total population, data from this study suggest that 4% to 5% of strokes in all women can be attributed to this risk factor. Reasons for increased ischemic stroke risk among women with POF or early menopause are not clear but early loss of ovarian function coupled with a prolonged low estrogen state is a plausible hypothesis.

The menopausal transition represents a change in endogenous hormones including decreasing estradiol levels several years before menopause and relative estrogen deficiency within 2 to 3 years of the final menstrual period.[22] Estrogen deficiency is thought to promote cardiovascular disease,[23] perhaps through functional or structural changes in the arteries,[24] and as such early onset of estrogen loss in women with POF may contribute to increased stroke risk. However, the role of estrogen deficiency has become controversial in

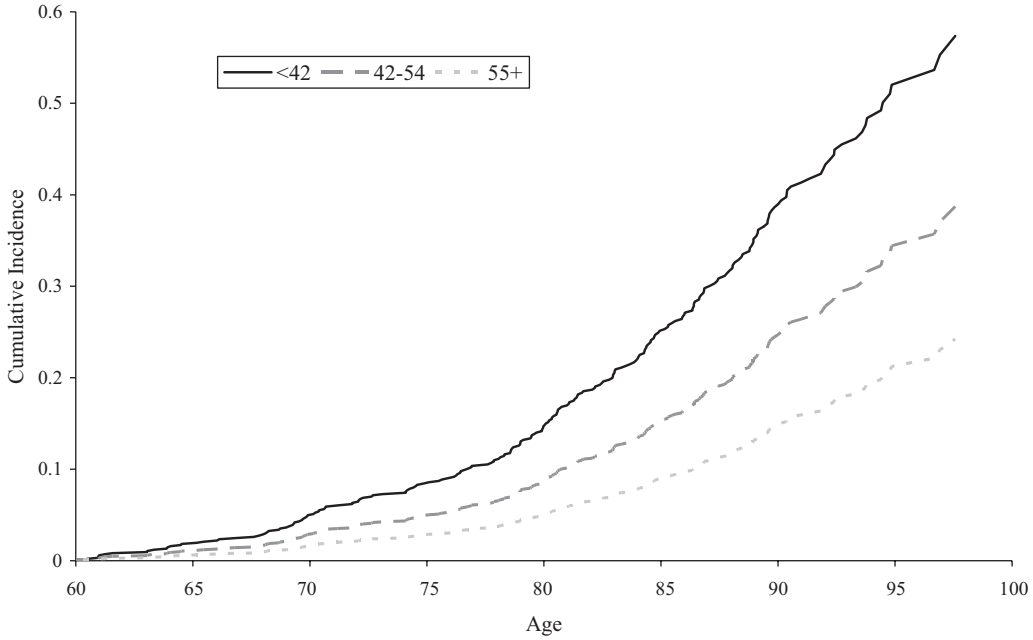


Figure. Cumulative incidence of ischemic stroke by age and age at natural menopause among women in the Framingham Heart Study (n=1430).

Table 2. Associations of Age at Natural Menopause and Risk of Incident Ischemic Stroke Among Women in the Framingham Heart Study (n=1430)

			Age-Adjusted			Multivariable-Adjusted*		
	No. of Participants	No. of Ischemic Strokes	HR	95% CI	*P*	HR	95% CI	*P*
Age at natural menopause								
<42	56	13	1.00			1.00		
42–54	1299	213	0.57	0.33–1.01	0.05	0.50	0.29–0.89	0.02
≥55	75	8	0.33	0.14–0.79	0.01	0.31	0.13–0.76	0.01
Never smokers at age 60								
<42	30	9	1.00			1.00		
42–54	605	96	0.38	0.19–0.76	0.01	0.39	0.20–0.78	0.01
≥55	40	6	0.34	0.12–0.96	0.04	0.40	0.14–1.12	0.08
Never estrogen users at age 60								
<42	51	12	1.00			1.00		
42–54	1158	193	0.54	0.30–0.97	0.04	0.48	0.27–0.87	0.02
≥55	63	7	0.31	0.12–0.79	0.01	0.30	0.12–0.77	0.01

*Adjusted for age, systolic blood pressure, atrial fibrillation, diabetes, current smoking, cardiovascular disease, and estrogen use.

light of the higher stroke risk associated with hormone replacement therapy (HRT) in clinical trials.[25–27] Recent analyses of Women's Health Initiative (WHI) randomized controlled trial data suggest that the timing of HRT initiation may modify the association of HRT and cardiovascular risk, with the effects of HRT being favorable in women initiating therapy in close proximity to menopause. Interestingly, this pattern does not hold for stroke, further complicating an understanding of the hormone-stroke association.[28,29]

No published study has assessed the association between endogenous estrogens and stroke risk. Studies of other nonstroke cardiovascular disease end points in postmenopausal women have found no association between endogenous estrogen and peripheral artery disease,[30] intima media thickness,[31] and cardiovascular disease.[32,33] In contrast, proxy measures of endogenous estrogen exposure, including measures of lifetime ovarian activity and BMD, have been associated with stroke risk in some[15,34] but not all studies.[16] Unlike previous studies, which suggested a

Table 3. Associations of Bone Mineral Density Measured at 3 Sites and Risk of Incident Ischemic Stroke Among Women in the Framingham Heart Study (n=654)

			Age-Adjusted			Multivariable-Adjusted*		
Quintile of BMD	No. of Participants	No. of Ischemic Strokes	HR	95% CI	*P*	HR	95% CI	*P*
Femoral neck								
Q1	122	18	1.37	0.69–2.73	0.36	1.33	0.66–2.68	0.42
Q2	124	18	1.25	0.63–2.38	0.53	1.30	0.65–2.59	0.45
Q3	126	15	1.00			1.00		
Q4	124	17	1.18	0.59–2.36	0.65	0.92	0.45–1.88	0.81
Q5	124	24	1.72	0.90–3.29	0.10	1.41	0.72–2.76	0.32
Trochanter								
Q1	125	25	2.17	1.07–4.38	0.03	2.36	1.15–4.83	0.02
Q2	127	19	1.13	0.51–2.52	0.76	1.07	0.48–2.41	0.87
Q3	128	9	1.00			1.00		
Q4	131	18	1.90	0.93–3.86	0.08	1.77	0.86–3.64	0.12
Q5	126	24	2.38	1.19–4.77	0.01	1.94	0.94–3.95	0.07
Radius								
Q1	122	22	3.18	1.48–6.82	0.00	2.92	1.55–6.34	0.01
Q2	123	12	2.10	0.95–4.65	0.07	1.94	0.87–4.32	0.11
Q3	125	12	1.00			1.00		
Q4	123	21	2.07	0.93–4.60	0.08	1.68	0.74–3.80	0.21
Q5	123	24	2.92	1.36–6.28	0.01	2.43	1.12–5.30	0.03

Q indicates quintile; BMD, bone mineral density.

*Adjusted for age, systolic blood pressure, atrial fibrillation, diabetes, current smoking, cardiovascular disease, estrogen use, and body mass index.

linear association of decreasing BMD and increasing stroke risk, we observed a U-shaped pattern. Women in the lowest quintiles of BMD (trochanter and radius) had elevated risk. This finding supports the estrogen deficiency-stroke hypothesis, although other explanations are possible. Bone metabolism and atherosclerosis share factors including osteopontin and osteocalcin, as well as other potential pathogenic contributors such as oxidized lipids and hypertension.[35,36] This link is supported by an association between low BMD and carotid plaques.[37] The finding of elevated stroke risk with the highest quintile of BMD is unexpected and could be real but could also be the result of misspecification of our model, residual confounding, or selection bias given the age at which BMD was measured in this study.

More research is needed to understand the impact of endogenous estrogen on stroke risk. However, given the harmful association of HRT with stroke in recent trials and negative findings of studies of endogenous estrogen and nonstroke cardiovascular disease end points, alternate hormonal pathways, including changes in androgens and sex-hormone binding globulin with menopause, should be explored. Lower levels of sex-hormone binding globulin (SHBG) and higher levels of free androgen index (FAI) have been associated with cardiovascular disease,[32] but again, data on stroke are lacking. Low SHBG and high FAI were also related to an adverse cardiovascular risk factor profile, including higher insulin, glucose, lipids, and hemostatic and inflammatory markers, in a study of perimenopausal women.[38] Estradiol was also associated with an adverse risk factor profile but to a lesser degree. These findings suggest that the association of age at menopause and stroke risk may be mediated through changes in risk factors which occur with menopause, although associations remained after adjustment for risk factors in this study. Alternatively, an adverse cardiovascular risk factor profile in premenopausal women may be associated with earlier menopause.[39]

Some limitations warrant discussion. The population was limited to white women who were recruited in 1948; therefore, results may not be generalizable to different populations or to more recent birth cohorts. Although age at natural menopause in the Framingham population is similar to estimates in more recent cohorts,[40,41] there have been temporal trends in increasing age at menopause.[42] Oral contraceptive use and use of hormone replacement therapy were uncommon in this cohort because of the study time period limiting generalizability to more recent birth cohorts with a greater prevalence of these medications. Similarly, secular trends in stroke risk factors or their treatment may limit generalizability. Women with stroke before age 60 were excluded. Although ischemic stroke was rare before 60, if early menopause is associated with stroke at younger ages, the association of age at menopause and stroke may differ from that presented. Similarly, the secondary analysis followed women prospectively from BMD measurement, which occurred on average at 76 years. Women who experienced stroke or who died before BMD measurement were not included. This may have introduced bias and suggests that our results should be confirmed in different populations and across a broader range of ages. Although we adjusted for confounders, with a focus on factors known to influence stroke in this population, there may be other unaccounted for confounders. For example, we did not include metabolic syndrome, measures of central adiposity or parity, which may be confounders, because they were not available for this population for the time frame under study. Sample sizes and numbers of events were small in some analyses, which may have limited power. This study relied on self-reported menopausal status which may be subject to recall bias, although the prospective biennial exams minimize this possibility.

Summary

Given the increased stroke burden in women, it is critical to understand risk factors unique to women so that new strategies for prevention can be considered. Results from the current study demonstrated an elevated risk of ischemic stroke in women with early menopause and possible POF and in women with low BMD. These findings raise the hypothesis that estrogen deficiency may play a role in ischemic stroke but current evidence regarding this hypothesis is inconsistent. Alternate hypotheses, including the role of androgens and/or a common cause of BMD and stroke, are also possible. Future studies, with measures of endogenous hormones, are needed to unravel the relationship between hormonal changes that occur with menopause, either premature or at the usual onset, and ischemic stroke.

Sources of Funding

Dr Lisabeth is funded by National Institute for Neurological Disorders and Stroke K23 NS050161. This work was supported by the Framingham Heart Study's National Heart, Lung and Blood Institute's contract (N01-HC-25195) and grants from the National Institute of Neurological Disorders and Stroke (5R01-NS 17950) and the National Institute of Arthritis and Musculoskeletal and Skin Diseases and the National Institute on Aging (R01 AR/AG 41398).

Disclosures

None.

References

1. Kung HC, Hoyert DL, Xu J, Murphy SL. Deaths: final data for 2005. *Natl Vital Stat Rep*. 2008;56:1–120.
2. Seshadri S, Beiser A, Kelly-Hayes M, Kase CS, Au R, Kannel WB, Wolf PA. The lifetime risk of stroke: estimates from the Framingham Study. *Stroke*. 2006;37:345–350.
3. Reeves MJ, Bushnell CD, Howard G, Gargano JW, Duncan PW, Lynch G, Khatiwoda A, Lisabeth L. Sex differences in stroke: epidemiology, clinical presentation, medical care, and outcomes. *Lancet Neurol*. 2008;7:915–926.
4. Atsma F, Bartelink ML, Grobbee DE, van der Schouw YT. Postmenopausal status and early menopause as independent risk factors for cardiovascular disease: a meta-analysis. *Menopause*. 2006;13:265–279.
5. Cooper GS, Sandler DP. Age at natural menopause and mortality. *Ann Epidemiol*. 1998;8:229–235.
6. de Kleijn MJ, van der Schouw YT, Verbeek AL, Peeters PH, Banga JD, van der Graaf Y. Endogenous estrogen exposure and cardiovascular mortality risk in postmenopausal women. *Am J Epidemiol*. 2002;155:339–345.
7. Hu FB, Grodstein F, Hennekens CH, Colditz GA, Johnson M, Manson JE, Rosner B, Stampfer MJ. Age at natural menopause and risk of cardiovascular disease. *Arch Intern Med*. 1999;159:1061–1066.
8. de Lecinana MA, Egido JA, Fernandez C, Martinez-Vila E, Santos S, Morales A, Martinez E, Pareja A, Alvarez-Sabin J, Casado I. Risk of ischemic stroke and lifetime estrogen exposure. *Neurology*. 2007;68:33–38.
9. Varenna M, Binelli L, Zucchi F, Ghiringhelli D, Gallazzi M, Sinigaglia L. Prevalence of osteoporosis by educational level in a cohort of postmenopausal women. *Osteoporos Int*. 1999;9:236–241.

10. Ito M, Yamada M, Hayashi K, Ohki M, Uetani M, Nakamura T. Relation of early menarche to high bone mineral density. *Calcif Tissue Int*. 1995; 57:11–14.
11. Nilas L, Christiansen C. Bone mass and its relationship to age and the menopause. *J Clin Endocrinol Metab*. 1987;65:697–702.
12. Parazzini F, Bidoli E, Franceschi S, Schinella D, Tesio F, La Vecchia C, Zecchin R. Menopause, menstrual and reproductive history, and bone density in northern Italy. *J Epidemiol Community Health*. 1996;50: 519–523.
13. Slemenda C, Longcope C, Peacock M, Hui S, Johnston CC. Sex steroids, bone mass, and bone loss: a prospective study of pre-, peri-, and postmenopausal women. *J Clin Invest*. 1996;97:14–21.
14. Sowers MR, Jannausch M, McConnell D, Little R, Greendale GA, Finkelstein JS, Neer RM, Johnston J, Ettinger B. Hormone predictors of bone mineral density changes during the menopausal transition. *J Clin Endocrinol Metab*. 2006;91:1261–1267.
15. Browner WS, Pressman AR, Nevitt MC, Cauley JA, Cummings SR. Association between low bone density and stroke in elderly women: the Study of Osteoporotic Fractures. *Stroke*. 1993;24:940–946.
16. Mussolino ME, Madans JH, Gillum RF. Bone mineral density and stroke. *Stroke*. 2003;34:e20–22.
17. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health*. 1951;41:279–281.
18. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. *Stroke*. 1991;22:312–318.
19. Wolf PA, D'Agostino RB, O'Neal MA, Sytkowski P, Kase CS, Belanger AJ, Kannel WB. Secular trends in stroke incidence and mortality: the Framingham Study. *Stroke*. 1992;23:1551–1555.
20. Luborsky JL, Meyer P, Sowers MF, Gold EB, Santoro N. Premature menopause in a multi-ethnic population study of the menopause transition. *Hum Reprod*. 2003;18:199–206.
21. Torgerson DJ, Thomas RE, Reid DM. Mothers and daughters menopausal ages: is there a link? *Eur J Obstet Gynecol Reprod Biol*. 1997;74:63–66.
22. Burger HG, Dudley EC, Hopper JL, Groome N, Guthrie JR, Green A, Dennerstein L. Prospectively measured levels of serum follicle-stimulating hormone, estradiol, and the dimeric inhibins during the menopausal transition in a population-based cohort of women. *J Clin Endocrinol Metab*. 1999;84:4025–4030.
23. van der Graaf Y, de Kleijn MJ, van der Schouw YT. Menopause and cardiovascular disease. *J Psychosom Obstet Gynaecol*. 1997;18:113–120.
24. Wildman RP, Colvin AB, Powell LH, Matthews KA, Everson-Rose SA, Hollenberg S, Johnston JM, Sutton-Tyrrell K. Associations of endogenous sex hormones with the vasculature in menopausal women: the Study of Women's Health Across the Nation (SWAN). *Menopause*. 2008;15: 414–421.
25. Anderson GL, Limacher M, Assaf AR, Bassford T, Beresford SA, Black H, Bonds D, Brunner R, Brzyski R, Caan B, Chlebowski R, Curb D, Gass M, Hays J, Heiss G, Hendrix S, Howard BV, Hsia J, Hubbell A, Jackson R, Johnson KC, Judd H, Kotchen JM, Kuller L, LaCroix AZ, Lane D, Langer RD, Lasser N, Lewis CE, Manson J, Margolis K, Ockene J, O'Sullivan MJ, Phillips L, Prentice RL, Ritenbaugh C, Robbins J, Rossouw JE, Sarto G, Stefanick ML, Van Horn L, Wactawski-Wende J, Wallace R, Wassertheil-Smoller S. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative Randomized Controlled Trial. *JAMA*. 2004;291:1701–1712.
26. Grady D, Herrington D, Bittner V, Blumenthal R, Davidson M, Hlatky M, Hsia J, Hulley S, Herd A, Khan S, Newby LK, Waters D, Vittinghoff E, Wenger N. Cardiovascular disease outcomes during 6.8 years of hormone therapy: Heart and Estrogen/progestin Replacement Study Follow-up (HERS II). *JAMA*. 2002;288:49–57.
27. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC, Kotchen JM, Ockene J. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative Randomized Controlled Trial. *JAMA*. 2002;288:321–333.
28. Rossouw JE, Prentice RL, Manson JE, Wu L, Barad D, Barnabei VM, Ko M, LaCroix AZ, Margolis KL, Stefanick ML. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *JAMA*. 2007;297:1465–1477.
29. Wassertheil-Smoller S, Hendrix SL, Limacher M, Heiss G, Kooperberg C, Baird A, Kotchen T, Curb JD, Black H, Rossouw JE, Aragaki A, Safford M, Stein E, Laowattana S, Mysiw WJ. Effect of estrogen plus progestin on stroke in postmenopausal women: The Women's Health Initiative: a randomized trial. *JAMA*. 2003;289:2673–2684.
30. Price JF, Lee AJ, Fowkes FG. Steroid sex hormones and peripheral arterial disease in the Edinburgh Artery Study. *Steroids*. 1997;62:789–794.
31. Golden SH, Maguire A, Ding J, Crouse JR, Cauley JA, Zacur H, Szklo M. Endogenous postmenopausal hormones and carotid atherosclerosis: a case-control study of the Atherosclerosis Risk in Communities cohort. *Am J Epidemiol*. 2002;155:437–445.
32. Rexrode KM, Manson JE, Lee IM, Ridker PM, Sluss PM, Cook NR, Buring JE. Sex hormone levels and risk of cardiovascular events in postmenopausal women. *Circulation*. 2003;108:1688–1693.
33. Barrett-Connor E, Goodman-Gruen D. Prospective study of endogenous sex hormones and fatal cardiovascular disease in postmenopausal women. *BMJ*. 1995;311:1193–1196.
34. Jorgensen L, Engstad T, Jacobsen BK. Bone mineral density in acute stroke patients: low bone mineral density may predict first stroke in women. *Stroke*. 2001;32:47–51.
35. Demer LL. Vascular calcification and osteoporosis: inflammatory responses to oxidized lipids. *Int J Epidemiol*. 2002;31:737–741.
36. Cappuccio FP, Meilahn E, Zmuda JM, Cauley JA. High blood pressure and bone-mineral loss in elderly white women: a prospective study. Study of Osteoporotic Fractures Research Group. *Lancet*. 1999;354:971–975.
37. Jorgensen L, Joakimsen O, Rosvold Berntsen GK, Heuch I, Jacobsen BK. Low bone mineral density is related to echogenic carotid artery plaques: a population-based study. *Am J Epidemiol*. 2004;160:549–556.
38. Sutton-Tyrrell K, Wildman RP, Matthews KA, Chae C, Lasley BL, Brockwell S, Pasternak RC, Lloyd-Jones D, Sowers MF, Torrens JI. Sex-hormone-binding globulin and the free androgen index are related to cardiovascular risk factors in multiethnic premenopausal and perimenopausal women enrolled in the Study of Women Across the Nation (SWAN). *Circulation*. 2005;111:1242–1249.
39. Kok HS, van Asselt KM, van der Schouw YT, van der Tweel I, Peeters PH, Wilson PW, Pearson PL, Grobbee DE. Heart disease risk determines menopausal age rather than the reverse. *J Am Coll Cardiol*. 2006;47: 1976–1983.
40. Hefler LA, Grimm C, Heinze G, Schneeberger C, Mueller MW, Muendlein A, Huber JC, Leodolter S, Tempfer CB. Estrogen-metabolizing gene polymorphisms and age at natural menopause in caucasian women. *Hum Reprod*. 2005;20:1422–1427.
41. Kok HS, Onland-Moret NC, van Asselt KM, van Gils CH, van der Schouw YT, Grobbee DE, Peeters PH. No association of estrogen receptor alpha and cytochrome p450c17alpha polymorphisms with age at menopause in a Dutch cohort. *Hum Reprod*. 2005;20:536–542.
42. Nichols HB, Trentham-Dietz A, Hampton JM, Titus-Ernstoff L, Egan KM, Willett WC, Newcomb PA. From menarche to menopause: trends among US women born from 1912 to 1969. *Am J Epidemiol*. 2006;164:1003–1011.

J Thromb Thrombolysis (2009) 28:106–116
DOI 10.1007/s11239-009-0324-4

Highlights from the I international symposium of thrombosis and anticoagulation in internal medicine, October 23–25, 2008, Sao Paulo, Brazil

Renato D. Lopes · Richard C. Becker · David Garcia · Elaine M. Hylek · Christopher B. Granger · Dayse M. Lourenço · Helena B. Nader · Francisco H. A. Maffei · Fernando H. Y. Cesena · José C. Nicolau · Ari Timmerman · Ana T. Rocha · Edison F. de Paiva · Helio P. Guimaraes · Antonio C. Carvalho · Elbio A. D'Amico · Carlos A. Moreira-Filho · José M. Aldrighi · Antonio C. Lopes

Published online: 26 March 2009

Abstract The importance of thrombosis and anticoagulation in clinical practice is rooted firmly in several fundamental constructs that can be applied both broadly and globally. Awareness and the appropriate use of anticoagulant therapy remain the keys to prevention and treatment. However, to assure maximal efficacy and safety, the clinician must, according to the available evidence, choose the right drug, at the right dose, for the right patient, under the right indication, and for the right duration of time. The first *International Symposium of Thrombosis and Anticoagulation in Internal Medicine* was a scientific program developed by clinicians for clinicians. The primary objective of the meeting was to educate, motivate and inspire internists, cardiologists and hematologists by convening national and international visionaries, thought-leaders and dedicated clinician-scientists in Sao Paulo, Brazil. This article is a focused summary of the symposium proceedings.

Keywords Thrombosis · Anticoagulation · Internal medicine

R. D. Lopes · R. C. Becker (✉) · C. B. Granger
Duke University School of Medicine, Durham, NC, USA
e-mail: becke021@mc.duke.edu

D. Garcia
University of New Mexico Health Sciences Center, Albuquerque, NM, USA

E. M. Hylek
Boston University Medical Center, Boston, MA, USA

D. M. Lourenço · H. B. Nader · H. P. Guimaraes · A. C. Carvalho · A. C. Lopes
Federal University of Sao Paulo, Sao Paulo, SP, Brazil

F. H. A. Maffei
Botucatu School of Medicine, Botucatu, SP, Brazil

F. H. Y. Cesena · J. C. Nicolau · E. F. de Paiva · E. A. D'Amico · C. A. Moreira-Filho · J. M. Aldrighi
University of Sao Paulo, Sao Paulo, SP, Brazil

A. Timmerman
Dante Pazzanese Institute of Cardiology, Sao Paulo, Brazil

A. T. Rocha
Federal University of Bahia, Salvador, BA, Brazil

The importance of thrombosis and anticoagulation in clinical practice is rooted firmly in several fundamental constructs that can be applied both broadly and globally. First, hemostasis, representing the physiological or protective phenotype of thrombosis is life-sustaining. Second, thrombotic disorders are common and occur in patients of all ages, races, ethnicities and medical/surgical conditions. Third, in many instances, thrombosis as the proximate cause of venous thromboembolism (VTE), stroke and myocardial infarction is preventable and treatable. Awareness and the appropriate use of anticoagulant therapy remain the keys to prevention and treatment. However, to assure maximal efficacy and safety, the clinician must, according to the available evidence, choose the right drug, at the right dose, for the right patient, under the right indication, and for the right duration of time.

The opportunity to share ideas, and advance the care of patients with thrombotic disorders, is the fundamental tenet of practicing clinicians worldwide. This can only be accomplished through knowledge gained from carefully designed, meticulously conducted and honestly interpreted translational and clinical research.

True to the lasting spirit of scholarly interchange, the first *International Symposium of Thrombosis and Anticoagulation in Internal Medicine* was a scientific program developed by clinicians for clinicians. The symposium was promoted by the Federal University of Sao Paulo together with the Brazilian Society of Internal Medicine and the Duke Clinical Research Institute of Duke University School of Medicine. It was also supported by the Brazilian and Paulista Societies of Cardiology. The chairmen of the meeting were Dr. Renato D. Lopes and Dr. Richard C. Becker, both from Duke University School of Medicine and the Duke Clinical Research Institute. The symposium took place in Sao Paulo, Brazil from the 23–25 of October, 2008.

The primary objective for the 3 days of academic presentations and open discussions was to educate, motivate and inspire internists, cardiologists and hematologists by convening national and international visionaries, thought-leaders and dedicated clinician-scientists in Sao Paulo, Brazil. The following is a focused summary of the symposium proceedings.

Thrombosis—what is the role of the endothelium?

It is widely recognized that the endothelium is not a static barrier between the vessel lumen and the vessel wall, but rather a dynamic organ that synthesizes, secretes and regulates a wide variety of substances, including nitric oxide (NO), cytokines, chemokines, adhesion molecules and mediators that affect the function of different cells.

The endothelium is the major regulator of vascular homeostasis. Under normal conditions, the endothelium promotes vasodilation and exerts antioxidant, anti-inflammatory effects, inhibiting leukocyte adhesion and transmigration. It also inhibits smooth muscle cell proliferation and migration, platelet adhesion and aggregation, and displays both anticoagulant and profibrinolytic properties.

Several traditional risk factors for atherosclerosis (systemic arterial hypertension, diabetes mellitus, hypercholesterolemia, smoking, ageing) adversely affect endothelial cell function, even before the development of obstructive atherosclerotic plaques. The resultant "endothelial dysfunction" is characterized by a propensity toward vasoconstriction, release of inflammatory mediators and predisposition to thrombosis (Fig. 1). Endothelial dysfunction is considered a marker of early atherosclerosis, and is considered the pathophysiological foundation for atherosclerotic plaque progression, which include impaired vascular repair leading to erosion and disruption.

Oxidative stress and the free-radical-mediated neutralization of nitric oxide (NO) are closely linked to atherothrombosis. Apart from NO, endothelial cells secrete

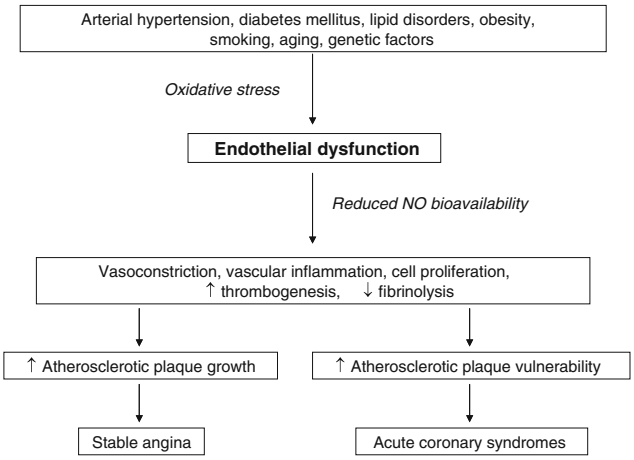


Fig. 1 Cardiovascular risk factors induce endothelial dysfunction, closely associated with oxidative stress and characterized by reduced nitric oxide (NO) bioavailability. The dysfunction of vascular endothelial cells establishes an environment characterized by a propensity toward vasoconstriction; the release of inflammatory mediators; and. a marked predisposition to thrombosis. Endothelial dysfunction is a fundamental pathophysiological alteration that governs initiation, growth and erosion/rupture of atherosclerotic plaque-the proximate cause of clinical syndromes, including stable or unstable angina and acute myocardial infarction

other antithrombotic substances, such as prostacyclin (PGI2), CD39, thrombomodulin, heparan sulfate, and tissue plasminogen activator (tPA).The vascular endothelium also synthesizes prothrombotic substances, such as von Willebrand factor, P-selectin, tissue factor and plasminogen activator inhibitor-1 (PAI-1). Under normal conditions, inhibitors of platelet activation and coagulalation predominate, allowing thrombin generation and fibrin formation to be tightly regulated. In contrast, when the endothelium is dysfunctional, a shift toward a prothrombotic state takes place. An increase in the expression of selectins promotes platelet adhesion to the endothelium. Adherent, activated platelets interact with and stimulate endothelial cells and monocytes, further amplifying the inflammatory environment that inherently typifies atherosclerosis.

There is ample evidence linking endothelial dysfunction and changes in NO metabolism to atherothrombosis, providing mechanistic support for an observed independent association between endothelial dysfunction and future cardiovascular events. Accordingly, treatment strategies targeting endothelium/NO pathways may promote vascular health and reduce thrombotic events. For example, risk factor control, statins, angiotensin-converting enzyme inhibitors, physical exercise, antioxidants and red wine improve endothelial performance. Cell-based therapeutics that target signaling pathways implicated in endothelial dysfunction may confer additional benefit and warrant further investigation.

Thrombosis and hemostasis

Fibrin clot formation, the basis for both protective hemostasis and pathological thrombosis, is a complex, cell-based process represented by several integrated biochemical steps designed to maintain blood fluidity and vascular integrity.

Following vascular injury, platelets tether and ultimately adhere to collagen fibers within the subendothelium-a physical event mediated by platelet membrane glycoproteins GPIaIIa, GP VI, GP 1b, GPIIbIIIa, and von Willebrand factor, which represent the predominant ligand for both transient and stable adhesion and fibrinogen that builds a "bridge" between adjacent platelets, establishing a stable aggregate.

Activated platelets expose phospholipids that, in turn, provide a surface for coagulation protein assembly. Tissue factor, in the presence of calcium ions and the exposed phospholipids on the activated platelet membrane, initiates activation of factor VII at the site of endothelial injury. Activated factor VII (VIIa) activates factor IX which, in the presence of factor VIII, forms the X-ase complex that activates factor X. Factor Xa, in the presence of factor Va forms the prothrombinase complex that cleaves the prothrombin molecule, forming a small amount of thrombin. Thrombin is capable of activating both factors V and VIII, thus creating a positive feedback loop that leads to additional thrombin formation.

Thrombin, by cleaving fibrinopeptides A and B from fibrinogen, generates fibrin monomers that polymerize to form the lattice of a fibrin clot. Finally, thrombin activates factor XIII, which promotes covalent binding within gamma chains of fibrin to stabilize the fibrin clot.

Several intrinsic regulatory mechanisms, such as fibinolysis by plasmin or the inactivation of coagulation proteins by endogenous anticoagulants like antithrombin III and activated protein C provide a counterbalance that, in most instances, prevents pathologic (or unwanted) thrombosis.

Pharmacokinetics and pharmacodynamics of vitamin K antagonists

Vitamin K antagonists (VKAs) have been the only oral anticoagulants available for clinical use until now, and have been used for more than 60 years. Their effectiveness has been demonstrated for primary and secondary VTE prophylaxis, prevention of systemic embolism in patients with atrial fibrillation, prosthetic cardiac valves, or large myocardial infarctions, particularly with mural thrombosis.

VKAs exert their effect by inhibiting vitamin K oxide reductase, thus limiting the amount of reduced vitamin K available for the γ-carboxylation of the glutamate residues on the N-terminal regions of coagulation proteins II, VII, IX and X. This specific carboxylation step is an absolute prerequisite for calcium- dependent binding to cofactors on phospholipid surfaces- its absence reduces the coagulant potential of the blood. VKAs also interfere with carboxylation of the anticoagulant proteins C, S and Z and several proteins synthesized in bone.

There are two VKAs available in Brazil: warfarin, the most commonly used, with a half-life of about 35 h and phenprocoumon, a much longer-acting agent, with a half-life of 5.5 days. Both preparations are metabolized by the cytochrome P450 system in the liver.

Individual response to VKAs varies greatly depending on genetic factors, concomitant diseases and both medication and food interactions. For these reasons, close monitoring of treatment with VKAs is necessary; to include INR determinations at least every 4 weeks, clinical interview, and ascertainment of new (or changes in) medications or foods. The INR provides a reliable and evidence-based parameter for effectiveness and risk of bleeding, and must be maintained in the range between 2 and 3 for most patients. The effect of VKAs may be attenuated or fully reversed by vitamin K administration.

Antithrombotics in acute coronary syndrome with st-segment elevation

Advances in the management of patients with STEMI have been achieved with antithrombotic pharmacotherapies. Further contributions are likely with new compounds such as oral, direct factor Xa inhibitors (otamixaban, apixaban, Du 183b, Tak 442 and rivaroxaban), platelet P2Y12 receptor blockers (cangrelor, AZD 6140, prasugrel) and platelet thrombin receptor (PAR-1) blocker. Among the approved drugs, important data has been presented in the last few years. A brief summary is presented below:

(A) *Antiplatelet drugs*: Clopidogrel was tested against placebo in two STEMI studies that included a total of nearly 49,000 patients. The primary endpoint in CLARITY was the composite of an occluded culprit coronary artery, death or reinfarction at the time of the coronary angiography or at hospital discharge. A 36% risk reduction ($P < 0.001$) in the main endpoint, in favor of clopidogrel was demonstrated. In COMMIT, the primary endpoint was a composite of death, reinfarction or stroke up to 28 days, and the results showed a 9% relative risk reduction ($P = 0.002$), also in favor of clopidogrel. In both studies, the bleeding rates were similar between clopidogrel and placebo. The ACC/AHA STEMI

guidelines recommend clopidogrel whether or not reperfusion therapy has been provided. More recently the TRITON Trial was published, and its sub-analysis for patients with STEMI was presented during the European Society of Cardiology Congress. The results were as follows: incidence of the primary endpoint (CV death, MI, stroke) at 30 days of 12.4% and 10% for clopidogrel and prasugrel, respectively ($P = 0.002$); stent thrombosis of 2.8% and 1.6%, respectively($P = 0.02$). There was no significant difference in bleeding between groups.

(B) *Antithrombin drugs*: In 2001 the HERO-2 study was published, showing that bivalirudin, in addition to streptokinase, had a similar 30-day mortality rate (primary endpoint) compared to UFH, but an increase in the incidence of bleeding. The CREATE study was published in 2005, and reported a lower incidence of death, reinfarction or stroke at 7 and 30 days (main endpoint—HR of 0.87, $P = 0.014$) for reviparin compared to placebo, but at the cost of increased major bleeds (HR = 2.49, $P = 0.001$). The OASIS-6 Trial, published in 2006 included 12,000 patients treated initially with either fibrinolytics or primary PCI, and then randomized to fondaparinux, placebo or UFH. The primary efficacy endpoint was the composite of death or MI at 30 days. The observed hazard-ratios were as follows: for the comparison of fondaparinux and placebo or UFH, 0.86 ($P = 0.008$); for the comparison of fondaparinux and placebo, 0.79 ($P < 0.05$); and, for the comparison of fondaparinux and UFH, 0.95 ($P =$ NS). Moreover, there was an unfavorable interaction ($P = 0.03$) between fondaparinux and primary PCI, with a hazard-ratio of 1.20 for patients undergoing PCI, and 0.88 for those without primary PCI. There were no significant differences in bleeding, between fondaparinux and UFH. A meta-analysis including 27,000 patients treated with LMWH or UFH, revealed a significant 16% net clinical benefit in favor of enoxaparin in relation to UFH.Based on the available information, the ACC/AHA guidelines recommend UFH, enoxaparin or fondaparinux for patients with STEMI, but caution that "because of the risk of catheter thrombosis, fondaparinux should not be used as the sole anticoagulant to support PCI".

Anticoagulation in acute coronary syndromse without ST-segment elevation

Platelets play a pivotal role in the transformation of a stable to an unstable atherosclerotic plaque. Disruption of an atherosclerotic plaque exposes the subendothelial matrix (e.g. collagen and tissue factor) to circulating blood.

Antiplatelet therapy, a cornerstone of therapy in NSTE-ACS, is directed at decreasing the formation of thromboxane A_2 (aspirin), inhibiting the P2Y12-mediated platelet activation (thienopyridines) and directly inhibiting platelet aggregation (GP IIb/IIIa inhibitors). In four randomized trials, the use of aspirin versus placebo was associated with a 50% reduction in death or MI. Therefore, after an initial dose of 162–325 mg, a dose of 75–100 mg daily is recommended in patients with ACS. Clopidogrel when added to aspirin, confers a 20% reduction in cardiovascular death, MI, or stroke, compared with aspirin alone, in both low- and high-risk patients with NSTE-ACS. The dose of clopidogrel for medical treatment is 300 mg, followed by 75 mg daily. The benefit of GP IIb/IIIa inhibitors is most evident when used in high-risk patients (e.g. elevated troponin, diabetes mellitus). Abciximab is currently approved only in patients undergoing PCI within 12 h of treatment initiation. Eptifibatide and tirofiban can be used in either a conservative or intervention-based strategy.

Anticoagulation, traditionally with unfractionated heparin (UFH) is another cornerstone of therapy for patients with NSTE-ACS. A meta-analysis showed a 33% reduction in death or MI comparing UFH plus aspirin versus aspirin alone. Low-molecular-weight heparins (LMWHs) combine factor IIa and Xa inhibition and thus inhibit both the action and generation of thrombin. LMWH has several potential advantages over UFH. In ESSENCE and TIMI 11B trials, enoxaparin was superior to UFH, with a statistically significant 20% reduction in events among moderate-risk patients. In the SYNERGY trial, including high-risk patients managed with an invasive strategy enoxaparin was found to be noninferior to UFH. The standard dose of enoxaparin is 1 mg/kg given subcutaneously (SC) every 12 h.

Fondaparinux is a synthetic, indirect, specific factor Xa inhibitor that requires antithrombin III for its pharmacodynamic activity. In the OASIS-5 trial, fondaparinux at a dose of 2.5 mg SC once daily produced similar rates of death, MI or refractory ischemia to enoxaparin, but with substantially less major bleeding.

Direct thrombin inhibitors have a theoretical advantage over heparin compounds; they do not require antithrombin III and can inhibit clot-bound thrombin; they do not interact with plasma proteins, they provide a very stable level of anticoagulation and do not cause thrombocytopenia. In the ACUITY trial, patients were managed with an early invasive strategy, and randomized to receive either bivalirudin alone, enoxaparin or UFH plus a GP IIb/IIIa inhibitor or bivalirudin plus a GP IIb/IIIa inhibitor. No differences were observed between the three treatment arms for the composite of death, MI or unplanned

revascularization at 30 days, but bivalirudin caused less bleeding compared with the other two arms (3% vs. 5% vs. 7%, respectively; $P < 0{,}001$).

Atrial fibrillation and acute coronary syndromes

Atrial fibrillation (AF) is a common complication of myocardial infarction (MI) with a reported incidence ranging from 5% to 23%. It is associated with worse in-hospital and long-term outcomes.

Although antithrombotic therapy is important in the treatment of patients with both AF and MI, the combined administration of aspirin, thienopyridines, and a vitamin K antagonist (triple therapy) increases the risk of bleeding. The current AF guidelines recommend VKA anticoagulant therapy for patients with a $CHADS_2$ score ≥ 2 as a class IA recommendation. Guidelines also recommend low dose aspirin (81 mg/day), clopidogrel, and warfarin (with target INR 2.0–2.5) after stenting for patients with acute coronary syndromes and a concomitant indication for oral anticoagulation. The available literature on the subject of "triple therapy" shows that patients with AF and ACS are not discharged from the hospital on a VKA. Paradoxically, patients deemed to be at highest risk of stroke ($CHADS_2$ score ≥ 2) are least likely to be treated due to physician concerns over the potential risk of bleeding.

New onset AF that develops in the setting of ACS continues to be a marker of poor short and long-term prognosis. There appears to be a "treatment-risk paradox" concerning VKA use, highlighting the need for additional investigation to better understand risk-benefit relationships and optimal management strategies.

Atrial fibrillation

The prevalence of atrial fibrillation (AF) is increasing worldwide. Older age, hypertension, heart failure, and obesity all increase the risk of developing AF. Atrial fibrillation is a potent risk factor for stroke raising the risk on average 5-fold. The seminal trials in nonvalvular AF demonstrated the remarkable efficacy of warfarin in stroke prevention with a risk reduction of 68%. Current guidelines recommend warfarin for patients with stroke, transient ischemic attack, or systemic embolism and for patients with two or more of the following risk factors: age 75 years and greater, hypertension, heart failure, diabetes mellitus. Aspirin or warfarin is recommended if only one of these risk factors is present, depending on patient preference. Despite the efficacy of warfarin, numerous studies have shown that only about one-half of patients are treated. Older age and perceived bleeding risk are the most often-cited negative predictive factors. Difficulty with warfarin monitoring is also a major obstacle to its use.

Recent randomized trials have focused on maintenance of sinus rhythm and the potential role of antiplatelet therapy in stroke prevention in AF. These studies have illustrated that prolonged maintenance of sinus rhythm remains elusive and antiplatelet agents, either as monotherapy or in combination (aspirin plus clopidogrel), is less efficacious than anticoagulant treatment in AF. Because the 30-day mortality of AF-related stroke is 24%, use of less efficacious agents mandates careful consideration. Current trials in AF will explore the efficacy of novel anticoagulant drugs as potential replacements for warfarin.

Hemorrhage is a serious adverse effect of antithrombotic therapy. The clinical dilemma in atrial fibrillation is that both risk of stroke and risk of hemorrhage increase with age. Older age is also associated with lower warfarin dose requirements and a slower return to therapeutic levels following an episode of excessive anticoagulation. Concomitant antiplatelet therapy also increases the risk of extracranial and intracranial hemorrhage. Strategies to minimize these risks include vigilant monitoring of the International Normalized Ratio (INR) with a target INR of 2.5, maintaining blood pressure less than 130/80 mmHg, and minimizing concomitant aspirin use. It is important to emphasize that the frequency and severity of stroke in AF outweigh the risk of warfarin-related hemorrhage for most patients.

New studies on atrial fibrillation

There are several ongoing phase II and phase III trials of antithrombotic therapy for stroke prevention in atrial fibrillation. They are summarized in Figs. 2, 3 and 4. The trials of warfarin (versus placebo) for stroke prevention in atrial fibrillation included approximately 3,800 patients. Recently completed and ongoing trials will include a total of 71,600 patients. The results of these trials are eagerly awaited and will be important to better understand this common disease and to improve patient care.

VTE prophylaxis in Brazil—a global perspective

VTE, as a public health problem, frequently affects hospitalized patients at risk for thromboembolic events and represents a major field for prophylaxis interventions. Guideline-recommended thromboprophylaxis reduces the burden of VTE, both in at-risk surgical and medical patients. However, multinational, prospective registries such as the IMPROVE show that risk factors are very common in medical patients (93% have at least one risk

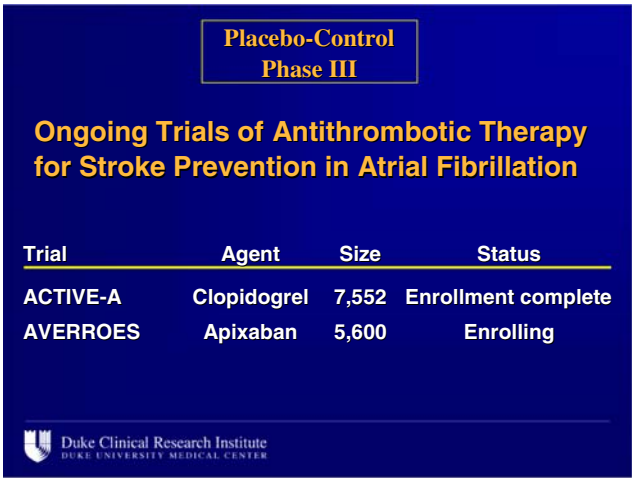


Fig. 2 Ongoing Placebo-control phase III trials of antithrombotic therapy for stroke prevention in atrial fibrillation

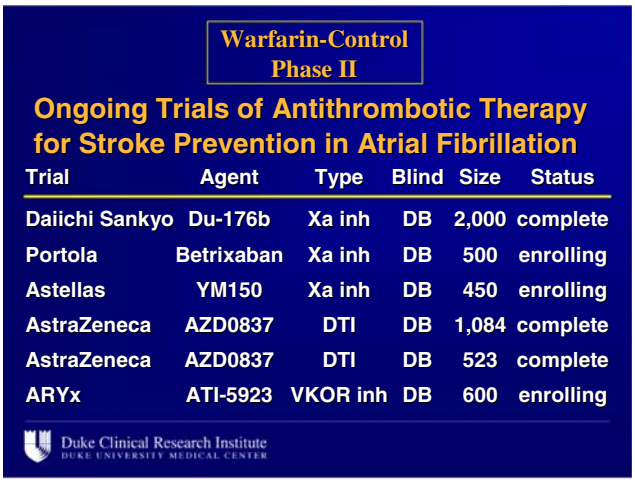


Fig. 3 Ongoing Warfarin-control phase II trials of antithrombotic therapy for stroke prevention in atrial fibrillation

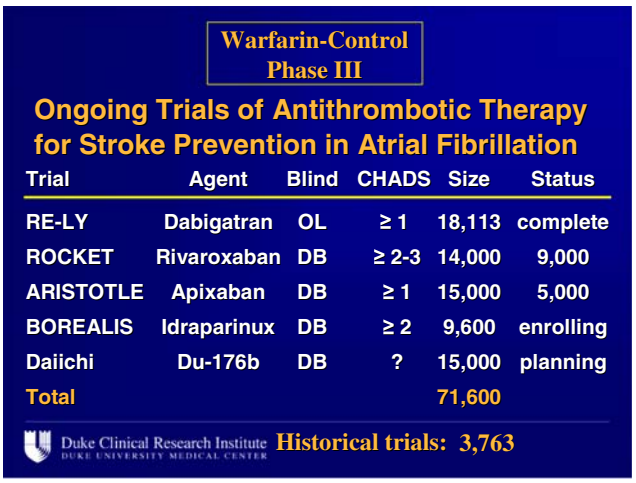


Fig. 4 Ongoing Warfarin-control phase III trials of antithrombotic therapy for stroke prevention in atrial fibrillation

factor), but that VTE prophylaxis is underutilized in the participating countries. In Brazil, the utilization of prophylaxis was significantly less than "rest of world" (36% vs. 51%), particularly among the public hospitals. The ENDORSE study, a large, global observational study of VTE prophylaxis in medical and surgical patients, included 32 countries, 358 hospitals and 68,183 patients. Using a cross-sectional design, the study showed that more than half of the hospitalized patients were at-risk of VTE and that prophylaxis was underutilized in both surgical and medical patients (59% and 40%, respectively). Therefore, there is a clear gap between guidelines and clinical practice, observed across many countries.

VTE prophylaxis in special patient populations

Given the coexisting risk of thrombosis and bleeding associated with acute CVA, the PREVAIL study evaluated the efficacy and safety of enoxaparin (40 mg SC daily) versus unfractionated heparin (5,000 IU SC 12–12 h) for the prevention of VTE. The study showed that enoxaparin was superior to UFH for the prevention of VTE and proximal deep vein thrombosis, reducing the overall incidence by 43% without increasing the risk of major bleeding. The EXCLAIM study investigated the potential benefit of prolonged VTE prophylaxis in acutely ill medical patients with recent reduced mobility. Hospitalized medical patients were randomized to enoxaparin (40 mg daily) or placebo for an additional 28 days after initial 10 day-prophylaxis with enoxaparin. The study showed a significant reduction in VTE (4.9% vs. 2.8%). Major bleeding was significantly more frequent in the enoxaparin group (0.6% vs. 0.1%). One could conclude that highly selected, acutely ill medical patients, including those with reduced mobility, age >75 years, malignancy or previous history of VTE, might benefit from extended prophylaxis beyond the recommended 10 ± 4 days, but at a cost of increased bleeding.

Brazilian guidelines for VTE prophylaxis in medical patients

VTE refers to deep vein thrombosis (DVT) and pulmonary embolism (PE), frequent complications responsible for 10% of deaths in the hospital. Although classically related to surgical procedures, fatal PE can occur in high-risk medical patients as well. VTE prophylaxis among hospitalized patients remains low, partially due to the lack of readily available risk assessment tools and widely implemented "triggers" for ordering prophylactic measures upon hospital admission.

In 2006, 12 Brazilian Medical Societies published the "Brazilian Guideline for VTE Prophylaxis in Medical

Patients" as part of the Guidelines Project of the Brazilian Medical Association (http://www.projetodiretrizes.org.br/volume_4.php). A systematic review was performed, with the objective to identify diseases and conditions associated with VTE and the optimal strategy for its prevention. An algorithm was subsequently developed to assist physicians in day-to-day clinical practice.

Risk assessment

Every medical patient admitted must have his/her VTE risk evaluated. Patients 40 years of age or older, with reduced mobility and at least one additional risk factor for VTE must be considered at risk. In the absence of contraindications, prophylaxis should be provided. Patients younger than 40 years, but having one or more risk factors for VTE may also benefit from prophylaxis.

Prophylaxis

For prophylaxis, once a day SC low molecular weight heparin (enoxaparin 40 mg, dalteparin 5.000 IU, or nadroparin 3,800 and 5,700 IU, respectively for patients weighing 70 kg or more), or SC unfractionated heparin, 5.000 IU three times a day, may be used. Prophylaxis should continue for 6–14 days, even if the patient resumes ambulation.

Improving VTE prophylaxis in medical patients

There are several barriers for implementation of an effective VTE prophylaxis program, beginning with lack of awareness of the recommendations, resistance to change, fear of inducing bleeding, absence of institutional policies, economical barriers, and lack of an adequate risk assessment tool. It is widely recognized that continuing medical education (CME) initiatives, including lectures and dissemination of guidelines are not effective. Multifaceted interventions, targeting specific barriers are more effective than single-strategy interventions. A combined approach must include formal presentations of the guidelines to hospital physicians; distribution of the printed guidelines; creation of a working group to identify local barriers to change; use of printed or electronic reminders, and constant evaluation of physician and institution performance.

VTE prophylaxis programs

Identification of interested personnel is an important first step in the development of a VTE prophylaxis program. The hospital administration must also be engaged and committed to the process and support the establishment of a Commission for VTE Prophylaxis (CVTEP) that should be multidisciplinary, with participation of physicians, nurses, pharmacists, physiotherapists, and hospital quality control personnel. The CVTEP should be proactive, performing a daily evaluation of prophylaxis utilization to include patient selection and dosing in every area of the hospital. The CVTEP should also be responsible for providing physician feedback and establishing mechanisms for continued quality improvement. Additional strategies for success include staff presentations emphasizing VTE prophylaxis in medical, surgical, and subspecialties areas, distribution of educational material, decision-support systems, risk assessment tools, and electronic alerts. An active, integrated, and multifaceted approach may be the key to achieving and maintaining long term compliance with VTE risk evaluation and prophylaxis.

Orthopedics surgery—How to prevent thromboembolic events?

Patients undergoing orthopedic surgery are at increased risk of venous thromboembolic events; proven prophylactic measures are available but are generally underused. The incidence of deep venous thrombosis (DVT) in patients undergoing orthopedic surgery is reported to range from 40% to 60% in patients who did not receive thromboprophylaxis.

Several studies provide evidence of a significant reduction in venous thromboembolic events using low-molecular-weight heparin (LMWH), unfractionated heparin, warfarin, or fondaparinux. A new class of oral anticoagulants, direct factor Xa inhibitors, appears particularly promising.

The RECORD program included several phase III trials that compared the efficacy and safety of rivaroxaban to enoxaparin in patients undergoing major orthopedic surgery. The RECORD 1 investigated VTE thromboprophylaxis in patients undergoing total hip arthroplasty. It was randomized, double-blind study that assigned 4,541 patients to receive either 10 mg of oral rivaroxaban once daily, beginning after surgery, or 40 mg of enoxaparin subcutaneously once daily, beginning the evening before surgery, plus a placebo tablet or injection. A total of 3,153 patients were included in the superiority analysis and 4,433 were included in the safety analysis. There was a highly statistically significant reduction in total VTE favoring rivaroxaban (absolute risk reduction, 2.6%; 95% confidence interval [CI], 1.5–3.7; $P < 0.001$). Major VTE was 0.2% in the rivaroxaban group and 2.0% in the enoxaparin group (absolute risk reduction, 1.7%; 95% CI, 1.0–2.5; $P < 0.001$).

The RECORD 2 study compared the use of rivaroxaban for extended thromboprophylaxis with short-term thromboprophylaxis with enoxaparin. 2,509 patients scheduled to

undergo elective total hip arthroplasty were randomly assigned, stratified according to centre to receive oral rivaroxaban 10 mg once daily for a total of 31–39 days or enoxaparin 40 mg given subcutaneously once daily for 10–14 days (with a placebo tablet given for 31–39 days). Analyses were done using a modified intention-to-treat population. The primary outcome occurred in 2.0% patients in the rivaroxaban group, compared with 9.3% in the enoxaparin group (absolute risk reduction 7.3%; 95% CI 5.2–9.4; $P < 0.0001$). The incidence of any on-treatment bleeding was similar in both groups (6.6% events in the rivaroxaban safety population vs. 5.5% in the enoxaparin safety population; $P = 0.25$).

Finally, in RECORD 3 the efficacy and safety of rivaroxaban in preventing venous thrombosis after total knee arthroplasty was studied. In this randomized, double-blind trial, 2,531 patients who were to undergo total knee arthroplasty received either oral rivaroxaban, 10 mg once daily, beginning 6–8 h after surgery, or subcutaneous enoxaparin, 40 mg once daily, beginning 12 h before surgery. The primary efficacy outcome occurred in 9.6% of rivaroxaban-treated patients and 18.9% of enoxaparin-treated patients (absolute risk reduction, 9.2%; 95% confidence interval [CI], 5.9–12.4; $P < 0.001$). Major VTE occurred in 1.0% and 2.6% of patients, respectively (absolute risk reduction, 1.6%; 95% CI, 0.4–2.8; $P = 0.01$). Major bleeding occurred in 0.6% of patients in the rivaroxaban group and 0.5% of patients in the enoxaparin group ($P =$ NS).

Bridging (perioperative) anticoagulation: risks and benefits

Patients taking VKA anticoagulant therapy may require interruption of treatment to undergo either surgery or an invasive procedure. During temporary discontinuations, the physician (and patient) must weigh the risks and benefits of administering "bridging" (short-acting) anticoagulants such as unfractionated heparin or low molecular weight heparin. Although there is a lack of evidence from randomized controlled trials that defines an optimal perioperative anticoagulation strategy, several prospective cohort studies suggest that, even for patients at high risk of thromboembolism (e.g. a recent pulmonary embolism or a prosthetic mechanical heart valve), peri-procedural LMWH is associated with a low rate of thromboembolic events. The major challenge for clinicians is an even greater paucity of information on the risk of thromboembolism without bridging therapy. Indeed two recently published studies suggest that, for many patients with atrial fibrillation, the risk of simple warfarin interruption may be quite low. The results of these observational studies especially when considered cumulatively with cost, inconvenience and bleeding risk conferred by peri-operative anticoagulants, highlight the need for a randomized, controlled trial of bridging therapy. Such a trial-the Bridge study, funded by the United States National, Heart, Lung and Blood Institute—is ongoing.

Oral anticoagulation in valvular heart disease

Valvular heart disease is associated with a risk of thromboembolism and resulting morbidity and mortality. Warfarin and Phenprocoumon are available in Brazil, decreasing blood levels of vitamin K dependent coagulation factors by 50–75% and biological activity of new factors being synthesized by 20–30%. Patient education is a critical component of treatment, with emphasis on what to do if bleeding occurs, when to perform blood tests, target INR, drug-drug interactions, food-drug interactions and the impact of exercise on warfarin response.

Our approach to patients requiring warfarin is as follows: we begin with a dose of 5 mg and perform an INR measurement on the third and seventh day, adjusting the dose accordingly to achieve the chosen INR. In the setting of acute thrombotic disorders, heparin is maintained until the target INR is achieved. We prefer that our patients continue their regular diet, rather than changing to a diet restricted in vitamin K-containing foods. The optimal INR varies by indication, but typically ranges from 2.0 to 4.0, with values less than 2.0 being associated with thrombosis risk, and those >4.0 posing a risk for serious bleeding. Increased bleeding episodes occur most often within the first 90 days of treatment initiation among patients with uncontrolled hypertension, INR > 4.0, previous bleeding episodes, occult malignancies and in those with medication noncompliance and poor follow-up for coagulation monitoring. Anticoagulants should be avoided in patients unable to understand all aspects of treatment; have inadequate resources;or in whom the potential risk of bleeding outweighs the benefit of treatment.

New anticoagulants

A brief overview of anticoagulants and platelet-directed therapies under development, and their respective targets of inhibition is summarized in Figs. 5 and 6.

Anticoagulation in patients with malignancy

The association between thrombosis and cancer was established by the French physician Armand Trousseau in

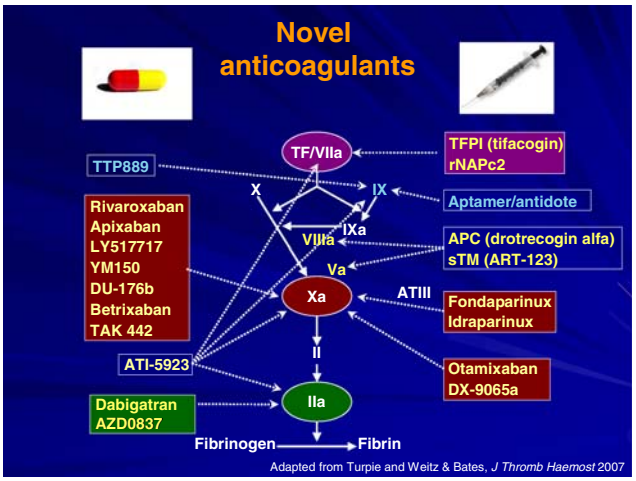


Fig. 5 Novel anticoagulants and targets of inhibition

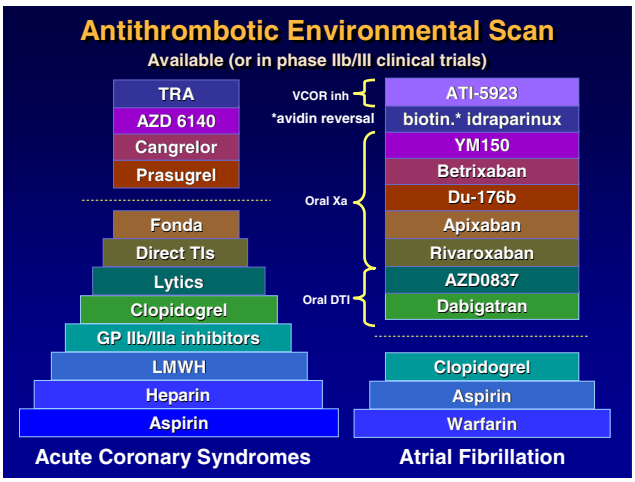


Fig. 6 Antithrombotic therapy for acute coronary syndrome and atrial fibrillation

1865, when he described thrombophlebitis or phlegmasia as a presenting sign of visceral malignancy. It is well established that tumor cells can secrete factors that initiate coagulation, including tissue factor and cancer procoagulant protein, which directly activates factor X. It has also been shown that "cryptic proteins" in the coagulation system, in the fibrinolytic system and secreted by platelets can affect angiogenesis, an essential process for tumor growth and metastasis.

VTE is a major complication of cancer and an important cause of morbidity and mortality. It has been estimated that VTE occurs in 4–20% of patients with cancer, and 14.3% of hospitalized cancer patients die as a direct result of pulmonary embolism (EP). Chemotherapy and hormonal treatment, particularly tamoxifen, increase the risk of VTE as does surgery.

Anticoagulant therapy is used in two classic conditions in cancer patients: for treatment of VTE episodes and as a prophylactic measure for hospitalized patients, particularly those who undergo surgery lasting more than 30 min. Although the American Society of Clinical Oncology does not recommend VTE prophylaxis for ambulatory cancer patients, several recent reports suggest that high-risk patients may in fact benefit from this approach.

Cancer patients frequently have long-term indwelling central venous catheters for administration of blood products, chemotherapy and parenteral nutrition. Catheter-related venous thrombosis is one of the most common complications, but routine antithrombotic prophylaxis is not recommended.

How to diagnose and treat heparin-induced thrombocytopenia

Heparin-induced thrombocytopenia (HIT) is a limb-and life-threatening immune-based disorder characterized by the formation of IgG antibodies against an antigenic complex consisting of heparin—a negatively-charged mucopolysaccharide and platelet factor 4 (PF4). Asymptomatic thrombocytopenia, defined as a platelet count either 50% or more below baseline or less than 150×10^9/l develops in 30–50% of patients who develop heparin-PF4 antibodies, In turn, thrombosis involving the arterial, venous and less often microcirculatory system occurs in 30–50% of patients with thrombocytopenia. In most instances, antibody production requires 3–5 days of daily heparin exposure; however, antibodies can develop within hours of exposure, particularly in patients with recent heparin treatment or weeks later—a condition known as delayed HIT.

Clinical suspicion is the key to diagnosis, with confirmation subsequently provided by documentation of heparin-PF4 antibodies using either a functional, ELISA-based or platelet serotonin release-determined assay.

The management of HIT must begin with complete cessation of all heparin products, followed by infusion of a direct thrombin inhibitor—lepirudin, argatroban or bivalirudin (for patients undergoing percutaneous coronary intervention). A vitamin K antagonist should *not* be instituted until a direct thrombin inhibitor is started and the platelet count has increased to baseline or to a level above 150×10^9/l. The recommended duration of vitamin K antagonist treatment is, at a minimum, 6 weeks and longer if a thrombotic condition dictates.

Future directions: pharmacogenomics

The majority of common diseases arise from interactions between innate and acquired genetic alterations, exposure

to varying environmental factors and life style. Accordingly, they are referred to as complex diseases. Common diseases such as cancer, cardiovascular disease and diabetes are examples of complex diseases. Variations in DNA sequence and gene expression, influenced by environmental factors, determine individual differences in susceptibility or protection to common diseases, as well as in the response to therapy. The development of drugs tailored specifically to the patient's genetic profile or "signature", minimizing adverse effects and maximizing treatment response, constitutes the overarching theme of pharmacogenomics.

The hereditary basis of individual variability for disease susceptibility and drug response were, for a long time, studied within the classic genetic paradigm, i.e. investigating polymorphisms or mutations in a particular gene and the co-segregation of these genes and the phenotype of interest along several generations. The sequencing and mapping of the human genome expanded the possibilities for studying genetic variability: 3 million genomic sites where individuals can differ by only one DNA nucleotide were identified and these variations, called single nucleotide polymorphisms or SNPs, have subsequently been associated to risk for or protection from several diseases. Through SNP analysis it was shown that polymorphisms in the genes CYP2C9 e VKORC1, which have an essential role in warfarin's metabolism and pharmacological profile, determine a patients' response to this oral anticoagulant. Whether knowledge of a patient's genotype will allow clinicians to reduce the rate of adverse events such as bleeding remains to be established through carefully designed clinical trials.

Acknowledgements The I International Symposium of Thrombosis and Anticoagulation in Internal Medicine was sponsored by an unrestricted grant from Sanofi-aventis. The symposium was organized by Meeting Eventos, and the chairmen of the meeting thank Cecilia Jahnel and Fernanda Prestes for all the support in the organization.

References

1. Hirsh J, Guyatt G, Albers GW, Harrington R, Schünemann HJ, American College of Chest Physicians (2008) Executive summary: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). Chest 133(6 Suppl): 71S–109S. doi:10.1378/chest.08-0693. No abstract available
2. Bonetti PO, Lerman LO, Lerman A (2003) Endothelial dysfunction: a marker of atherosclerotic risk. Arterioscler Thromb Vasc Biol 23(2):168–175. doi:10.1161/01.ATV.0000051384.43104.FC
3. Cai H, Harrison DG (2000) Endothelial dysfunction in cardiovascular diseases: the role of oxidant stress. Circ Res 87(10): 840–844
4. Alheid U, Frölich JC, Förstermann U (1987) Endothelium-derived relaxing factor from cultured human endothelial cells inhibits aggregation of human platelets. Thromb Res 47(5):561–571. doi:10.1016/0049-3848(87)90361-6
5. Libby P (2008) The molecular mechanisms of the thrombotic complications of atherosclerosis. J Intern Med 263(5):517–527. doi:10.1111/j.1365-2796.2008.01965.x
6. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G, American College of Chest Physicians (2008) Executive summary: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). Chest 133(6 Suppl): 160S–198S. doi:10.1378/chest.08-0670
7. Antman EM, Hand M, Armstrong PW, Bates ER, Green LA, Halasyamani LK, Hochman JS, Krumholz HM, Lamas GA, Mullany CJ, Pearle DL, Sloan MA, Smith SC Jr (2008) 2004 writing committee members, Anbe DT, Kushner FG, Ornato JP, Jacobs AK, Adams CD, Anderson JL, Buller CE, Creager MA, Ettinger SM, Halperin JL, Hunt SA, Lytle BW, Nishimura R, Page RL, Riegel B, Tarkington LG, Yancy CW. 2007 focused update of the ACC/AHA 2004 guidelines for the management of patients with ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines: developed in collaboration with the Canadian Cardiovascular Society endorsed by the American Academy of Family Physicians: 2007 writing group to review new evidence and update the ACC/AHA 2004 guidelines for the management of patients with ST-elevation myocardial infarction, writing on behalf of the 2004 writing committee. Circulation 117(2):296–329. doi:10.1161/CIRCULATIONAHA.107.188209. No abstract available
8. Anderson JL, Adams CD, Antman EM, Bridges CR, Califf RM, Casey DE Jr, Chavey WE II, Fesmire FM, Hochman JS, Levin TN, Lincoff AM, Peterson ED, Theroux P, Wenger NK, Wright RS, Smith SC Jr, Jacobs AK, Halperin JL, Hunt SA, Krumholz HM, Kushner FG, Lytle BW, Nishimura R, Ornato JP, Page RL, Riegel B, American College of Cardiology (2007) American Heart Association Task Force on Practice Guidelines (writing committee to revise the 2002 guidelines for the management of patients with unstable angina/non ST-elevation myocardial infarction); American College of Emergency Physicians; Society for Cardiovascular Angiography and Interventions; Society of Thoracic Surgeons; American Association of Cardiovascular and Pulmonary Rehabilitation; Society for Academic Emergency Medicine. ACC/AHA 2007 guidelines for the management of patients with unstable angina/non ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (writing committee to revise the 2002 guidelines for the management of patients with unstable angina/non ST-elevation myocardial infarction): developed in collaboration with the American College of Emergency Physicians, the Society for Cardiovascular Angiography and Interventions, and the Society of Thoracic Surgeons: endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation and the Society for Academic Emergency Medicine. Circulation 116(7):e148–e304. doi:10.1161/CIRCULATIONAHA.107.181940. No abstract available
9. Lopes RD, Pieper KS, Horton JR, Al-Khatib SM, Newby LK, Mehta RH, Van de Werf F, Armstrong PW, Mahaffey KW, Harrington RA, Ohman EM, White HD, Wallentin L, Granger CB (2008) Short- and long-term outcomes following atrial fibrillation in patients with acute coronary syndromes with or without ST-segment elevation. Heart 94(7):867–873. doi:10.1136/hrt.2007.134486
10. Lopes RD, Piccini JP, Hylek EM, Granger CB, Alexander JH (2008) Antithrombotic therapy in atrial fibrillation: guidelines translated for the clinician. J Thromb Thrombolysis 26(3):167–174. doi:10.1007/s11239-008-0272-4
11. Hylek EM, Becker RC (2008) Atrial fibrillation and stroke severity: expanding the mechanistic exemplar, clinical phenotype, and goals of anticoagulant pharmacotherapy. Ann Neurol 64(5):480–483. doi:10.1002/ana.21551

12. Tapson VF, Decousus H, Pini M, Chong BH, Froehlich JB, Monreal M, Spyropoulos AC, Merli GJ, Zotz RB, Bergmann JF, Pavanello R, Turpie AG, Nakamura M, Piovella F, Kakkar AK, Spencer FA, Fitzgerald G, Anderson FA Jr (2007) IMPROVE Investigators. Venous thromboembolism prophylaxis in acutely ill hospitalized medical patients: findings from the International Medical Prevention Registry on Venous Thromboembolism. Chest 132(3):936–945. doi:10.1378/chest.06-2993
13. Cohen AT, Tapson VF, Bergmann JF, Goldhaber SZ, Kakkar AK, Deslandes B, Huang W, Zayaruzny M, Emery L, Anderson FA Jr (2008) ENDORSE Investigators. Venous thromboembolism risk and prophylaxis in the acute hospital care setting (ENDORSE study): a multinational cross-sectional study. Lancet 371(9610):387–394. doi:10.1016/S0140-6736(08)60202-0
14. Sherman DG, Albers GW, Bladin C, Fieschi C, Gabbai AA, Kase CS, O'Riordan W, Pineo GF (2007) PREVAIL Investigators. The efficacy and safety of enoxaparin versus unfractionated heparin for the prevention of venous thromboembolism after acute ischaemic stroke (PREVAIL study): an open-label randomised comparison. Lancet 369(9570):1347–1355. doi:10.1016/S0140-6736(07)60633-3
15. Spencer FA, Lessard D, Emery C, Reed G, Goldberg RJ (2007) Venous thromboembolism in the outpatient setting. Arch Intern Med 167(14):1471–1475. doi:10.1001/archinte.167.14.1471
16. Rocha AT, Paiva EF, Lichtenstein A, Milani R Jr, Cavalheiro CF, Maffei FH (2007) Risk-assessment algorithm and recommendations for venous thromboembolism prophylaxis in medical patients. Vasc Health Risk Manag 3(4):533–553. Review
17. Eriksson BI, Borris LC, Friedman RJ, Haas S, Huisman MV, Kakkar AK, Bandel TJ, Beckmann H, Muehlhofer E, Misselwitz F, Geerts W (2008) RECORD1 Study Group. Rivaroxaban versus enoxaparin for thromboprophylaxis after hip arthroplasty. N Engl J Med 358(26):2765–2775. doi:10.1056/NEJMoa0800374
18. Kakkar AK, Brenner B, Dahl OE, Eriksson BI, Mouret P, Muntz J, Soglian AG, Pap AF, Misselwitz F, Haas S (2008) RECORD2 Investigators. Extended duration rivaroxaban versus short-term enoxaparin for the prevention of venous thromboembolism after total hip arthroplasty: a double-blind, randomised controlled trial. Lancet 372(9632):31–39. doi:10.1016/S0140-6736(08)60880-6
19. Lassen MR, Ageno W, Borris LC, Lieberman JR, Rosencher N, Bandel TJ, Misselwitz F, Turpie AG (2008) RECORD3 Investigators. Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty. N Engl J Med 358(26):2776–2786. doi:10.1056/NEJMoa076016
20. Garcia DA, Regan S, Henault LE, Upadhyay A, Baker J, Othman M, Hylek EM (2008) Risk of thromboembolism with short-term interruption of warfarin therapy. Arch Intern Med 168(1):63–69. doi:10.1001/archinternmed.2007.23
21. Weitz JI, Hirsh J, Samama MM, American College of Chest Physicians (2008) New antithrombotic drugs: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). Chest 133(6 Suppl):234S–256S. Erratum in: Chest. 2008 Aug;134. 2:473. doi:10.1378/chest.08-0673
22. Douketis JD, Berger PB, Dunn AS, Jaffer AK, Spyropoulos AC, Becker RC, Ansell J, American College of Chest Physicians (2008) The perioperative management of antithrombotic therapy: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). Chest 133(6 Suppl):299S–339S. doi:10.1378/chest.08-0675
23. Salem DN, O'Gara PT, Madias C, Pauker SG, American College of Chest Physicians (2008) Valvular and structural heart disease: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). Chest 133(6 Suppl):593S–629S. doi:10.1378/chest.08-0724
24. Warkentin TE, Greinacher A, Koster A, Lincoff AM, American College of Chest Physicians (2008) Treatment and prevention of heparin-induced thrombocytopenia: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). Chest 133(6 Suppl):340S–380S. doi:10.1378/chest.08-0677
25. Lyman GH, Khorana AA, Falanga A, Clarke-Pearson D, Flowers C, Jahanzeb M, Kakkar A, Kuderer NM, Levine MN, Liebman H, Mendelson D, Raskob G, Somerfield MR, Thodiyil P, Trent D, Francis CW, American Society of Clinical Oncology (2007) American Society of Clinical Oncology guideline: recommendations for venous thromboembolism prophylaxis and treatment in patients with cancer. J Clin Oncol 25(34):5490–5505. doi:10.1200/JCO.2007.14.1283. Review
26. Gage BF, Lesko LJ (2008) Pharmacogenetics of warfarin: regulatory, scientific, and clinical issues. J Thromb Thrombolysis 25(1):45–51. doi:10.1007/s11239-007-0104-y. Review

American Journal of Epidemiology
Vol. 169, No. 7
DOI: 10.1093/aje/kwn403
Advance Access publication January 29, 2009

Original Contribution

Life-Course Socioeconomic Position and Incidence of Coronary Heart Disease

The Framingham Offspring Study

Eric B. Loucks, John W. Lynch, Louise Pilote, Rebecca Fuhrer, Nisha D. Almeida, Hugues Richard, Golareh Agha, Joanne M. Murabito, and Emelia J. Benjamin

Initially submitted June 30, 2008; accepted for publication December 5, 2008.

Cumulative exposure to socioeconomic disadvantage across the life course may be inversely associated with coronary heart disease (CHD); the mechanisms are not fully clear. An objective of this study was to determine whether cumulative life-course socioeconomic position (SEP) is associated with CHD incidence in a well-characterized US cohort that had directly assessed childhood and adulthood measures of SEP and prospectively measured CHD incidence. Furthermore, analyses aimed to evaluate whether adjustment for CHD risk factors reduces the association between cumulative life-course SEP and CHD. The authors examined 1,835 subjects who participated in the Framingham Heart Study Offspring Cohort from 1971 through 2003 (mean age, 35.0 years; 52.4% women). Childhood SEP was measured as father's education; adulthood SEP was assessed as own education and occupation. CHD incidence included myocardial infarction, coronary insufficiency, and coronary death. Cox proportional hazards analyses indicated that cumulative SEP was associated with incident CHD after adjustment for age and sex (hazard ratio = 1.82, 95% confidence interval: 1.17, 2.85 for low vs. high cumulative SEP score). Adjustment for CHD risk factors reduced that magnitude of association (hazard ratio = 1.29, 95% confidence interval: 0.78, 2.13). These findings underscore the potential importance of CHD prevention and treatment efforts for those whose backgrounds include low SEP throughout life.

cohort studies; coronary disease; myocardial ischemia; social class; socioeconomic factors

Abbreviations: CHD, coronary heart disease; CI, confidence interval; HDL, high density lipoprotein; HR, hazard ratio; SEP, socioeconomic position.

Coronary heart disease (CHD) remains a major cause of mortality in the United States and worldwide, responsible for 10% of the disability-adjusted life years lost in developing countries and 18% in developed countries (1, 2). Strong inverse socioeconomic gradients in CHD exist in many developed countries, where adulthood socioeconomic position (SEP) is typically measured as participants' own education, occupation, and income (3, 4). Evidence is fairly consistent that childhood SEP (often measured as parents' occupation or education) is also inversely associated with CHD in developed countries (5, 6). There also tend to be socioeconomic gradients in the expected directions for CHD risk factors including smoking, diabetes, blood pressure, cholesterol, and, for women, obesity (3, 7–10).

To better understand how SEP may influence CHD, it is informative to conceptualize SEP across the life course (11–13). People experience a certain set of socioeconomic circumstances at every phase of their lives; each period may theoretically influence the course of chronic disease. The ''accumulation-of-risk'' SEP framework focuses on the total amount of (i.e., cumulative) exposure to socioeconomic disadvantage (11, 12). Initial evidence suggests that, in a number of studies in Europe using case-control designs (14) or nationally available death records (13, 15–18), cumulative SEP is inversely associated with cardiovascular disease. Less is known about the association of cumulative SEP with incident CHD in the United States. Furthermore, little is known about whether specific CHD risk factors may be

Correspondence to Dr. Eric B. Loucks, Douglas Mental Health Institute, McGill University, 6875 LaSalle Blvd., Montreal, QC, Canada H4H 1R3 (e-mail: eric.loucks@mcgill.ca).

particularly important in explaining life-course socioeconomic gradients in CHD.

An objective of this study was to determine whether cumulative life-course SEP is associated with CHD incidence in a well-characterized US cohort (The Framingham Offspring Study) that had directly assessed childhood and adulthood measures of SEP and prospectively measured CHD incidence. Furthermore, analyses aimed to evaluate whether adjustment for CHD risk factors reduces the association between cumulative life-course SEP and CHD. Exploratory analyses further evaluated whether any specific CHD risk factors (e.g., smoking, systolic blood pressure, cholesterol, fasting glucose, body mass index) may be particularly important explanatory mechanisms for the association of life-course SEP with CHD incidence.

MATERIALS AND METHODS

Study sample

The Framingham Heart Study is a community-based, longitudinal, observational cohort study initiated in 1948 to prospectively investigate risk factors for CHD. The Framingham Offspring Study began in 1971 with recruitment of 5,124 men and women who were offspring (or offspring's spouses) of the Original Cohort of the Framingham Heart Study. The design and selection criteria of the Framingham Offspring Study have been described elsewhere (19). At each examination visit, participants underwent medical history, physical examination, anthropometry, and laboratory assessment of CHD risk factors, as previously described (19). Framingham participants signed informed consent, and the Framingham Study is reviewed annually by the Boston University Medical Center Institutional Review Board.

There were 5,124 participants who completed Offspring Study examination 1 (during 1971–1975), and 4,989 agreed for their data to be in the open-access data set. Of these, 2,136 had no father in the Original Cohort of the Framingham Heart Study and hence were excluded from analyses, leaving 2,853 participants. Of these participants, 119 had fathers whose education variable was missing. A further 818 were missing their own education or occupation variables (150 died between examinations 1 and 2, 509 did not attend examination 2 or 3, and 159 did not answer the education/occupation question), leaving 1,916 participants. We further restricted participants to those aged $\geq$28 years at the time their own educational attainment and occupation were measured, and we excluded 21 participants with baseline CHD events, resulting in a final sample of 1,835 participants. Analyses on excluded ($n = 3{,}154$) versus included participants found that excluded participants were more likely to be older (age 36.9 vs. 35.0 years, respectively; $P < 0.0001$), to have slightly higher fasting glucose levels (103 mg/dL vs. 101 mg/dL, $P < 0.0001$), and to be taking antihypertensive medications (3.7% vs. 2.5% of participants, $P = 0.03$). Included and excluded participants were similar regarding other variables including sex, cigarette smoking, body mass index, systolic blood pressure, total:high density lipoprotein (HDL) cholesterol ratio, cholesterol-lowering medication use, and incident CHD events.

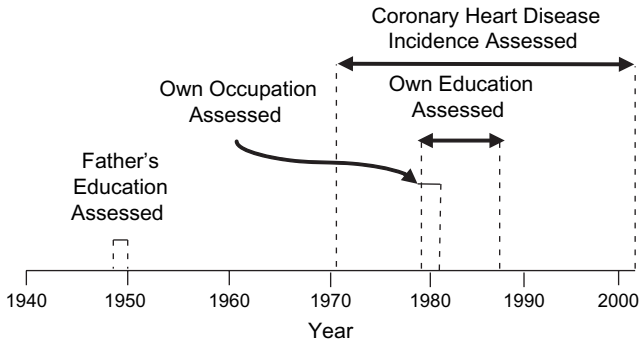


Figure 1. Time line of assessments for exposures (father's education, own occupation, own education) and outcome (coronary heart disease incidence). Covariates were assessed at examination 1 of the Framingham Heart Study Offspring Cohort (1971–1975), Framingham, Massachusetts. Enrollment and initiation of examination 1 for the Original Cohort of the Framingham Heart Study took place during 1948–1950. Enrollment and initiation of examination 1 for the Offspring Cohort took place during 1971–1975. Initiation of Offspring Cohort examination 2 took place during 1979–1982. Initiation of Offspring Cohort examination 3 occurred during 1984–1987.

Childhood SEP: father's education

Childhood SEP was measured by father's educational attainment, obtained directly from Offspring Study cohort participants' fathers who were enrolled in the Framingham Heart Study Original Cohort. Father's educational level was measured at enrollment between 1948 and 1950 when their mean age was 44 years (range: 28–62) (Figure 1). Father's education was ascertained directly from the father as a 6-category variable: $\leq$eighth grade, some high school (i.e., did not graduate from high school), high school graduate, some college (i.e., did not graduate from college), college graduate, and a final category including postgraduate school, business college, nursing school, music school, and art school. For analyses, father's education was categorized into 3 groups: <high school, high school, and >high school.

We explored the use of mother's education as a measure of childhood SEP. Mother's educational attainment was not associated with incident CHD in this cohort (hazard ratio (HR) = 1.25, 95% confidence interval (CI): 0.84, 1.90 for mother's education <high school vs. >high school, after adjusting for age and sex). The vast majority of mothers (84%) in this cohort had the same occupation (homemaker), likely because of the historical time period when Offspring Study participants were children (approximately during the 1930s–1950s), when it was less common for mothers to work outside the home. Consequently, father's education was used as a measure of childhood SEP.

Adulthood SEP: own education and occupation

Own education was measured directly from Framingham Offspring Study participants at examination 3 (1984–1987);

if examination 3 education was missing, the examination 2 assessment (1979–1982) was used (Figure 1). Education was available in 6 categories of years of education: 0–4, 5–8, 9–11, 12, 13–16, ≥17. For analyses, own education was collapsed into 3 groups: ≤12, 13–16, and ≥17 years of education. Own occupation was measured at examination 2 (1979–1982) by asking what kind of work the participants do (or did), categorized as professional, executive, supervisory, technical, laborer, clerical, sales, and housewife. To obtain higher levels of education or occupation, participants were restricted to those aged ≥28 years when educational attainment and occupation were measured to allow 10 years from likely completion of high school (at age 18 years on average). Sensitivity analyses using data on participants aged ≥40 years were performed, as described below.

Accumulation-of-risk SEP framework

Analyses utilizing 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/homemaker = 1, executive/professional/supervisory/technical = 2). Higher cutpoints were used for educational categories of Offspring Study compared with Original Cohort fathers to account for secular trends of increased normative levels of education across generations.

Coronary heart disease

At all examinations, participants underwent standardized physician-administered medical history assessments. All participants suspected of experiencing stroke were invited back for a detailed neurologic examination. Hospital and nursing home records as well as outside medical office records were routinely sought for all cardiovascular events and all deaths. In addition, Framingham Study personnel surveyed the only hospital in town daily for participant emergency room visits and hospitalizations. Suspected cardiovascular disease events and deaths were reviewed by a panel of 3 investigators, who examine all relevant available data (Framingham Study clinic data; outside medical, nursing home, and hospitalization records) and make event determinations by using previously published criteria (20).

CHD incidence was identified as occurring in participants diagnosed since onset of the Framingham Offspring Study (1971–1975) until 2003 (Figure 1). Clinically validated CHD events included myocardial infarction, coronary insufficiency, and coronary death (sudden and nonsudden). Secondary analyses investigated the outcome cardiovascular disease, which included clinically validated measures of myocardial infarction, coronary insufficiency, cerebrovascular events (including cerebral embolism, intracerebral hemorrhage, subarachnoid hemorrhage, and other cerebrovascular accident), heart failure requiring hospitalization, and death due to the aforementioned outcomes.

Covariates

Risk factors were measured at baseline, Offspring Study examination 1 (1971–1975) (Figure 1). Cigarette smoking was determined by self-report and was defined as smoking regularly in the year prior to the examination. Systolic blood pressure was calculated as the average of the clinic physician's 2 seated systolic blood pressure measurements. Body mass index was calculated as weight in kilograms divided by the square of height in meters (kg/m^2). Body weight was measured to the nearest 1 pound with a standing beam balance and with subjects wearing examination robes and undergarments. This measurement was then converted to kilograms (1 pound = 0.454 kg). Height was measured with the use of a stadiometer (to the nearest 0.25 inch and was then converted to meters (1 inch = 0.025 m). Fasting glucose was measured with a hexokinase reagent kit (A-gent glucose test; Abbott, South Pasadena, California). Glucose assays were run in duplicate, and the intraassay coefficient of variation ranged from 2% to 3% depending on the assayed glucose concentration. HDL and total cholesterol concentrations were measured by automated enzymatic techniques (21). Medication use was self-reported.

Statistical analyses

Sex- and age-adjusted descriptive statistics (predicted means and percent prevalences) were generated for CHD and CHD risk factors (systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, body mass index, cigarette smoking, and antihypertensive medication use) according to father's education, own education, and own occupation.

Cox proportional hazards analyses evaluated the association of SEP with incidence of CHD. Secondary analyses used cardiovascular disease instead of CHD as the outcome. Analyses were adjusted for potential confounders including age and sex, as well as for CHD risk factors (described above). Cholesterol medication use was not included in analyses because only 5 participants used these medications at baseline. Pearson correlation coefficients and variance inflation were used to evaluate collinearity, which found that systolic and diastolic blood pressure variables were highly collinear ($r = 0.81$). The 3 SEP variables (father's education, own education, and own occupation) were found to have minimal variance inflation and were not correlated highly enough to be of concern to simultaneously adjust for all 3 in a single multivariable model (correlation coefficients ranged from 0.27 to 0.51). Consequently, all 3 measures of SEP were simultaneously adjusted for in analyses to evaluate whether any specific SEP measures contributed more strongly to CHD risk.

Marginal hazards models were run by using the procedure PHREG in SAS version 9.1 software (SAS Institute, Inc., Cary, North Carolina) with option COVSANDWICH to account for clustering of outcomes by family. Sex-specific analyses were underpowered and could not be conducted because only 44 CHD events in females and 100 events in males occurred. Formal interaction tests did not show evidence of effect modification by sex. Consequently, data for males and females were pooled in analyses. Power analyses

were performed by using the computer program PS: Power and Sample Size Calculations, version 2.1.31 (Vanderbilt Medical Center, Nashville, Tennessee) according to criteria reported by Dupont and Plummer (22).

Sensitivity analyses were performed on a sample further restricted to participants ≥40 years of age (rather than ≥28 years of age) at baseline to assess associations between occupation and CHD incidence among participants who had more time to attain higher occupational levels. Further sensitivity analyses investigated associations between cumulative life-course SEP and CHD incidence during the time frame after which all SEP measures were obtained (1988–2003 rather than 1971–2003) (Figure 1).

RESULTS

The Framingham Heart Study Offspring participants included in the present study were a mean age of 35.0 years at baseline, and 52.4% were women. The age range at examination 1 (1971–1975) was 19–62 years. Father's education was inversely associated with several CHD risk factors, including smoking, body mass index, systolic blood pressure, total:HDL cholesterol ratio, and fasting glucose. Own education was inversely associated with smoking, systolic blood pressure, and total:HDL cholesterol ratio. Own occupation was inversely related to smoking and body mass index. Furthermore, unequal proportions of females were represented in the own education and occupation categories (Table 1).

Age- and sex-adjusted Cox proportional hazards models showed that cumulative SEP across the life course was inversely associated with CHD incidence (HR = 1.82, 95% CI: 1.17, 2.85 for low vs. high cumulative SEP score) (Table 2). Further adjustment for CHD risk factors reduced the association (HR = 1.29, 95% CI: 0.78, 2.13). In exploratory analyses adjusting for individual CHD risk factors, smoking was associated with the greatest reduction in the point estimate (age-, sex-, and smoking-adjusted HR = 1.43, 95% CI: 0.90, 2.27). Adjusting for body mass index, systolic blood pressure, and HDL:total cholesterol ratio also somewhat reduced the effect size; adjusting for fasting glucose and antihypertensive medication use had little effect on the strength of association between cumulative SEP and CHD incidence (Table 2).

In an effort to provide information regarding whether any of the 3 subcomponents of the cumulative SEP score may be contributing particularly strongly to the gradient between cumulative SEP and CHD incidence, we analyzed the individual association of father's education, own education, and own occupation with CHD incidence. Father's education and own education were inversely associated with CHD incidence after adjusting for age and sex (HR = 1.65, 95% CI: 1.02, 2.66 for father's education <high school vs. >high school, and HR = 1.85, 95% CI: 1.05, 3.27 for own education ≤12 years vs. ≥17 years; Table 3). Further adjustment for other SEP measures reduced the estimated effect sizes only a small amount for both father's education and own education (HR = 1.53, 95% CI: 0.92, 2.55 for participants whose father's education was <high school vs. >high school, and HR = 1.62, 95% CI: 0.85, 3.09 for participants whose own education was ≤12 years vs. ≥17 years). However, the 95% confidence intervals encompassed both a null effect (i.e., HR = 1.0) and a large effect (e.g., HR = 2.5), indicating that the statistical power was low ($1 - \beta = 0.43$ for a cumulative 30-year incidence of CHD of 7.5% vs. 5% for father's education <high school ($n = 958$) vs. father's education >high school ($n = 444$), with $\alpha = 0.05$). Occupation was not associated with CHD incidence (Table 3).

We conducted a series of secondary analyses. Analyses were repeated by using cardiovascular disease instead of CHD as an outcome (Web Tables 1 and 2; this information is described in the first 2 of 3 supplementary tables, all of which are posted on the *Journal*'s website (http://aje.oupjournals.org/)). Findings were generally similar when either outcome was used. Sensitivity analyses were performed to assess the association of occupation with CHD and cardiovascular disease incidence among participants aged ≥40 years rather than aged ≥28 years, as used in the analyses described above. Effect sizes were similar in both sets of analyses (Web Table 3). Because own education (1979–1987) and occupation (1979–1982) were assessed after commencement of CHD incidence measurements (1971–2003), we performed sensitivity analyses restricted to the years after completion of all SEP measures (1988–2003) to evaluate whether the timing of SEP and CHD measures had an impact on the findings. Analyses showed similar socioeconomic gradients in CHD (HR for high vs. low cumulative SEP score = 1.94, 95% CI: 1.19, 3.17 after adjusting for age and sex).

DISCUSSION

Evidence from our study supported an inverse association of cumulative life-course SEP with CHD incidence. Further adjustment for CHD risk factors substantially attenuated the strength of association.

Prior literature

With regard to adulthood socioeconomic disparities in CHD, a systematic review showed consistent inverse associations between adulthood SEP and CHD in developed nations since the 1970s (4). A separate systematic review reported inverse associations of childhood SEP with risk of cardiovascular disease in 31 of 40 studies (23). The accumulation-of-risk SEP framework suggests that as the duration and severity of socioeconomic disadvantage increase, resulting cumulative damage could place individuals at higher risk of CHD (12). Our study provides evidence to support this hypothesis, in that higher life-time exposure to socioeconomic deprivation was associated with increased risk of CHD. Adjusting for CHD risk factors reduced the strength of association and rendered it nonsignificant, which was not unexpected because CHD risk factors are candidate pathways by which SEP may influence CHD. For life-course SEP, a systematic review found consistently inverse associations between the accumulation-of-risk SEP framework and risk of cardiovascular disease (13), which is in agreement with our findings. In the latter review, the associations of early-life SEP (independent of adulthood SEP)

Table 1. Age- and Sex-adjusted Baseline Characteristics of Participants According to Life-Course Socioeconomic Position, Framingham Heart Study Offspring Cohort, United States, 1971–1975[a]

	Father's Education		
	<High School (n = 958)	High School (n = 433)	>High School (n = 444)
Age, years[b]	37.3 (36.8, 37.9)	31.8 (31.0, 32.6)	33.1 (32.3, 33.9)
Sex (% female)[b]	52.3	55.0	50.2
Body mass index, kg/m^2	25.5 (25.2, 25.7)	25.0 (24.6, 25.4)	24.6 (24.2, 25.0)
Systolic blood pressure, mm Hg	122 (121, 123)	122 (121, 124)	120 (119, 122)
Total:HDL cholesterol ratio	4.3 (4.2, 4.4)	4.1 (4.0, 4.2)	4.0 (3.9, 4.1)
Fasting glucose, mg/dL	102 (101, 102)	101 (100, 102)	100 (99, 102)
Taking antihypertensive medication, %	1.6 (1.0, 2.7)	2.2 (1.2, 4.2)	0.5 (0.2, 1.6)
Current smoker, %	45.7 (42.5, 49.0)	44.9 (40.2, 49.7)	36.7 (32.3, 41.4)
	Own Education		
	≤12 years (n = 741)	13–16 years (n = 777)	≥17 years (n = 317)
Age, years[b]	37.0 (36.3, 37.6)	34.1 (33.5, 34.7)	32.5 (31.7, 33.4)
Sex (% female)[b]	57.9	55.3	32.9
Body mass index, kg/m^2	25.3 (25.0, 25.6)	25.1 (24.8, 25.3)	25.0 (24.6, 25.4)
Systolic blood pressure, mm Hg	122 (121, 123)	122 (121, 123)	120 (119, 122)
Total:HDL cholesterol ratio	4.2 (4.1, 4.3)	4.2 (4.1, 4.3)	3.9 (3.8, 4.1)
Fasting glucose, mg/dL	102 (101, 103)	100 (100, 101)	101 (100, 103)
Taking antihypertensive medication, %	1.5 (0.9, 2.6)	1.5 (0.9, 2.6)	1.3 (0.5, 3.2)
Current smoker, %	51.6 (47.9, 55.3)	43.4 (40.0, 47.0)	24.2 (19.8, 29.2)
	Own Occupation		
	Laborer (n = 401)	Housewife/ Clerical/Sales (n = 775)	Supervisory/Technical/ Professional/Executive (n = 659)
Age, years[b]	36.9 (36.0, 37.8)	35.5 (34.8, 36.1)	33.3 (32.6, 33.9)
Sex (% female)[b]	21.7	85.2	32.6
Body mass index, kg/m^2	25.7 (25.3, 26.1)	25.0 (24.7, 25.4)	25.0 (24.7, 25.3)
Systolic blood pressure, mm Hg	122 (120, 123)	122 (121, 123)	121 (120, 126)
Total:HDL cholesterol ratio	4.2 (4.0, 4.3)	4.2 (4.1, 4.3)	4.1 (4.0, 4.2)
Fasting glucose, mg/dL	101 (100, 103)	101 (100, 102)	102 (101, 103)
Taking antihypertensive medication, %	1.8 (0.9, 3.4)	1.2 (0.6, 2.2)	1.7 (0.9, 3.0)
Current smoker, %	52.3 (47.1, 57.5)	44.1 (40.2, 48.1)	36.9 (33.2, 40.9)

Abbreviation: HDL, high density lipoprotein.
[a] Data are expressed as predicted mean value or percent prevalence (95% confidence interval).
[b] Calculated by using univariate analyses.

and social mobility were less consistently associated with measures of cardiovascular disease (13).

Potential mechanisms

Cumulative life-course SEP may influence CHD through a number of mechanisms. Low adulthood SEP is typically inversely associated with many risk factors for CHD, such as smoking (7), blood pressure (8), diabetes (9), and, for women, obesity (10). Childhood SEP has been shown to be inversely associated with several CHD risk factors in adulthood, including smoking, blood pressure, cholesterol, and adiposity (24–26). Other potential mechanisms not measured in this study, including depression and stress, were not able to be evaluated.

Few if any studies on the association of cumulative SEP with CHD adjusted for CHD risk factors separately (13–18); consequently, little is known about which risk factors may be particularly important in explaining socioeconomic gradient in CHD. Note that methodological biases can be induced by adjusting for mediators; therefore, these mechanistic findings should be interpreted with caution (27).

In our study, adjusting for smoking reduced the strength of association between life-course SEP and CHD. Point estimates for the association of cumulative SEP with CHD incidence were lower after adjusting for smoking than for

Table 2. Cox Proportional Hazards Ratios Demonstrating the Association of Cumulative Life-Course Socioeconomic Position Score With Incidence of Coronary Heart Disease, Framingham Offspring Study, United States, 1971–2003

Cumulative SEP Score[a]	No.	No. of Events	Model Adjustment[b]															
			Baseline		Cigarette Smoking		Body Mass Index		Systolic Blood Pressure		Total:HDL Cholesterol Ratio		Fasting Glucose		Antihypertensive Medication Use		All CHD Risk Factors[c]	
			HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
0 or 1	573	66	1.82	1.17, 2.85	1.43	0.90, 2.27	1.75	1.10, 2.79	1.71	1.09, 2.69	1.70	1.07, 2.71	1.85	1.17, 2.93	1.82	1.17, 2.85	1.29	0.78, 2.13
2 or 3	647	48	1.62	1.01, 2.61	1.45	0.90, 2.32	1.60	0.99, 2.59	1.56	0.97, 2.49	1.49	0.92, 2.42	1.67	1.03, 2.70	1.62	1.01, 2.61	1.38	0.85, 2.25
4 to 6	615	30	1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00	

Abbreviations: CHD, coronary heart disease; CI, confidence interval; HDL, high density lipoprotein; HR, hazard ratio; SEP, socioeconomic position.

[a] Analyses used a cumulative SEP score including father's education, own education, and own occupation. Scores were calculated for each SEP measure separately and then summed (range, 0–6): father's education: <high school = 0, high school = 1, >high school = 2; own education: ≤12 years = 0, 13–16 years = 1, ≥17 years = 2; own occupation: laborer = 0, clerical/sales/homemaker = 1, executive/professional/supervisory/technical = 2.

[b] All models were adjusted for age and sex.

[c] CHD risk factors include smoking, body mass index, systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, and antihypertensive medication use.

other CHD risk factors. However, given the wide 95% confidence intervals for these point estimates, it was not possible to ascertain whether smoking was more important than other CHD risk factors in reducing the association. We found strong inverse socioeconomic gradients in smoking in our study. Some studies on socioeconomic gradients in CHD, using other measures of SEP (only adulthood SEP or only childhood SEP), adjusted individually for smoking. For example, in a study on male physicians (who consequently had similar adulthood SEP), low childhood SEP was associated with a 2.40 (95% CI: 1.21, 4.74) higher relative risk of developing CHD before the age of 50 years. Adjustment for smoking only slightly and nonsignificantly reduced the association (relative risk = 2.24, 95% CI: 1.11, 4.51) (28). In the Whitehall II study, the magnitude of association of occupational class with CHD was nonsignificantly reduced by 18% after adjusting for smoking. Adjusting individually for hypertension, high cholesterol, and diabetes reduced the effect sizes by 14%, 3%, and 6%, respectively (29).

In a recent study on adulthood (that did not include childhood) socioeconomic gradients in mortality in 22 European countries, smoking-related conditions accounted for 22% and 6% of the socioeconomic gradient in the all-cause death rate among men and women, respectively (30). In another study in the United States and 11 European countries, adulthood (the study did not include childhood) socioeconomic gradients in cardiovascular disease mortality were highly associated with socioeconomic gradients in cigarette smoking and excessive alcohol consumption, unlike overweight, moderate alcohol consumption, and lack of fresh vegetables, which were not strongly associated with socioeconomic gradients in cardiovascular disease mortality (31). Overall, studies using SEP measures other than cumulative SEP provide limited evidence to suggest that smoking is a particularly important risk factor in explaining the socioeconomic gradients in CHD. Replication of findings using cumulative SEP is needed in other study samples to better ascertain the role of smoking in explaining cumulative socioeconomic gradients in CHD.

Strengths and limitations

Strengths of our study include that childhood SEP was directly assessed from parents. A review found that those studies that measured SEP in childhood showed stronger associations between childhood SEP and disease outcomes compared with studies that measured adult recall of childhood SEP, probably because of reductions in measurement error (32). Furthermore, the measures of CHD used only clinically validated outcomes. CHD risk factors were routinely and directly assessed by using measures with good validity and reliability.

Weaknesses of this study include the relatively small sample size (n = 1,835) compared with larger studies; consequently, we had lower statistical power. Furthermore, this study included a community-based population of individuals of European descent (representing the demographics of the city of Framingham, Massachusetts, at study onset) residing in the northeastern United States; thus, generalizability of results to other communities, races, and ethnicities is

Table 3. Cox Proportional Hazards Ratios for the Association of Socioeconomic Position With Incidence of Coronary Heart Disease, Framingham Offspring Study, United States, 1971–2003

SEP Measure and SEP Level	No. of Events	Model Adjustment					
		Age, Sex		Age, Sex, Other SEP Measures[a]		Age, Sex, CHD Risk Factors[b]	
		HR	95% CI	HR	95% CI	HR	95% CI
Father's education							
<High school	97	1.65	1.02, 2.66	1.53	0.92, 2.55	1.35	0.81, 2.23
High school	26	1.56	0.88, 2.77	1.50	0.84, 2.69	1.49	0.83, 2.69
>High school	21	1.00		1.00		1.00	
Own education							
≤12 years	71	1.85	1.05, 3.27	1.63	0.86, 3.11	1.20	0.66, 2.12
13–16 years	58	1.81	1.02, 3.21	1.76	0.98, 3.17	1.31	0.71, 2.40
>16 years	15	1.00		1.00		1.00	
Own occupation							
Laborer	50	1.20	0.80, 1.78	0.98	0.62, 1.55	0.90	0.59, 1.36
Homemaker, clerical, or sales	44	1.02	0.64, 1.64	0.92	0.57, 1.49	0.84	0.52, 1.36
Professional, executive, supervisory, or technical	50	1.00		1.00		1.00	

Abbreviations: CHD, coronary heart disease; CI, confidence interval; HR, hazard ratio; SEP, socioeconomic position.

[a] "Other SEP measures" refers to adjustment for measures of SEP other than the exposure of interest. For example, analyses of father's education were adjusted for own education and own occupation.

[b] CHD risk factors include smoking, body mass index, systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, and antihypertensive medication use.

uncertain. A limitation of the accumulation-of-risk SEP framework, as described by Pollitt et al. (13), is that cumulative life-course SEP measurements conflate SEP measures at specific times in the life course (e.g., SEP in early, middle, and late life); therefore, it is unknown which time period may be particularly important in influencing health. Furthermore, cumulative SEP measures typically give equal weighting to each subcomponent of SEP, which may not reflect true contributions of SEP to health. To provide information on relative contributions of each subcomponent of the cumulative SEP index to CHD incidence, we provided point estimates for each subcomponent. These analyses suggested that measures of father's education and own education were more important risk factors for CHD incidence than own occupation, at least as measured in this cohort.

In summary, this study found that directly assessed cumulative SEP across the life course was inversely associated with incident CHD in Framingham Offspring Study participants. Adjustment for CHD risk factors reduced the magnitude of association. These findings underscore the potential importance of CHD prevention and treatment efforts for those whose backgrounds include low SEP throughout their life course.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada (Eric B. Loucks, John W. Lynch, Rebecca Fuhrer, Nisha D. Almeida, Golareh Agha); Department of Psychiatry, Douglas Mental Health Institute, McGill University, Montreal, Canada (Eric B. Loucks, Nisha D. Almeida); School of Health Sciences, University of South Australia, South Australia, Australia (John W. Lynch); Department of Medicine, McGill University, Montreal, Canada (Louise Pilote, Hugues Richard); Boston University and National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts (Joanne M. Murabito, Emelia J. Benjamin); Departments of Medicine and Preventive Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts (Joanne M. Murabito, Emelia J. Benjamin); Cardiology Department and Whitaker Cardiovascular Institute, School of Medicine, Boston University, Boston, Massachusetts (Emelia J. Benjamin); and Department of Epidemiology, School of Public Health, Boston University, Boston, Massachusetts (Emelia J. Benjamin). Joanne M. Murabito and Emelia J. Benjamin contributed equally to the study.

The Framingham Heart Study is supported by National Institutes of Health (NIH)/National Heart, Lung, and Blood Institute (NHLBI) contract N01-HC-25195. The research was supported by Canadian Institutes of Health operating grant MOP81239 (E. B. L.); Canadian Institutes of Health Interdisciplinary Capacity Enhancement grant HOA80072 (J. W. L.); NIH grants HL064753, HL076784, and AG028321 (E. J. B.); and a Canadian Institutes of Health New Investigator Award (E. B. L.).

The authors are grateful to the NHLBI for providing them with the limited access data set, in conjunction with

Framingham Heart Study investigators who provided other important variables for analyses.

Conflict of interest: none declared.

REFERENCES

1. Kung HC, Hoyert DL, Xu J, et al. Deaths: final data for 2005. *Natl Vital Stat Rep*. 2008;56(10):1–120.
2. Mackay J, Mensah G. *The Atlas of Heart Disease and Stroke*. Geneva, Switzerland: World Health Organization; 2004.
3. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation*. 1993; 88(4 pt 1):1973–1998.
4. González MA, Rodríguez Artalejo F, Calero JR. Relationship between socioeconomic status and ischaemic heart disease in cohort and case-control studies: 1960–1993. *Int J Epidemiol*. 1998;27(3):350–358.
5. Galobardes B, Lynch JW, Davey Smith G. Childhood socioeconomic circumstances and cause-specific mortality in adulthood: systematic review and interpretation. *Epidemiol Rev.* 2004;26:7–21.
6. Galobardes B, Lynch JW, Smith GD. Is the association between childhood socioeconomic circumstances and cause-specific mortality established? Update of a systematic review. *J Epidemiol Community Health*. 2008;62(5):387–390.
7. Gilman SE, Martin LT, Abrams DB, et al. Educational attainment and cigarette smoking: a causal association? *Int J Epidemiol*. 2008;37(3):615–624.
8. Colhoun HM, Hemingway H, Poulter NR. Socio-economic status and blood pressure: an overview analysis. *J Hum Hypertens*. 1998;12(2):91–110.
9. Maty SC, Everson-Rose SA, Haan MN, et al. Education, income, occupation, and the 34-year incidence (1965–99) of type 2 diabetes in the Alameda County Study. *Int J Epidemiol*. 2005;34(6):1274–1281.
10. McLaren L. Socioeconomic status and obesity. *Epidemiol Rev.* 2007;29:29–48.
11. Lynch J, Smith GD. A life course approach to chronic disease epidemiology. *Annu Rev Public Health*. 2005;26:1–35.
12. Kuh D, Ben-Shlomo Y. Introduction: a life course approach to the aetiology of adult chronic disease. In: Kuh D, Ben-Shlomo Y, eds. *A Life Course Approach to Chronic Disease Epidemiology.* 2nd ed. Oxford, United Kingdom: Oxford University Press; 2004:3–14.
13. Pollitt RA, Rose KM, Kaufman JS. Evaluating the evidence for models of life course socioeconomic factors and cardiovascular outcomes: a systematic review [electronic article]. *BMC Public Health*. 2005;5:7.
14. Wamala SP, Lynch J, Kaplan GA. Women's exposure to early and later life socioeconomic disadvantage and coronary heart disease risk: the Stockholm Female Coronary Risk Study. *Int J Epidemiol*. 2001;30(2):275–284.
15. Davey Smith G, Hart C, Blane D, et al. Lifetime socioeconomic position and mortality: prospective observational study. *BMJ*. 1997;314(7080):547–552.
16. Davey Smith G, Hart C. Life-course socioeconomic and behavioral influences on cardiovascular disease mortality: the collaborative study. *Am J Public Health*. 2002;92(8): 1295–1298.
17. Claussen B, Davey Smith G, Thelle D. Impact of childhood and adulthood socioeconomic position on cause specific mortality: the Oslo Mortality Study. *J Epidemiol Community Health*. 2003;57(1):40–45.
18. Pensola TH, Martikainen P. Cumulative social class and mortality from various causes of adult men. *J Epidemiol Community Health*. 2003;57(9):745–751.
19. Kannel WB, Feinleib M, McNamara PM, et al. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979;110(3): 281–290.
20. Kannel W, Wolf P, Garrison R. *The Framingham Study: An Epidemiological Investigation of Cardiovascular Disease, Section 34*. Bethesda, MD: National Heart, Lung, and Blood Institute; 1987. ((NIH) 87–2703).
21. McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. *Clin Chim Acta*. 1987;166(1):1–8.
22. Dupont WD, Plummer WD Jr. Power and sample size calculations: a review and computer program. *Control Clin Trials*. 1990;11(2):116–128.
23. Galobardes B, Smith GD, Lynch JW. Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. *Ann Epidemiol*. 2006; 16(2):91–104.
24. Blane D, Hart CL, Smith GD, et al. Association of cardiovascular disease risk factors with socioeconomic position during childhood and during adulthood. *BMJ*. 1996;313(7070): 1434–1438.
25. Poulton R, Caspi A, Milne BJ, et al. Association between children's experience of socioeconomic disadvantage and adult health: a life-course study. *Lancet*. 2002;360(9346): 1640–1645.
26. Gilman SE, Abrams DB, Buka SL. Socioeconomic status over the life course and stages of cigarette use: initiation, regular use, and cessation. *J Epidemiol Community Health*. 2003; 57(10):802–808.
27. Cole SR, Hernán MA. Fallibility in estimating direct effects [comment]. *Int J Epidemiol*. 2002;31(1):163–165.
28. Kittleson MM, Meoni LA, Wang NY, et al. Association of childhood socioeconomic status with subsequent coronary heart disease in physicians. *Arch Intern Med*. 2006;166(21): 2356–2361.
29. Singh-Manoux A, Nabi H, Shipley M, et al. The role of conventional risk factors in explaining social inequalities in coronary heart disease: the relative and absolute approaches to risk. *Epidemiology.* 2008;19(4):599–605.
30. Mackenbach JP, Stirbu I, Roskam AJ, et al. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med.* 2008;358(23):2468–2481.
31. Mackenbach JP, Cavelaars AE, Kunst AE, et al. Socioeconomic inequalities in cardiovascular disease mortality; an international study. *Eur Heart J*. 2000;21(14):1141–1151.
32. Davey Smith G, Lynch JW. Socioeconomic differentials. In: Kuh D, Ben-Shlomo Y, eds. *A Life Course Approach to Chronic Disease Epidemiology.* 2nd ed. Oxford, United Kingdom: Oxford University Press; 2004.

doses of a much older therapeutic strategy, that of RAAS blockade, in addition to background contemporary practice for chronic heart failure (excepting ACE inhibitors). In the absence of efficacious novel agents, we clearly need to focus on improving the therapeutic use of existing drugs. HEAAL reminds us that much still needs to be learned in this area, specifically a renewed focus on seeking the optimum dose of such agents in patients with chronic heart failure.

Henry Krum
Centre of Cardiovascular Research and Education in Therapeutics, Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC 3004, Australia
henry.krum@med.monash.edu

I have received payment for EMPHASIS-HF (Pfizer) and ATMOSPHERE (Novartis) committee work.

1 Pitt B, Poole-Wilson PA, Segal R, et al, on behalf of the ELITE II investigators. Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial—the Losartan Heart Failure Survival Study ELITE II. *Lancet* 2000; **355:** 1582–87.

2 Dickstein K, Kjekshus J, and the OPTIMAAL Steering Committee for the OPTIMAAL Study Group. Effects of losartan and captopril on mortality and morbidity in high-risk patients after acute myocardial infarction: the OPTIMAAL randomised trial. *Lancet* 2002; **360:** 752–60.

3 Brenner BM, Cooper ME, de Zeeuw D, et al, for the RENAAL Study Investigators. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. *N Engl J Med* 2001; **345:** 861–69.

4 Dahlöf B, Devereux RB, Kjeldsen SE, et al, for the LIFE Study Group. Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. *Lancet* 2002; **359:** 995–1003.

5 Konstam MA, Neaton JD, Dickstein K, et al, for the HEAAL Investigators. Effects of high-dose versus low-dose losartan on clinical outcomes in patients with heart failure (HEAAL study): a randomised, double-blind trial. *Lancet* 2009; published online Nov 17. DOI:10.1016/S0140-6736(09)61913-9.

6 Brunner-La Rocca HP, Weilenmann D, Kiowski W, Maly FE, Candinas R, Follath F. Within-patient comparison of effects of different dosages of enalapril on functional capacity and neurohormone levels in patients with chronic heart failure. *Am Heart J* 1999; **138:** 654–62.

7 Pacher R, Stanek B, Globits S, et al. Effects of two different enalapril dosages on clinical, haemodynamic and neurohumoral response of patients with severe congestive heart failure. *Eur Heart J* 1996; **17:** 1223–32.

8 Gullestad L, Aukrust P, Ueland T, et al. Effect of high- versus low-dose angiotensin converting enzyme inhibition on cytokine levels in chronic heart failure. *J Am Coll Cardiol* 1999; **34:** 2061–67.

9 The NETWORK Investigators. Clinical outcome with enalapril in symptomatic chronic heart failure; a dose comparison. *Eur Heart J* 1998; **19:** 481–89.

10 Packer M, Poole-Wilson PA, Armstrong PW, et al, on behalf of the ATLAS Study Group. Comparative effects of low and high doses of the angiotensin-converting enzyme inhibitor, lisinopril, on morbidity and mortality in chronic heart failure. *Circulation* 1999; **100:** 2312–18.

11 Nanas JN, Alexopoulos G, Anastasiou-Nana MI, et al, for the High Enalapril Dose Study Group. Outcome of patients with congestive heart failure treated with standard versus high doses of enalapril: a multicenter study. *J Am Coll Cardiol* 2000; **36:** 2090–95.

Health and human rights: no miracle in postconflict Chechnya

Nearly 15 years after the wars began between Chechen separatist and Russian troops, Chechnya and its capital Grozny have seen reconstruction efforts and economic recovery over the past several years that some there even praise as "the Chechen miracle". Meanwhile, violent incidents in Chechnya have become dramatically more prevalent since the war over South Ossetia between Georgia and Russia last August. Violent deaths, compared with those last summer, have almost doubled in number since, from 84 to 152.[1]

Last April federal Russian military forces officially terminated what they termed a counter-terrorism operation and what human rights organisations considered "a regional system of torture, forced confessions, and fabricated trials".[2] Documented by human rights organisations as well as the national and international press, Russian forces continue their practices of enforced disappearance, torture, and other ill-treatment of alleged terrorists, while Russian authorities have faced a wave of killings of their representatives by increasingly organised insurgents.[3] The European Court of Human Rights, as of October, 2009, has ruled in over 100 cases of Chechen citizens that the Russian Government failed to properly investigate and prosecute human rights abuses by its forces, or found Russia directly responsible for human rights violations ranging from inhuman treatment to secret detention and deaths. Thousands more cases are pending.

The ongoing human rights violations from both sides of the conflict have direct health implications for the majority of Chechens still suffering from physical and mental war sequelae. Since the wars officially ended almost 10 years ago, reports from non-governmental organisations and data from UN organisations have portrayed desperate living and public health conditions in the Northern Caucasus, even compared with other disadvantaged regions within the Russian Federation.

The printed journal includes an image merely for illustration

AP

The infant mortality rate in Chechnya is at least twice the Russian average of 13 per 1000 livebirths, mainly due to diarrhoea and respiratory infections, and although trends have been improving, many thousands of women and children lack basic social services.[4]

After the wars, Chechnya's inefficient health system was left overburdened and its infrastructure literally bombed. 3 years ago, the Kremlin made reconstruction efforts in Chechnya a policy priority, and the Chechen Government reportedly invested heavily into a programme (*Sdorovye*) for health facilities and the procurement of medical equipment. International organisations and the Chechen health ministry officially announced and launched a series of programmes for health and education.[4,5]

The official portrayal of successful health-system reconstruction, whose standards still lack behind even Russian standards, particularly for subspecialty care, leaves out the many qualified health professionals who have left for Moscow or abroad in search of better work and life conditions. Patients requiring services beyond primary care need to rely on a quota system that provides a limited number of grants to refer cases for specialised treatment that cannot be provided in Chechnya.[6] International organisations offering assistance with human rights investigations[7] are confronted with the refusal of the federal government to cooperate,[8] and the few non-governmental organisations operative in postconflict Chechnya are hindered in their work by an excessive burden of administrative hurdles created by recently imposed legal requirements.[9]

Similarly, those health professionals who stay in Chechnya are reluctant to speak out for improved health-services delivery, fearing to be regarded "unpatriotic", a label which can soon become dangerous in the context of counter-terrorist activities. In a societal climate that analysts call one of the most repressive in the world,[10] health-care delivery and programming that meets the needs of the ill and vulnerable (eg, targeting women and infants, tuberculosis patients, or the 30 000 internally displaced persons still in Chechnya) are clearly challenging.

When Natalia Estemirova from the local human rights organisation Memorial was abducted and killed in Chechnya in July, 2009,[7] after a series of abductions and murders of prominent and less high-profile human rights activists, Memorial suspended its work there. The absence of independent monitoring mechanisms to hold policy makers and law-enforcement services accountable, and the lack of international attention and response, could create a dangerous nurturing ground for extremism. Spreading violence could become a threat beyond the region—and from there endanger global security.

Violence has resurged in Chechnya,[1,3] menacing the health and health systems of its war-burdened population. The legitimate goal of controlling armed violence and stabilising a disadvantaged region will have to respect human rights to reverse this trend and bring sustainable improvements in health for this postconflict society. There is no miracle in Chechnya, but a situation that calls for international pressure from academia, parliamentary, governmental, and non-governmental organisations, and journalists.

Karsten Lunze
Preventive Medicine, Boston University, Boston, MA 02118, USA
karsten.lunze@post.harvard.edu

I thank Jennifer Leaning from the Harvard Humanitarian Initiative for her contributions to this Comment. I declare that I have no conflicts of interest.

1 Mendelson SE, Malarkey M, Moore L. Violence in the North Caucasus: summer 2009. August, 2009. http://csis.org/publication/violence-north-caucasus-4 (accessed Oct 13, 2009).
2 Council of Europe. Legal remedies for human rights violations in the North Caucasus. April 15, 2008. http://www.reliefweb.int/rw/rwb.nsf/db900sid/MMAH-7DW46X?OpenDocument (accessed Oct 13, 2009).
3 Council of Europe. Committee on Legal Affairs and Human Rights, situation in the North Caucasus region: security and human rights. Sept 29, 2009. http://assembly.coe.int/CommitteeDocs/2009/ajdoc43_2009.pdf (accessed Oct 13, 2009).
4 UNICEF. Issue facing children in Russia. http://www.unicef.org/infobycountry/russia_background.html (accessed Oct 13, 2009).
5 WHO Regional Office for Europe. First Results of EU Recovery and Development Programme for Northern Caucasus discussed recently in Kislovodsk. December, 2007. http://www.euro.who.int/emergencies/fieldwork/20080109_4 (accessed Oct 13, 2009).
6 WHO. Health action in the North Caucasus. February–March, 2005. http://www.who.int/hac/crises/rus/sitreps/Health%20Action%20Feb_Mar_2005%20%20.pdf (accessed Oct 13, 2009).
7 UN. UN experts offer help in investigating killings of rights defenders. *UN Daily News* July 21, 2009. http://www.un.org/news/dh/pdf/english/2009/21072009.pdf (accessed Oct 13, 2009).
8 RIA News Agency. Russia rules out UN expert probe into Chechen activist's murder. July 22, 2009. http://en.rian.ru/russia/20090722/155588682.html (accessed Oct 13, 2009).
9 Kamhi A. The Russian NGO Law: potential conflicts with international, national, and foreign legislation. *Int J Not-for-Profit Law* 2006; **9:** 34–57.
10 Freedom House. Worst of the worst: the world's most repressive societies 2009. 2009. http://www.freedomhouse.org/uploads/WoW09/WOW%202009.pdf (accessed Oct 13, 2009).

European Heart Journal (2009) **30**, 850–856
doi:10.1093/eurheartj/ehn573

Association of pericardial fat, intrathoracic fat, and visceral abdominal fat with cardiovascular disease burden: the Framingham Heart Study

Amir A. Mahabadi[1], Joseph M. Massaro[2,4], Guido A. Rosito[1,5], Daniel Levy[4], Joanne M. Murabito[3,4], Philip A. Wolf[3,4], Christopher J. O'Donnell[4,6], Caroline S. Fox[4,7], and Udo Hoffmann[1]*

[1]Cardiac MR PET CT Program, Department of Radiology, Massachusetts General Hospital, Harvard Medical School Boston, MA, USA; [2]Department of Mathematics, Boston University, Boston, MA, USA; [3]Departments of Medicine and Neurology, Boston University School of Medicine, Boston, MA, USA; [4]Framingham Heart Study of the National Heart, Lung, and Blood Institute, Framingham, MA, USA; [5]Federal Foundation School of Medical Science of Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil; [6]Cardiology Division, Massachusetts General Hospital, Harvard Medical School Boston, MA, USA; and [7]Department of Medicine and Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA

Received 6 August 2008; revised 23 November 2008; accepted 27 November 2008; online publish-ahead-of-print 9 January 2009

Aims The aim of this study was to assess whether pericardial fat, intrathoracic fat, and visceral abdominal adipose tissue (VAT) are associated with the prevalence of cardiovascular disease (CVD).

Methods and results Participants from the Framingham Heart Study Offspring cohort underwent abdominal and chest multidetector computed tomography to quantify volumes of pericardial fat, intrathoracic fat, and VAT. Relations between each fat depot and CVD were assessed using logistic regression. The analysis of 1267 participants (mean age 60 years, 53.8% women, 9.7% with prevalent CVD) demonstrated that pericardial fat [odds ratio (OR) 1.32, 95% confidence interval (CI) 1.11–1.57; $P = 0.002$] and VAT (OR 1.35, 95% CI 1.11–1.57; $P = 0.003$), but not intrathoracic fat (OR 1.14, 95% CI 0.93–1.39; $P = 0.22$), were significantly associated with prevalent CVD in age–sex-adjusted models and after adjustment for body mass index and waist circumference. After multivariable adjustment, associations were attenuated ($P > 0.14$). Only pericardial fat was associated with prevalent myocardial infarction after adjusting for conventional measures of adiposity (OR 1.37, 95% CI 1.03–1.82; $P = 0.03$).

Conclusion Pericardial fat and VAT, but not intrathoracic fat, are associated with CVD independent of traditional measures of obesity but not after further adjustment for traditional risk factor. Taken together with our prior work, these findings may support the hypothesis that pericardial fat contributes to coronary atherosclerosis.

Keywords Pericardial fat • Visceral abdominal fat • Cardiovascular disease • Framingham Heart Study • Epidemiology

Introduction

Obesity currently affects nearly one-third of the population in the industrialized world.[1,2] Traditionally, anthropometric measures such as body mass index (BMI) or waist circumference (WC) have been used to quantify overall adiposity. However, regional fat depots may be of greater importance than overall adiposity.[3–8] Several studies have highlighted pericardial fat and abdominal visceral adipose tissue (VAT) as unique, pathogenic fat depots.[9–17] Abdominal VAT is the largest visceral fat depot in the human body with more than 10 times the volume of pericardial fat.[18] It is significantly correlated with cardiovascular disease (CVD) risk factors, the metabolic syndrome,[19,20] and systemic markers of inflammation.[21] Therefore, VAT is hypothesized to have a systemic effect on atherosclerosis.

In contrast, pericardial fat is a smaller fat depot, but surrounds the coronary arteries and the myocardium and therefore may have a local (paracrine) effect on the development of coronary

* Corresponding author. Massachusetts General Hospital, Cardiac MR PET CT Program, 165 Charles River Plaza, Suite 400, Boston, MA 02114, USA. Tel: +1 617 726 1255, Fax: +1 617 724 4152, Email: uhoffmann@partners.org

artery disease. Recently, we showed that pericardial fat is associated with metabolic risk factors.[18] Furthermore, we demonstrated that pericardial fat is correlated with the presence of coronary artery calcification.[18] Additional studies have identified an association between pericardial fat and the severity of coronary artery disease.[22,23] Taken together, these findings support the hypothesis that the magnitude of perivascular fat tissue may be a determinant of the extent of atherosclerosis in the coronary arteries.

In contrast, the impact of intrathoracic fat on total CVD, because of its small volume and lack of close proximity to the coronary arteries, may be limited, despite its correlation with CVD risk factors. Therefore, the aim of this study was to assess the association of pericardial, intrathoracic, and visceral abdominal fat with the prevalence of CVD in the Offspring population of the Framingham Heart Multi-detector Computed Tomography (MDCT) Study. We hypothesized that pericardial fat, because it may have a local effect on the coronary arteries and abdominal VAT and a systemic effect on atherosclerosis as the largest visceral fat depot, are associated with total CVD. Further, we hypothesized that intrathoracic fat, which is a small fat depot not in local anatomic contact with the coronary arteries, would be less likely to be associated with CVD.

Methods

Study sample

Participants of this study were drawn from the Framingham Heart MDCT Study sample. Offspring participants underwent MDCT imaging between June 2002 and April 2005. Exclusion criteria were pregnancy, age <40 years for women and age <35 years for men, and weight >320 pounds. Study design has been described previously.[24,25]

Overall, 1422 subjects underwent CT scanning of the chest and abdomen between June 2002 and April 2005. Of these, 1342 had non-missing or interpretable pericardial fat, intrathoracic fat, and VAT measures, and 1320 of those attended the seventh examination cycle. Of the remaining 1320, 47 subjects were excluded due to a history of prior open heart surgery, and an additional six subjects were excluded due to a missing covariate profile, resulting in a total sample size of 1267 participants.

The study was approved by the Institutional Review Boards of the Boston University Medical Center and the Massachusetts General Hospital. All subjects provided written informed consent.

Multi-detector computed tomography imaging protocol

Participants underwent radiographic assessment of their chest and abdomen in the supine position within one procedure using an eight-slice MDCT scanner (LightSpeed Ultra, General Electric, Milwaukee, WI, USA). The thoracic scan was performed during inspiratory breath hold, with an average scan length of 18 s (tube voltage 120 kVp, tube current 320 mA <220 lbs and 400 mA >220 lbs, gantry rotation time 500 ms, temporal resolution 330 ms). Image acquisition was prospectively triggered with the centre of the acquisition at 70% of the R-R-interval. Images were reconstructed with a slice thickness of 2.5 mm without overlap and a field of view of 25 cm. On average, 48 contiguous slices were taken for volume coverage from the carina to the diaphragm. For the abdominal scan, 25 contiguous slices were reconstructed with a slice thickness of 5 mm without overlap, starting 150 mm above the upper edge of S1, and a field of view of 35 cm (tube voltage: 120 kVp, tube current 320 mA <220 lbs and 400 mA >220 lbs, gantry rotation time 500 ms, pitch 1.33).

Fat tissue measurements

Pericardial fat, total thoracic fat, and VAT volumes were assessed using a dedicated workstation (Aquarius 3D Workstation, TeraRecon, San Matteo, CA, USA). Fat volumes were measured by a semi-automatic segmentation technique. The reader was required to manually trace a region of interest. Within the region of interest, fat was defined as pixels within a window of -195 to -45 Houndsfield units (HU) and a window centre of -120 HU. Pericardial fat volume was defined as any adipose tissue located within the pericardial sac. Total thoracic fat volume was defined as any adipose tissue located within the thorax from the level of the right pulmonary artery to the diaphragm and from the chest wall to the descending aorta in addition to fat inside the pericardial sac (*Figure 1*). For VAT, the muscular abdominal wall was manually traced to separate VAT from the subcutaneous fat. Inter- and intra-observer reproducibilities were excellent for VAT (ICC ≥ 0.99),[19] pericardial fat (ICC 0.95), and total thoracic fat (ICC 0.98).[18] Intrathoracic fat was defined as the difference between total thoracic fat and pericardial fat in order to create a unique fat depot that was non-overlapping with the pericardial fat compartment. This is in contrast to our prior work in which intrathoracic fat referred to total thoracic fat.

Risk factor and covariate assessment

Risk factors and covariates were measured at the seventh examination cycle (1998–2001). BMI was defined as weight (in kilograms) divided by the square of height (in metres). WC was measured at the level of the umbilicus. Fasting plasma glucose, total and high-density cholesterol, and triglycerides were measured on fasting morning samples. Diabetes was defined as fasting plasma glucose level of ≥ 126 mg/dL or treatment with either insulin or a hypoglycaemic agent. Hypertension was defined as systolic blood pressure above 140 mmHg, diastolic blood pressure above 90 mmHg, or treatment. Participants were considered current smokers if they had smoked at least one cigarette per day within the previous year. Alcohol use was defined as a consumption of more than seven drinks per week for women and more than 14 dinks per week for men. Women were considered menopausal if their periods had stopped for ≥ 1 year.

CVD included coronary heart disease (CHD, defined as recognized or unrecognized myocardial infarction, stable or unstable angina pectoris, and coronary insufficiency), stroke (defined as atherothrombotic infarction, transient ischaemic attack, cerebral embolism, intracerebral haemorrhage, and subarachnoid haemorrhage), intermittent claudication (defined as the presence of exertional calf discomfort that was relieved with rest), and congestive heart failure (CHF, defined as the presence of two major criteria or one major and two minor criteria according to the Framingham Heart Study criteria for the diagnosis of CHF[26]). All suspected CVD events were previously adjudicated by a panel of three Framingham investigators after review of all available Framingham Heart Study examination records, hospitalization records, and physician notes, as described previously in detail.[27]

Statistical analysis

Pericardial fat, intrathoracic fat, and visceral abdominal fat were normally distributed. Associations between the fat tissues were assessed using the Pearson's correlation coefficient. Associations between the amount of each fat depot and the prevalence of CVD were assessed

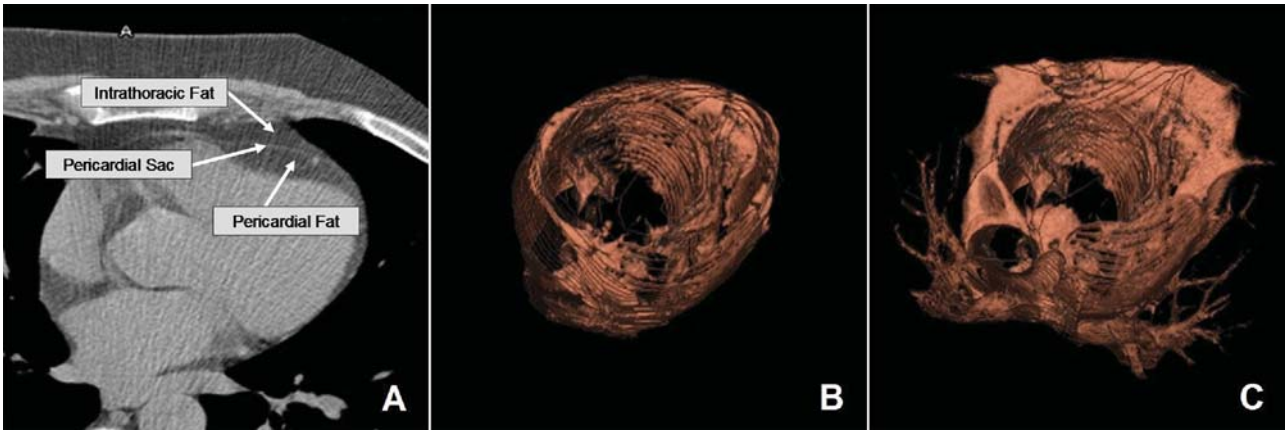


Figure 1 Pericardial fat and intrathoracic fat in an axial image, showing that the pericardial sac was defined as the border between pericardial and intrathoracic fat (*A*) and three-dimensional reconstructions of pericardial fat (*B*) and total thoracic fat (*C*). Intrathoracic fat was calculated by subtracting pericardial fat from total thoracic fat. For this subject, pericardial fat volume was 153 cm^3, intrathoracic fat volume was 287 cm^3, and VAT volume was 1865 cm^3.

using multivariable logistic regression. First, all fat depots were standardized to a mean of 0 and a standard deviation of 1 to facilitate direct comparisons of effect sizes across different fat depots. For models examining CVD as an outcome, we made the following covariate adjustments: (i) age and sex adjustment; (ii) age, sex, BMI, and WC adjustment; (iii) age, sex, BMI, WC, and multivariable adjustment, including systolic blood pressure, hypertension treatment, total cholesterol/HDL cholesterol, lipid-lowering therapy, diabetes, smoking, alcohol use, menopausal status, and hormone replacement therapy. Separate logistic regression analyses were preformed for each fat tissue. To assess for a potential clustering of the data due to subjects from the same family, we determined the *P*-values for the association of pericardial fat, intrathoracic fat, and VAT with overall CVD using general estimate equations (GEEs). For the GEE, we assumed the exchangeable compound symmetry correlation structure between members of the same nuclear family, and robust standard errors were used. In GEE, we observed similar *P*-values and statistical significance for the same models as using generalized linear models. For the three fat variables of interest (pericardial fat, intrathoracic fat, and VAT), we assess if there was a significant quadratic trend. For all three variables, the quadratic trend was not significant ($P > 0.14$).

In secondary analyses, we tested the association of pericardial fat and VAT with CHD, myocardial infarction, and stroke. Because of the relatively small number of events (myocardial infarction, $n = 39$ and stroke, $n = 19$), these models were limited to the following covariate adjustments: (i) age and sex; (ii) age, sex, BMI, and WC.

Sex interaction was tested in all models. SAS version 9.13 was used for all computations. A two-tailed *P*-value of <0.05 was considered statistically significant.

Results

Overall, 1267 participants (mean age: 60 $\pm$ 9 years, 53.8% women) were included in this evaluation. Mean pericardial fat volume was 124 $\pm$ 50 cm^3, mean intrathoracic fat volume was 115 $\pm$ 63 cm^3, and mean visceral abdominal fat volume was 2091 $\pm$ 1099 cm^3. Detailed sample characteristics are shown in *Table 1*. The three fat depots were highly correlated (Pearson's correlation coefficient: 0.68 for pericardial fat vs. intrathoracic fat, 0.63 for pericardial fat vs. VAT, and 0.74 for intrathoracic fat vs. VAT; $P < 0.001$ for all).

Fat depots and the association with cardiovascular disease

In an age- and sex-adjusted model, both pericardial fat and VAT but not intrathoracic fat were significantly associated with CVD (*Table 2*). The association of pericardial fat and visceral abdominal fat remained statistically significant after further adjustment for BMI and WC, but was attenuated after multivariable adjustment. There was no statistically significant difference in the magnitude of the associations between pericardial fat and VAT with CVD ($P = 0.99$ for the model adjusted for age and sex).

Visceral fat tissues and the association with coronary heart disease, myocardial infarction, and stroke

After observing a significant association of pericardial fat and VAT with overall CVD, we performed secondary analysis to further investigate the association of both fat tissues with CHD and myocardial infarction as an outcome variable that may be influenced by a potential local effect of pericardial fat and stroke as an outcome variable that is influenced by a systemic effect of VAT but is not as likely to be influenced by pericardial fat.

Both pericardial fat and VAT were significantly associated with the prevalence of CHD in an age- and sex-adjusted model and after further adjustment for BMI and WC. However, associations were stronger for pericardial fat than for VAT.

In an age- and sex-adjusted model, both pericardial fat and VAT were significantly associated with the prevalence of myocardial infarction. This association remained significant for pericardial fat after further adjustment for BMI and WC, but was not significant for VAT.

Table 1 Study sample characteristics

Characteristic	*n* = 1267
Age (years)	60 ± 9
Women (%)	53.8 (682)
Body mass index (kg/m^2)	28.2 ± 5.1
Waist circumference (cm)	94.2 ± 13.7
Pericardial fat (cm^3)	124 ± 50
Intrathoracic fat (cm^3)	115 ± 63
Visceral abdominal fat (cm^3)	2091 ± 1099
Systolic blood pressure (mmHg)	126 ± 18
Hypertensive treatment (%)	29.1 (369)
HDL cholesterol (mg/dL)	53.2 ± 16.0
Total cholesterol (mg/dL)	201 ± 36
Lipid-lowering treatment (%)	18.4 (233)
Diabetes[a] (%)	9.8 (124)
Smoking (%)	
Current	10.2 (129)
Former	51.9 (657)
Never	38.0 (481)
Alcohol use[b] (%)	16.3 (207)
Post-menopausal (%)	82.7 (564)
Hormone replacement therapy (%)	36.7 (250)
All CVD (%)	9.7 (123)
CHD (%)	6.3 (80)
Myocardial infarction (%)	3.1 (39)
Stroke (%)	1.5 (19)
CHF (%)	0.2 (3)
Intermittent claudication (%)	1.9 (24)

Data presented as mean ± standard deviation for continuous traits or per cent (*n*) for dichotomous traits. CVD, cardiovascular disease; CHD, coronary heart disease; HDL, high density lipoprotein; CVD, cardiovascular disease; CHF, congestive heart failure.
[a]Defined as fasting plasma glucose ≥126 mg/dL or treatment with either insulin or a hypoglycaemic agent.
[b]Defined as more than 14 drinks per week (men) or more than seven drinks per week (women).

Only VAT but not pericardial fat was associated with stroke in an age- and sex-adjusted model. The association of VAT with stroke remained statistically significant after further adjustment for BMI and WC (*Table 3*).

There was no significant sex interaction for any fat depot ($P > 0.07$ for all age- and sex-adjusted models).

Discussion

In this cross-sectional study, we examined associations of pericardial fat, intrathoracic fat, and VAT with CVD in a community-based sample and found that pericardial fat and VAT were both associated with CVD in age- and sex-adjusted models. These associations remained significant after adjustment for BMI and WC.

The finding that pericardial fat, a very small fat depot, is associated with CVD supports the hypothesis that pericardial fat may have a paracrine role in the pathogenesis of CVD. In contrast, the strong association of VAT with CVD is likely due to powerful associations between VAT and CVD risk factors, including hypertension, dyslipidaemia, and diabetes, that we and others have reported previously, suggesting that VAT has systemic effects as a pathogenic fat depot. These hypotheses are supported by our finding that pericardial fat is predominantly associated with CHD and myocardial infarction, whereas only VAT is associated with stroke. However, the lacking association of pericardial fat with stroke may be caused by a smaller number of strokes compared with myocardial infarction (MI). The finding that pericardial fat and VAT were similarly associated with overall CVD may indicate that the potential paracrine effect of pericardial fat and the systemic effect of VAT might be of comparable effect size.

However, associations were attenuated upon adjustment for CVD risk factors, suggesting that ultimately relations between fat depots and CVD are due to shared risk factors. Also, we cannot rule out that several findings may be influenced by diminished statistical power to detect modest effect sizes. Further, the association between pericardial fat, VAT, and CVD events may be mediated via shared risk factors in the pathogenesis between fat depots and actual events.

Notably, we observed different associations to CVD of pericardial fat, intrathoracic fat, and VAT, despite their close correlation.

Table 2 Association of pericardial fat, intrathoracic fat, and visceral abdominal fat with cardiovascular disease per standard deviation of fat tissue[a]

Models for all three exposures	Pericardial fat		VAT		Intrathoracic fat	
	Odds ratio (95% CI)	*P*-value	Odds ratio (95% CI)	*P*-value	Odds ratio (95% CI)	*P*-value
Age and sex	1.32 (1.11–1.57)	0.002	1.35 (1.11–1.57)	0.003	1.14 (0.93–1.39)	0.22
Age, sex, BMI, and WC	1.31 (1.08–1.59)	0.006	1.46 (1.11–1.92)	0.007	1.08 (0.85–1.59)	0.54
Age, sex, BMI, WC, and multivariable adjustment[b]	1.17 (0.95–1.45)	0.14	1.23 (0.92–1.63)	0.16	0.96 (0.75–1.24)	0.76

BMI, body mass index; WC, waist circumference; VAT, visceral abdominal fat; CI, confidence interval.
[a]All fat depots have been standardized to a mean of 0 and standard deviation of 1 to facilitate comparisons across depots.
[b]Includes systolic blood pressure, hypertension treatment, total cholesterol/HDL, lipid-lowering therapy, diabetes, smoking, alcohol use, menopausal status, and hormone replacement therapy.

Table 3 Association of pericardial fat and visceral abdominal fat with coronary heart disease, myocardial infarction, and stroke per standard deviation of fat tissue[a]

		Pericardial fat		VAT	
		OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
CHD	Age and sex adjusted	1.97 (1.32–2.94)	0.0009	1.34 (1.09–1.66)	0.006
	Age, sex, BMI, and WC adjusted	1.92 (1.23–3.02)	0.004	1.36 (1.02–1.82)	0.04
Myocardial infarction	Age and sex adjusted	1.48 (1.15–1.90)	0.002	1.55 (1.14–2.11)	0.005
	Age, sex, BMI, and WC adjusted	1.37 (1.03–1.82)	0.03	1.38 (0.91–2.08)	0.13
Stroke	Age and sex adjusted	1.43 (0.99–2.07)	0.06	1.82 (1.18–2.80)	0.006
	Age, sex, BMI, and WC adjusted	1.29 (0.85–1.98)	0.23	1.83 (1.01–3.30)	0.046

CHD, coronary heart disease; BMI, body mass index; WC, waist circumference; VAT, visceral abdominal fat; CI, confidence interval.
[a]All fat depots have been standardized to a mean of 0 and standard deviation of 1 to facilitate comparisons across depots.

In the context of the current literature

Our finding that pericardial fat is associated with the prevalence of CVD and myocardial infarction after adjustment for age and sex and traditional measures of obesity are supported by our prior work, demonstrating a significant association of pericardial fat volume with traditional risk factors and the presence of coronary artery calcium.[18] Likewise, Taguchi *et al.*[22] found a significant association between pericardial fat volume and the prevalence of coronary artery disease in non-obese Japanese subjects. In a study of 203 participants from Korea, a close association between epicardial fat thickness and the severity of coronary artery disease was found.[23] Overall, our results are consistent with the hypothesis that perivascular fat may be associated with local vascular injury. In this context, our observation that VAT and stroke are strongly associated warrants further exploration.

Potential mechanisms

The specific composition and metabolic activity of visceral fat tissues such as pericardial fat and VAT are widely recognized as differing from subcutaneous fat. Visceral fat tissues have smaller adipocyte size,[28] higher protein content,[29] high rate of fatty acid incorporation,[30] and fast insulin-induced fatty acid breakdown[29] and secrete several pro- and anti-inflammatory mediators and cytokines such as adiponectin, interleukin-6, and TNF-α.[10–16,21,31] The amount of adiponectin, a stabilizer of the inhibitor of NF-κB released from pericardial fat,[11] decreases with an increased amount of fat.[32] The decrease in adiponectin enhances the activity of NF-κB, which leads to an increase in TNF-α and hence to a local increase of inflammation.[11] A mismatch of pro- and anti-inflammatory mediators and cytokines secreted by pericardial fat is suspected to have a local influence on the underlying coronary arteries. Increased CD45 mRNA expression in the pericardial fat of subjects with coronary artery disease, representing elevated macrophage infiltration,[13] and an increase in mast cells in the adventitia of coronary lesions[14] have been observed. The hypothesis of an impact on local inflammation of pericardial fat and its role in the pathogenesis of atherosclerosis of the coronary arteries are supported by our findings.

Despite the relative size, VAT differentiates from pericardial fat in blood supply and drainage. Intra-abdominal mesenteric fat (VAT) has a circulatory communication path to the liver via the portal circulation and thus may be highly associated with insulin resistance of the liver and hepatic production of inflammatory factors such as high sensitivity-C-reactive protein. VAT is associated with metabolic risk factors,[5,15,16,31,33–36] traditional CVD risk factors,[19] and systemic inflammatory markers.[21] These associations further emphasize the importance of VAT as a mediator of systemic CVD risk factors.

In contrast, intrathoracic and pericardial fat depots are substantially smaller than VAT and are unlikely to release substances that can be detected systemically. Therefore, their hypothesized role is more likely to be paracrine via their local effect on inflammation in the underlying tissue.[10,13,14,21]

Implications

Together with our previous findings of visceral fat tissues being associated with risk factors and vascular calcification,[18,19] these findings suggest that visceral fat depots are associated with CVD. Further research is warranted to establish the incremental value of fat measurements to traditional CV risk factors and the causal relationship between pericardial fat and VAT and the development of CVD.

Strength and limitations

The strengths of our study include a community-based sample not selected for adiposity-related traits. Fat volumes were quantified using a highly reproducible volumetric CT-based measure. Limitations of our study include the predominantly white Framingham Offspring Study, hence generalization to other ethnic groups is uncertain. Also, we were only able to assess prevalent CVD and by definition could not assess the relation between fat depots and fatal CVD. Given the cross-sectional study design, we cannot establish causality. Although our data are plausible with the biological hypothesis, a cause-and-effect relationship between pericardial fat and myocardial infarction cannot be established in this cross-sectional study. However, our results are consistent with our previous data, describing a cross-sectional association of pericardial fat and coronary artery calcium in subjects without known CVD. We included individuals with CHF as outcomes and recognize that these patients may have an underlying pathology different from CAD. However, there were only three participants with CHF: one

patient had angina pectoris about 2 years after the CHF occurred; one patient had an MI about 10 years before the CHF occurred, and the third patient only had CHF. Thus, it is unlikely to have affected our results. Further, we excluded all participants who had previously undergone cardiac bypass surgery, due to the unreliability of pericardial fat in this setting. Therefore, we have excluded participants with the most informative physiology and biased our results towards the null. A further limitation of our study is the diminished statistical power to detect modest effect sizes. Thus, our findings that need to be confirmed in studies with larger numbers of CVD events, including prospective studies, are warranted to determine whether metabolic fat depots are independently associated with CVD after adjustment for traditional CVD risk factors.

Conclusion

Pericardial fat and VAT, but not intrathoracic fat, are associated with CVD independent of traditional measures of obesity. However, none of these fat depots are independently associated with CVD after further adjustment for traditional risk factors. Taken together with our prior work, these findings may support the hypothesis that pericardial fat contributes to coronary atherosclerosis but needs to be confirmed in larger studies.

Funding

This work was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (N01-HC-25195). A.A.M. is supported by a grant from the German National Academic Foundation. G.A.R. is supported by a grant from the CAPES, Brazil.

Conflict of interest: none declared.

References

1. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999–2004. *JAMA* 2006;**295**:1549–1555.
2. Poirier P, Giles TD, Bray GA, Hong Y, Stern JS, Pi-Sunyer FX, Eckel RH. Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. *Circulation* 2006;**113**:898–918.
3. Despres JP, Lemieux I. Abdominal obesity and metabolic syndrome. *Nature* 2006;**444**:881–887.
4. Snijder MB, Dekker JM, Visser M, Bouter LM, Stehouwer CD, Yudkin JS, Heine RJ, Nijpels G, Seidell JC. Trunk fat and leg fat have independent and opposite associations with fasting and postload glucose levels: the Hoorn study. *Diabetes Care* 2004;**27**:372–377.
5. Poirier P, Despres JP. Waist circumference, visceral obesity, and cardiovascular risk. *J Cardiopulm Rehabil* 2003;**23**:161–169.
6. Despres JP. Is visceral obesity the cause of the metabolic syndrome? *Ann Med* 2006;**38**:52–63.
7. Lissner L, Bjorkelund C, Heitmann BL, Seidell JC, Bengtsson C. Larger hip circumference independently predicts health and longevity in a Swedish female cohort. *Obes Res* 2001;**9**:644–646.
8. Fujimoto WY, Jablonski KA, Bray GA, Kriska A, Barrett-Connor E, Haffner S, Hanson R, Hill JO, Hubbard V, Stamm E, Pi-Sunyer FX. Body size and shape changes and the risk of diabetes in the diabetes prevention program. *Diabetes* 2007;**56**:1680–1685.
9. de Vos AM, Prokop M, Roos CJ, Meijs MF, van der Schouw YT, Rutten A, Gorter PM, Cramer MJ, Doevendans PA, Rensing BJ, Bartelink ML, Velthuis BK, Mosterd A, Bots ML. Peri-coronary epicardial adipose tissue is related to cardiovascular risk factors and coronary artery calcification in post-menopausal women. *Eur Heart J* 2008;**29**:777–783.
10. Mazurek T, Zhang L, Zalewski A, Mannion JD, Diehl JT, Arafat H, Sarov-Blat L, O'Brien S, Keiper EA, Johnson AG, Martin J, Goldstein BJ, Shi Y. Human epicardial adipose tissue is a source of inflammatory mediators. *Circulation* 2003;**108**:2460–2466.
11. Ouchi N, Kihara S, Arita Y, Okamoto Y, Maeda K, Kuriyama H, Hotta K, Nishida M, Takahashi M, Muraguchi M, Ohmoto Y, Nakamura T, Yamashita S, Funahashi T, Matsuzawa Y. Adiponectin, an adipocyte-derived plasma protein, inhibits endothelial NF-kappaB signaling through a cAMP-dependent pathway. *Circulation* 2000;**102**:1296–1301.
12. Ouchi N, Kihara S, Funahashi T, Nakamura T, Nishida M, Kumada M, Okamoto Y, Ohashi K, Nagaretani H, Kishida K, Nishizawa H, Maeda N, Kobayashi H, Hiraoka H, Matsuzawa Y. Reciprocal association of C-reactive protein with adiponectin in blood stream and adipose tissue. *Circulation* 2003;**107**:671–674.
13. Baker AR, Silva NF, Quinn DW, Harte AL, Pagano D, Bonser RS, Kumar S, McTernan PG. Human epicardial adipose tissue expresses a pathogenic profile of adipocytokines in patients with cardiovascular disease. *Cardiovasc Diabetol* 2006;**5**:1.
14. Laine P, Kaartinen M, Penttila A, Panula P, Paavonen T, Kovanen PT. Association between myocardial infarction and the mast cells in the adventitia of the infarct-related coronary artery. *Circulation* 1999;**99**:361–369.
15. Yatagai T, Nagasaka S, Taniguchi A, Fukushima M, Nakamura T, Kuroe A, Nakai Y, Ishibashi S. Hypoadiponectinemia is associated with visceral fat accumulation and insulin resistance in Japanese men with type 2 diabetes mellitus. *Metabolism* 2003;**52**:1274–1278.
16. Saijo Y, Kiyota N, Kawasaki Y, Miyazaki Y, Kashimura J, Fukuda M, Kishi R. Relationship between C-reactive protein and visceral adipose tissue in healthy Japanese subjects. *Diabetes Obes Metab* 2004;**6**:249–258.
17. Kremen J, Dolinkova M, Krajickova J, Blaha J, Anderlova K, Lacinova Z, Haluzikova D, Bosanska L, Vokurka M, Svacina S, Haluzik M. Increased subcutaneous and epicardial adipose tissue production of proinflammatory cytokines in cardiac surgery patients: possible role in postoperative insulin resistance. *J Clin Endocrinol Metab* 2006;**91**:4620–4627.
18. 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.
19. 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.
20. Vega GL, Adams-Huet B, Peshock R, Willett D, Shah B, Grundy SM. Influence of body fat content and distribution on variation in metabolic risk. *J Clin Endocrinol Metab* 2006;**91**:4459–4466.
21. Pou KM, Massaro JM, Hoffmann U, Vasan RS, Maurovich-Horvat P, Larson MG, Keaney JF Jr, Meigs JB, Lipinska I, Kathiresan S, Murabito JM, O'Donnell CJ, Benjamin EJ, Fox CS. Visceral and subcutaneous adipose tissue volumes are cross-sectionally related to markers of inflammation and oxidative stress: the Framingham Heart Study. *Circulation* 2007;**116**:1234–1241.
22. Taguchi R, Takasu J, Itani Y, Yamamoto R, Yokoyama K, Watanabe S, Masuda Y. Pericardial fat accumulation in men as a risk factor for coronary artery disease. *Atherosclerosis* 2001;**157**:203–209.
23. Jeong JW, Jeong MH, Yun KH, Oh SK, Park EM, Kim YK, Rhee SJ, Lee EM, Lee J, Yoo NJ, Kim NH, Park JC. Echocardiographic epicardial fat thickness and coronary artery disease. *Circ J* 2007;**71**:536–539.
24. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci* 1963;**107**:539–556.
25. Shurtleff D. Some characteristics related to the incidence of cardiovascular disease and death: Framingham Study, 18-year follow-up: section 30. In: Kannel W, Fordon T, eds. *The Framingham Study: and Epidemiological Investigation of Cardiovascular Disease*. Washington, DC: Department of Health, Education, and Welfare; 1973. DHEW publication No. NIH 74-599.
26. Ho KK, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: the Framingham Study. *J Am Coll Cardiol* 1993;**22**(Suppl. 4A):6A–13A.
27. Cupples L, D'Agostino R, Kiely D. *The Framingham Heart Study, Section 35. An Epidemiological Investigation of Cardiovascular Disease Survival Following Cardiovascular Events: 30 Year Follow-up*. Bethesda, MD: National Heart, Lung and Blood Institute; 1988.
28. Sons HU, Hoffmann V. Epicardial fat cell size, fat distribution and fat infiltration of the right and left ventricle of the heart. *Anat Anz* 1986;**161**:355–373.
29. Marchington JM, Mattacks CA, Pond CM. Adipose tissue in the mammalian heart and pericardium: structure, foetal development and biochemical properties. *Comp Biochem Physiol B* 1989;**94**:225–232.

30. Marchington JM, Pond CM. Site-specific properties of pericardial and epicardial adipose tissue: the effects of insulin and high-fat feeding on lipogenesis and the incorporation of fatty acids *in vitro*. *Int J Obes* 1990;**14**:1013–1022.
31. Wajchenberg BL. Subcutaneous and visceral adipose tissue: their relation to the metabolic syndrome. *Endocr Rev* 2000;**21**:697–738.
32. Arita Y, Kihara S, Ouchi N, Maeda K, Kuriyama H, Okamoto Y, Kumada M, Hotta K, Nishida M, Takahashi M, Nakamura T, Shimomura I, Muraguchi M, Ohmoto Y, Funahashi T, Matsuzawa Y. Adipocyte-derived plasma protein adiponectin acts as a platelet-derived growth factor-BB-binding protein and regulates growth factor-induced common postreceptor signal in vascular smooth muscle cell. *Circulation* 2002;**105**:2893–2898.
33. Klein S. The case of visceral fat: argument for the defense. *J Clin Invest* 2004;**113**:1530–1532.
34. Matsuzawa Y. Therapy insight: adipocytokines in metabolic syndrome and related cardiovascular disease. *Nat Clin Pract* 2006;**3**:35–42.
35. Bacha F, Saad R, Gungor N, Arslanian SA. Adiponectin in youth: relationship to visceral adiposity, insulin sensitivity, and beta-cell function. *Diabetes Care* 2004;**27**:547–552.
36. Nielsen S, Guo Z, Johnson CM, Hensrud DD, Jensen MD. Splanchnic lipolysis in human obesity. *J Clin Invest* 2004;**113**:1582–1588.

CARDIOVASCULAR FLASHLIGHT

doi:10.1093/eurheartj/ehn610
Online publish-ahead-of-print 23 January 2009

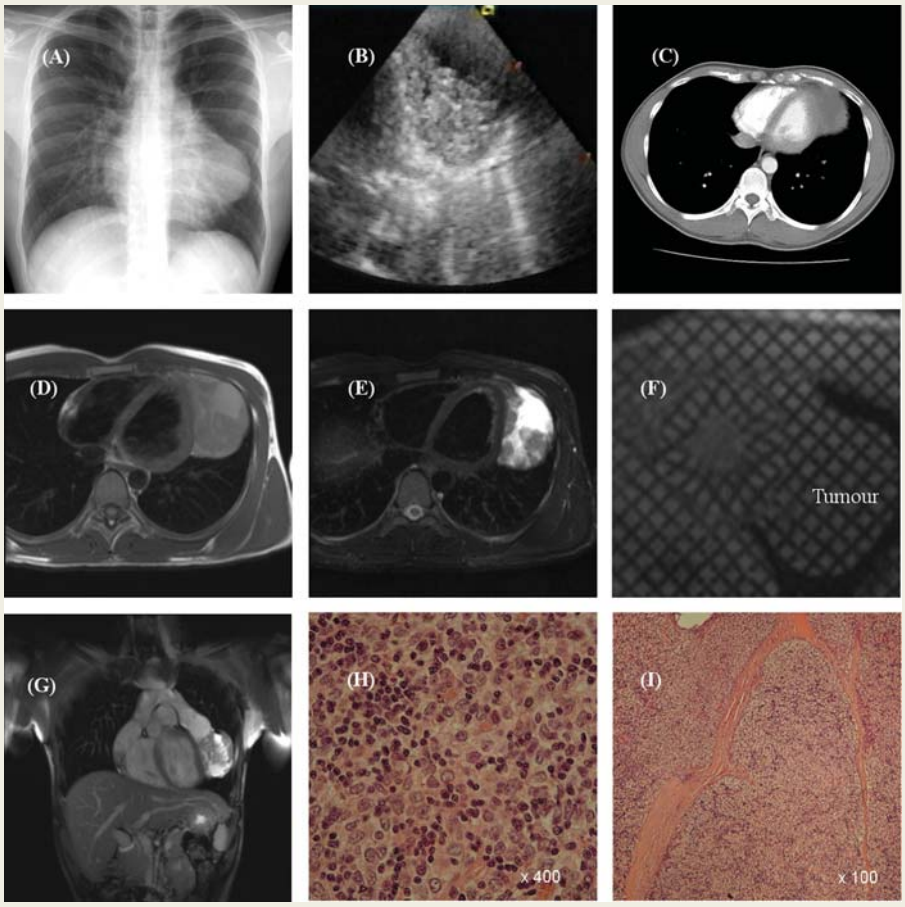
Paracardiac mass with possible cardiac infiltration: the incremental clinical value of multimodality non-invasive imaging

Bryan Wai*, Maurice Molan, and Piyush M. Srivastava

Cardiology Department, Austin Health, 145 Studley Road, Heidelberg, Melbourne, Victoria 3084, Australia

* Corresponding author. Tel: +61 394965527, Fax: +61 394964368, Email: bryan.wai@austin.org.au

An asymptomatic 24-year-old man was referred for further investigation with an incidental finding of a mass over the left heart border on an immigration chest X-ray (Panel A). Physical examination and laboratory results were normal. Transthoracic echocardiogram showed a cystic structure along the anterolateral left ventricular (LV) border with no pericardial effusion (Panel B). A CT scan showed a lobulated mass in the anterior and middle mediastinum. Cardiac infiltration could not be excluded (Panel C), resulting in an MRI being performed.



On cardiac MRI, flattening of the anterolateral LV wall with free movement of the LV myocardium against the lesion was noted (Panel G). Although heterogenous in signal, the T1 sequences (Panel D) showed that the mass was isointense to myocardium but had increased signal intensity with STIR sequences (Panel E). Myocardial tagging with breaking of tag lines at end systole suggested that there was no infiltration of the myocardium (Panel F).

This information assisted the surgical team in pre-operative planning; at resection, the mass was found to be well circumscribed without myocardial involvement.

Pathological findings were of a cystic-encapsulated structure attached to the thymus gland, measuring 7 × 6.5 × 3 cm. Histological specimens showed an encapsulated tumour with lobulation of architecture, fibrous septation, and a biphasic composition of small lymphocytes and polygonal epithelial cells consistent with thymoma (Panels H and I). No invasion of mediastinal fat was identified.

This case serves to highlight the incremental clinical value of non-invasive imaging modalities in delineating cardiac involvement of paracardiac masses.

Preventive Medicine 48 (2009) 593–595

Contents lists available at ScienceDirect

Preventive Medicine

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

Factors associated with sexual assault and time to presentation ☆

Jennifer S. McCall-Hosenfeld [a,*], Karen M. Freund [b,c], Jane M. Liebschutz [c,d]

[a] Pennsylvania State University College of Medicine, Departments of Medicine and Public Health Sciences, Hershey, PA, USA
[b] Women's Health Unit and Women's Health Interdisciplinary Research Center, Evans Department of Medicine, Boston University School of Medicine, Boston, MA, USA
[c] Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA
[d] Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA

ARTICLE INFO

Available online 31 March 2009

Keywords:
Rape
Sexual assault
Women
Forensic medicine
Domestic violence

ABSTRACT

Objective. Delay to medical care after sexual assault can be associated with adverse consequences for the assault survivor. Few studies examine factors associated with timely presentation to care after sexual assault. Using data from the Massachusetts Sexual Assault Nurse Examiner (SANE) program, we examine sexual assault and survivor characteristics and their association with time to presentation after sexual assault.

Method. Cross-sectional data were collected during forensic exam for all patients presenting to 24 SANE-affiliated hospitals in Massachusetts between July 2003 and June 2005. Data included patient demographics, assailant information, and assault characteristics. A Cox proportional hazards model described factors associated with delayed presentation for post-assault care.

Results. 478 females presented to SANE hospitals over two years. 66% were white, non-Hispanic; 14% Hispanic and 13% black; 39% were between 18 and 24 years old. The median time from onset of assault to presentation was 16 h. In multivariable analysis, assault by a known assailant was associated with delayed presentation (hazard ratio = 0.71, 95% confidence interval = 0.57, 0.88).

Conclusion. Most women who present for exam following sexual assault do so expeditiously. If an assailant is a family member or date, a woman is more likely to delay post-assault care. These findings can inform public health interventions.

Introduction

Lifetime prevalence of rape is 18% among US adult women, with annual prevalence between 0.3 (Tjaden and Thoennes, November 1998) and 1.1% (Moracco et al., 2007). Few rape survivors seek immediate medical attention, even with serious injury (Tjaden and Thoennes, November 1998). Delayed presentation may result in loss of forensic evidence and postpones treatment for sexually transmitted infections, postexposure prophylaxis for HIV, and emergency contraception, which are maximally effective if given early after assault (Resnick et al., 2000). Rape sequelae include illness and increased healthcare utilization (Suris et al., 2004). Medical care early after sexual assault may reduce these sequelae (Resnick et al., 2000).

Little prior data describes factors associated with delay to presentation for post-assault care. Prior study has retrospectively compared those who engage in post-assault medical care to non-presenters (Resnick et al., 2000). In a single institution study (Millar et al., 2002), severe assault and an unknown perpetrator were associated with earlier presentation.

Under Massachusetts law, medical providers treating sexual assault survivors must complete an anonymous forensic encounter form (Fallon et al., 2006). In Massachusetts, Sexual Assault Nurse Examiners (SANEs) complete at least 24% of these encounter forms, (Fallon et al., 2006) covering 24 hospital emergency rooms. SANEs are more likely than non-SANE providers to provide post-assault STD prophylaxis and emergency contraception (Campbell et al., 2005). SANEs collect higher quality forensic evidence. When cases are prosecuted, SANE exams are associated with more convictions (Campbell et al., 2005). SANE care may be psychologically beneficial to sexual assault survivors (Campbell et al., 2005).

Using data from the Massachusetts SANE program, we describe characteristics of the assault, assailant and survivor, and examine the association of these characteristics with time to presentation after sexual assault.

Methods

Our sample included all subjects for whom a forensic encounter form was submitted by a SANE for the two-year period from July 2003 through June 2005. We limited analyses to female assault survivors age 12 and older.

☆ Preliminary results from this work were presented at the Society of General Internal Medicine 31st Annual Meeting, Pittsburgh, PA on April 10, 2008.

* Corresponding author. Milton S. Hershey Medical Center, College of Medicine, Division of General Internal Medicine, Mail Code HU15, P.O. Box 850, Hershey, PA 17033-0850, USA. Fax: +1 717 531 7726.
E-mail address: jsm31@psu.edu (J.S. McCall-Hosenfeld).

0091-7435/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ypmed.2009.03.016

Table 1
Females presenting to Massachusetts emergency rooms, SANE forensic encounter form, 7/03–6/05.

	$N=478$	%
Age		
12–17	132	28
18–24	185	39
25–59	153	32
60+	8	2
Race/ethnicity		
White, non-Hispanic	317	66
Hispanic	67	14
Black	60	13
Asian	9	2
Other/unsure	25	5
Assault location		
House/dorm	310	65
Outdoors	64	13
Other/unsure	104	22
Region ($N=465$)		
Boston	126	27
Central	11	2
Northeast	61	13
Southeast	179	38
Western	88	19
Penetration	471	99
Ejaculation	467	98
Severe violence	217	45
Verbal threat	137	29
Intoxicant exposure	76	16
More than 1 assailant		
Yes	76	16
No	337	71
Unknown	65	14
Perpetrator relationship		
Stranger	158	33
Family	23	5
Acquaintance/date	238	50
Other/unsure	59	12
Condom use		
Yes	64	13
No	234	49
Unsure/unknown	180	38
	Median	IQR
Hours to ED presentation ($n=392$)	15.5	6.08–31.25
	Minimum, maximum	Range
	0.5, 117	116.5

SANE = sexual assault nurse examiner.
IQR = interquartile range.
ED = emergency department.

Dependent variable: Time to presentation was measured as the difference in hours between assault time and date and time and date of presentation for SANE exam, recorded on the forensic encounter form.

Independent variables: Age was categorized based Massachusetts Rape Crisis Center groupings as 12–17, 18–24, 25–59 and 60 years or greater. Assault survivors reported number of assailants and relationship of the assailant to the survivors (stranger, family member, or date or acquaintance).

We created dichotomous variables based on theoretical (common characteristics) and statistical (cells with small numbers) factors: race/ethnicity (white, non-Hispanic versus nonwhite); age (teenage [ages 12–17] versus adult [18 or older]), relationship of the assailant to the survivor (known [family, date or acquaintance] versus unknown), severe violence (beating, physical restraints, burns, bites or use of a weapon versus none), verbal threats (versus none), intoxicant exposure (voluntary or involuntary use of intoxicants including alcohol versus none), assault surroundings (home or dorm versus another location), and multiple assailants (greater than one versus one).

Statistical analyses

We computed frequencies for categorical data. For continuous variables, we computed means with standard deviations for parametric and medians with interquartile ranges for nonparametric distributions.

We used the Cochrane–Mantel–Hantzel chi-square test to perform bivariate tests of whether pertinent assault characteristics (severe violence, verbal threats, intoxicant exposure, assault surroundings, and multiple assailants) differed by age or race/ethnicity of the assault survivor, or by the relationship of the assailant to the survivor.

To determine the factors associated with time to presentation after sexual assault, we performed bivariate comparisons of median time to presentation using the Wilcoxon Rank Sum. A Cox proportional hazards model described independent factors associated with time to presentation. A decreased hazard ratio for presenting to the emergency department is indicative of a relative delay in presentation among those who are positive relative to those who are negative on a given factor.

Our final model included prespecified demographic criteria (age and race/ethnicity) and all variables with $p<0.10$ in bivariate analyses. To confirm the robustness of our variable selection, we used multiple selection procedures including stepwise (entry criterion $p<0.05$, retention criterion $p<0.10$), backwards (retention criterion $p<0.10$), and best subsets selection. Although assault in a house or dorm was significant in bivariate analysis, we excluded this from the multivariable model due to high association with the known assailant variable ($\chi^2=54$, $p<0.001$). All analyses were performed using SAS software, Version 9 (SAS Institute, Cary, NC).

Results

During the two-year interval, 478 SANE forensic encounter forms were filed on women. Assault characteristics are shown in Table 1.

In bivariate associations, 55% of those who reported a known assailant reported severe violence, compared to 44% of those who did not know their assailant ($p<0.01$). Teenagers (age 12–17) were more likely to be assaulted by more than one assailant compared to adults, (OR = 2.15, 95% CI = 1.28, 3.62). Subjects assaulted in a house or dorm were more likely to know their assailant (OR = 4.07, 95% CI = 2.73, 6.07). None of the remaining bivariate associations was significant.

Time to presentation data were available for 392 subjects. There were no significant differences in assault characteristics or demographics between these 392 subjects and those for whom time to presentation data were not available. The median time to presentation was 16 h; 95% presented within 72 h.

Analysis of bivariate comparisons of median time to presentation by potential explanatory variables yielded the following: severe violence was associated with earlier presentation (median time 13 h versus 17 h among those not reporting severe violence, $p<0.01$), verbal threats were associated with earlier presentation (10 h versus 17 h, $p<0.01$). Intoxicant exposure (18 h versus 14 h, $p<0.05$) and assault at home (16 h versus 12 h, $p<0.05$) were associated with later presentation. A known assailant was suggestive of an association with later presentation (17 h versus 15 h, $p=0.056$). White, non-Hispanic race/ethnicity, teenage status, penetration, ejaculation, condom use, and multiple assailants were not significantly associated with time to presentation.

Table 2 shows multivariable analysis for time to presentation. A known assailant was associated with delayed presentation to the emergency department, and verbal threats were associated with earlier presentation. The remaining covariates were not independently associated with time to presentation.

Table 2
Time to presentation after sexual assault, females presenting to Massachusetts emergency rooms, SANE forensic encounter form, 7/03–6/05 ($N=392$).

Variable	Number	Total person-hours	aHR	95% CI	
White, non-Hispanic	267	6071	1.20	0.95	1.53
Nonwhite	125	3147	Ref.		
Age 12–17	110	2744	0.92	0.72	1.18
Adult	277	6291	Ref.		
Known assailant	215	5552	**0.71**[a]	**0.57**[a]	**0.88**[a]
Unknown assailant	133	2460	Ref.		
Severe violence	181	3786	1.12	0.88	1.41
No severe violence	211	5432	Ref.		
Verbal threat	115	2251	**1.29**[a]	**1.00**[a]	**1.65**[a]
No verbal threat	277	6967	Ref.		
Intoxicant exposure	67	1708	0.83	0.60	1.14
No intoxicant exposure	325	7510	Ref.		

SANE = sexual assault nurse examiner.
aHR = adjusted hazard ratio. An adjusted hazard ratio of less than one indicates a delay to presentation for care compared to the reference group, whereas an adjusted hazard ratio of greater than one indicates earlier presentation to care compared to the reference group.
CI = confidence interval.
[a] $p<0.05$.

Discussion

Among women who present for SANE evaluation following sexual assault in Massachusetts, most do so expeditiously. Most of our sample would be eligible for collection of forensic evidence, emergency contraception (Resnick et al., 2000), and HIV chemoprophylaxis (Smith et al., 2005), all recommended within 72 h of sexual assault. This is important given the high rates of body cavity penetration and ejaculation and low rate of condom use among perpetrators in this cohort.

The high rate of severe violence (55%) reported by subjects who knew their assailants contests the conventional view that severe violence is associated with stranger assault (Stermac et al., 1995). However, need for medical care due to severe violence may differentially increase presentation for care among subjects who know their assailants. Concordant with prior report, (McDermott et al., 2008, Resnick et al., 2000) our data indicate that a majority of assault survivors knew their assailant.

Study limitations and strengths

Our data have several limitations. Providers may underreport treatment of sexual assault survivors. These data reflect a portion of those sexual assault survivors who present for post-assault care, and may not be generalizable to survivors who present to non-SANE hospitals or who do not present for immediate care.

Older persons are underrepresented in our data (2% over age 60) compared to census data for Massachusetts in 2005 (12%). However, the frequency of sexual assault among older persons in our study is similar to previous data (Zink and Fisher, 2006), and consistent with prior report that younger individuals are disproportionately likely to be sexually assaulted (McDermott et al., 2008). Frequency of Black (13%) or Hispanic (14%) race/ethnicity is identical to census data. Our data indicate a lower rate of intoxicant exposure (16%) compared to an Irish sample (32%, McDermott et al., 2008).

The strongest independent predictor of delay to presentation was a known assailant. These data are a logical extension of prior report that women assaulted by intimates are less likely to ever seek medical care (Resnick et al., 2000).

Conclusions

Because most sexual assailants are known to survivors, delaying care among those who know their assailant reflects a substantial public health risk. Preventive efforts must include public health campaigns to better inform survivors that assault by a known assailant is a crime, and early access to medical care may mitigate adverse outcomes.

Conflict of interest statement
The authors declare that there are no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the Sexual Assault Nurse Examiners of Massachusetts for their work caring for the survivors of sexual assault. The authors thank Carlene A. Pavlos, Director of the Division of Violence and Injury Prevention for the Massachusetts Department of Public Health, for access to this rich data set, Lucia Zuniga, SANE Program Director, Massachusetts Office for Victim Assistance, for critical review of the manuscript and for her leadership of the SANE program, and Vera Mouradian, PhD of the Massachusetts Department of Public Health, for additional manuscript review. Dr. McCall-Hosenfeld was supported by a Department of Veterans Affairs Special Fellowship in the Health Issues of Women Veterans while this research was conducted. Dr. Liesbchutz was supported by National Institute on Drug Abuse K23 DA016665. The views expressed herein are those of the authors and do not necessarily reflect the opinions of the Department of Veterans Affairs, the Massachusetts Department of Public Health, the Massachusetts Sexual Assault Nurse Examiner Program, or the National Institute on Drug Abuse.

References

Campbell, R., Patterson, D., Lichty, L.F., 2005. The effectiveness of Sexual Assault Nurse Examiner (SANE) programs: a review of the psychological, medical, legal and community outcomes. Trauma Violence Abuse 6, 313–329.
Fallon, R., Lawrence, S., Penman, S., 2006. Understanding Sexual Victimization: Using Medical Provider Data to Describe the Nature and Context of Sexual Crime in Massachusetts. Massachusetts Executive Office of Public Safety.
McDermott, S.D., McBride, B.M., Lee-Gorman, M., 2008. Sexual assault statistics from the republic of Ireland for 2004–2005. Med. Sci. Law 48, 142–147.
Millar, G., Stermac, L., Addison, M., 2002. Immediate and delayed treatment seeking among adult sexual assault victims. Women Health 35, 53–64.
Moracco, K.E., Runyan, C.W., Bowling, J.M., Earp, J.A.L., 2007. Women's experiences with violence: a national study. Womens Health Issues 17, 3–12.
Resnick, H.S., Holmes, M.M., Kilpatrick, D.G., et al., 2000. Predictors of post-rape medical care in a national sample of women. Am. J. Prev. Med. 19, 214–219.
Smith, D.K., Grohskopf, L.A., Black, R.J., et al., 2005. Antiretroviral postexposure prophylaxis after sexual, injection-drug use, or other nonoccupational exposure to HIV in the United States: recommendations from the U.S. Department of Health and Human Services. MMWR Recomm. Rep. 54, 1–20.
Stermac, L.E., DuMont, J.A., Kalemba, V., 1995. Comparison of sexual assaults by strangers and known assailants in an urban population of women. Can. Med. Assoc. J. 153, 1089–1094.
Suris, A., Lind, L., Kashner, T.M., Borman, P.D., Petty, F., 2004. Sexual assault in women veterans: an examination of PTSD risk, health care utilization, and cost of care. Psychosom. Med. 66, 749–756.
Tjaden, P., Thonnes, N., November 1998. Prevalence, Incidence and Consequences of Violence Against Women: Findings from the National Violence Against Women Survey. Washington, DC, U.S. Department of Justice, Report No. NCJ 172837.
Zink, T., Fisher, B.S., 2006. The prevalence and incidence of intimate partner and interpersonal mistreatment in older women in primary care offices. J. Elder Abuse Negl. 18, 83–105.

JOURNAL OF WOMEN'S HEALTH
Volume 18, Number 6, 2009

DOI: 10.1089/jwh.2008.0987

Sexual Assault in the Military and Its Impact on Sexual Satisfaction in Women Veterans: A Proposed Model

Jennifer S. McCall-Hosenfeld, M.D., M.Sc.,[1] Jane M. Liebschutz, M.D., M.P.H.,[2]
Avron Spiro, III, Ph.D.,[3] and Margaret R. Seaver, M.D., M.P.H.[4]

Abstract

Aims: Sexual assault in the military (SAIM) is associated with decreased sexual satisfaction. However, mediators of this association have not been fully described.
Methods: Using a retrospective analysis of cross-sectional data collected for the national Veterans Affairs (VA) Women's Health Survey, we propose a mediator model to explain the association between SAIM and decreased sexual satisfaction among women veterans. Four mediators of the association between SAIM and decreased sexual satisfaction are tested: (1) emotional health-related quality of life, (2) physical health-related quality of life, (3) lack of a close partner, and (4) gynecological illness. These mediators were chosen to encompass independent domains potentially relevant to sexual satisfaction, including emotional, physical, and relational.
Results: Of 3161 women (87%) who answered the sexual satisfaction question, the mean age was 45 (SD 15) years; 85% were white. Twenty-four percent reported a history of SAIM, and 39% reported sexual dissatisfaction. In age-adjusted logistic regression analyses, both SAIM and sexual dissatisfaction were strongly associated with each of the proposed mediators. However, of the four mediators, emotional health-related quality of life most strongly attenuated the association between SAIM and sexual dissatisfaction. After including all mediators, the association between SAIM and decreased sexual satisfaction was markedly attenuated.
Conclusions: SAIM's negative impact on sexual satisfaction in women veterans operates both directly and through its physical and mental health sequelae. Of the proposed mediators in this association, the most prominent is mental health-related quality of life; the other proposed mediators were minimally related.

Introduction

Sexual satisfaction is a complex construct that incorporates multiple domains, including physical and emotional satisfaction with sexual activity and satisfaction with sexual interpersonal relationships.[1–5] Sexual satisfaction is defined as "an affective response arising from one's subjective evaluation of the positive and negative dimensions associated with one's sexual relationship"[6] but may incorporate both partnered sexual activity and self-stimulation.[2] Although sexual satisfaction is increasingly acknowledged as an important domain of health-related quality of life[7] and emotional and relational health,[8] the correlates of sexual satisfaction are not well described.[2]

Women veterans report high rates of decreased sexual satisfaction.[9,10] This may be due to the disproportionate burden of risk factors for sexual dissatisfaction found among women veterans. A history of sexual assault in the military (SAIM) is unique to women veterans and highly associated with decreased sexual satisfaction.[10] SAIM may be more traumatizing than other forms of sexual violence because of assault by a close or trusted colleague, use of a weapon, or perception of inadequate response by the judicial system.[11] More than half of the women who report a history of SAIM were not satisfied with their sex life, compared with 34% of women who did not report a history of SAIM.[10] Sexual trauma may predispose to decreased sexual satisfaction through interference with several domains of sexual satisfaction, including emotional

[1]Pennsylvania State University College of Medicine, Hershey, Pennsylvania.
[2]Boston University School of Medicine, Boston, Massachusetts.
[3]VA Boston Healthcare System and Boston University School of Public Health, Boston, Massachusetts.
[4]Lahey Clinic, Burlington, Massachusetts.

satisfaction, physical sexual function, including gynecological functioning, and relational and interpersonal satisfaction.

SAIM is highly associated with adverse mental health consequences.[12,13] Sexual assault is associated with posttraumatic stress disorder (PTSD) at higher rates than are found in victims of other types of physical violence,[14,15] and military sexual trauma more strongly predicts the development of PTSD than other types of trauma.[16–18] Mental health disorders are highly associated with decreased sexual functioning,[11,12,19–22] which may reciprocally lower satisfaction with emotional elements of sexual life.

Medical sequelae of sexual trauma include physical injuries, a greater burden of chronic somatic complaints,[11,23] and decreased overall health-related quality of life.[23] Increased physical illness[24,25] and decreased health-related quality of life[26] are likewise associated with sexual dysfunction. Satisfaction with physical aspects of sexuality may be compromised by physical illness and overall decreased health-related quality of life, which can result from sexual trauma.

An extensive literature describes the adverse impact of sexual assault on gynecological and reproductive functioning. Chronic gynecological complaints are common among rape survivors,[27,28] many of whom suffer vaginal and perineal tears and the late sequelae of sexually transmitted diseases (STDs).[23] Sequelae of rape that may directly affect sexual satisfaction include lack of pleasure with a sexual encounter and dyspareunia.[27–29] Other medically unexplained gynecological complaints that are strongly associated with sexual assault history include unexplained menstrual irregularity,[27] dysmenorrhea,[27] and increased hysterectomy.[30] The role of medically unexplained gynecological symptoms as a potential mediator for sexual dissatisfaction has not been established.

Traumatic experiences adversely affect interpersonal relations among trauma survivors,[31,32] including disruption in the ability to form close partnerships. Relational factors, such as intimacy,[4] marital stability and quality,[33] marital satisfaction,[1] and satisfaction with nonsexual aspects of a relationship,[1] are associated with sexual satisfaction. Thus, trauma may mediate decreased sexual satisfaction through its effect on the relational domain of sexual satisfaction.

Prior work has linked decreased emotional health, decreased physical health, including gynecological function, and disrupted interpersonal relations to both sexual trauma and sexual dissatisfaction. However, these factors have never previously been assessed for their potential mediation of the association between sexual trauma and sexual dissatisfaction. In this study, we develop and test a conceptual model to describe the association of SAIM and sexual dissatisfaction through four proposed mediators: (1) lower emotional health-related quality of life, (2) lower physical health-related quality of life, (3) gynecological problems, and (4) disruption in interpersonal relationships as measured by lack of a close partner. We use the data of Skinner et al.[9–12] to examine the hypothesis that these physical and psychosocial correlates of sexual trauma mediate the association between SAIM and decreased sexual satisfaction among women veterans.

Materials and Methods

Subjects

We used data from the Veterans Affairs (VA) Womens' Health Project, a cross-sectional, national survey of 3632 women veterans[11] that was designed to characterize the health-related quality of life of women who use VA ambulatory services. The design of this study has been described in detail elsewhere.[9] Briefly, a randomly selected subset of all women veterans who had at least one outpatient VA visit between July 1, 1994, and June 30, 1995, was identified. Eligible subjects were mailed a self-administered questionnaire and returned it in a postage-paid envelope.

Dependent variable

Sexual satisfaction was measured using a single, face-valid, item asking: Overall how satisfied are you with your sex life? 1, very dissatisfied; 2, dissatisfied; 3, satisfied; 4, very satisfied. Responses were dichotomized into satisfied (answer 3 or 4) vs. dissatisfied (1 or 2), as described in other studies.[2,7,34] The sexual satisfaction construct is deliberately broad to measure the emotional, physical, gynecological, and relational domains relevant to sexual satisfaction. Use of a single, face-valid item to measure overall sexual satisfaction has been correlated in other studies to domains of functional sexuality, including sexual interest and number of daily sexual thoughts,[34] and relational and emotional factors, including family affection and partner initiation and communication.[2]

Independent variable

Self-reported SAIM was measured by an affirmative response to the question: Did you ever have an experience where someone used force or the threat of force to have sexual relations with you against your will while you were in the military?[11,12] This definition conforms to the U.S. Merit System Protection Board guidelines.[12]

Mediator variables

We defined four potential mediators of the association between SAIM and sexual satisfaction: emotional health-related quality of life, physical health-related quality of life, gynecological illness, and absence of a close partner.

Emotional health-related quality of life was measured with the mental health composite score (MCS), and physical health-related quality of life was measured with the physical health composite score (PCS) of the Short Form (SF)-36. The SF-36 is composed of eight subscales that aggregate to two higher-order clusters. Vitality, social functioning, role limitations due to emotional health, and mental health aggregate to the MCS. Physical functioning, role limitations due to physical functioning, bodily pain, and general health aggregate to the PCS. Reliability statistics for these measures exceed 80%, and validity has been established by comparison of these measures to other accepted clinical indicators for multiple disease states.[35–37] In these measures, scores range from 0 to 100, with a higher score indicating a more favorable health state.

Gynecological illness was a composite variable created by endorsing at least one of the following gynecological conditions: current endometriosis; vaginitis or yeast infections; abnormal, heavy, or irregular periods; chronic pelvic pain or painful periods; or a history of problems getting pregnant, hysterectomy, or abnormal Pap smears. In the construction of the composite gynecological variable, we first independently associated each age-adjusted element of the composite

gynecological variable with sexual trauma and then did the same for sexual satisfaction. Each element of this composite variable was significantly ($p < 0.05$) and independently associated with both sexual satisfaction and sexual trauma. All the associations were in the expected direction (greater gynecological complaints among subjects endorsing sexual satisfaction and sexual trauma, respectively), suggesting each could logically be placed in the mediator pathway.

Close partner was created from two survey line items. Subjects endorsed that they had a spouse or partner and subsequently that this spouse or partner was someone they "feel very close and intimate with." This definition was chosen because these two face-valid questions incorporate necessary elements of a published definition of a close interpersonal relationship, defined as "repeated interactions over time characterized by enduring bonds, emotional attachment, personal need fulfillment and irreplaceability."[38]

Covariates

Covariates included self-reported age, educational status, household income, and smoking. Alcohol abuse was defined by a score of three or more on a validated five-item screening test.[12]

Statistical methods

We compared demographic variables between sexually satisfied and dissatisfied women. We used *t* tests to compare continuous variables and chi-square to compare frequencies between categorical variables. An alpha ≤ 0.05 determined significance for all statistics.

To assess the clinical significance of statistical differences, we calculated an effect size measure. The effect size is a proportion of a standard deviation (SD), defined as the absolute value of the difference between the mean scores of the satisfied vs. dissatisfied women, divided by the SD of the reference group (satisfied).[39] Using previously published guidelines, an effect size of 0.20–0.40 is considered small, 0.50–0.79 is moderate, and ≥ 0.80 is large.[40]

We used logistic regression to separately model the association between each of the four proposed mediators and the exposure (SAIM), adjusting for age. We similarly modeled the association between each proposed mediator and the outcome (sexual satisfaction), adjusting for age. Lack of an independent association of any proposed mediator with either the exposure or outcome would suggest that the proposed mediator is incorrectly identified as a mediator.

After determining that each of the mediators was independently associated with both SAIM and sexual satisfaction, we used a nested approach to test the proposed mediation model, fitting a sequence of logistic regression models. In nested modeling, each successive model includes significant or prespecified predictors from the previous model and adds potential predictors from the next domain. We first tested an unadjusted odds ratio (OR) for decreased sexual satisfaction, given SAIM (model 1). Next (model 2), we tested and retained all demographic and health-related variables that changed this unadjusted effect estimate by $\geq 10\%$. Next we tested four models, one for each proposed mediator. If a potential mediator reduced or eliminated the effect of SAIM on sexual satisfaction, this was consistent with our hypothesis and retained in our mediator model. If a potential mediator was either not statistically significantly associated or did not alter the effect estimate of SAIM on sexual satisfaction by at least 10% or more, it was inconsistent with our model. In our third model (model 3), we created a fully adjusted model, retaining all proposed mediators and demographic and health-related variables that were statistically significant or changed the unadjusted effect estimate by $\geq 10\%$. All statistical analyses were run using SAS version 9.1 (SAS Institute, Cary, NC).

Results

As has been reported previously, 3632 women completed questionnaires, a 58.4% response rate.[11] Of these, 3181 women met inclusion criteria for our analyses, of whom 24% endorsed a history of SAIM.[11]

In unadjusted analyses (Table 1), respondents who reported dissatisfaction were more likely than those who reported satisfaction to be older, unmarried, more educated, and of lower household income. Respondents who reported dissatisfaction were more likely to endorse more smoking and alcohol abuse.

All proposed mediators were associated with sexual dissatisfaction. The mean mental health composite score among the sexually satisfied was 44, compared with 35 among the sexually dissatisfied. The effect size for emotional health-related quality of life on sexual satisfaction was 0.73, representing a moderate clinical difference. The mean physical health composite score among the sexually satisfied was 39, compared with 37 among the sexually dissatisfied. This effect size was 0.16, a clinically inconsequential effect.

A high proportion of our sample endorsed any gynecological problem. However, this was more frequent among the sexually dissatisfied (86%) than the satisfied (76%). The proportion of respondents endorsing no close partner was greater among the sexually dissatisfied (46%) versus the satisfied (25%).

In Table 2, each of the proposed mediators was significantly associated with both the exposure to SAIM and the outcome of sexual dissatisfaction (all $p < 0.001$), and each of the associations was in the expected direction. Thus, each of the proposed mediators independently met criteria for inclusion in our adjusted final model describing the association between SAIM and sexual satisfaction.

Table 3 shows the results of our nested logistic regression analyses to determine the association of our proposed mediators between SAIM and decreased sexual satisfaction. In model 1 (unadjusted), we found that SAIM was associated with 2.3 greater odds of decreased sexual satisfaction. After adjusting for demographics, including age, marital status, race, education, income, and the health behaviors of smoking and alcohol abuse (shown in model 2), we found modest attenuation of the association of SAIM and sexual dissatisfaction, with SAIM conferring 1.8 greater odds of decreased sexual satisfaction.

We then created four models, adding each proposed mediator independently. Addition of the mental health composite score produced the most marked reduction in the association between SAIM and sexual dissatisfaction, reducing the OR from 1.78 to 1.42. This is a 20% change in the overall effect estimate, suggesting that emotional health-related quality of life is a substantial mediator of this association.

Table 1. Demographics of Female Veteran Outpatients, by Sexual Satisfaction

Variables	*% sexually satisfied*[a] n = *1937 (61.3%)*	*% sexually dissatisfied*[a] n = *1244 (37.8%)*
Sexual assault in the military***	18	33
Demographics		
Age, years, mean (SD)***	46 (16)	43 (13)
Marital status***		
Married/partnered	45	33
Divorced/separated	29	39
Widowed	8	6
Never married	19	21
Race ($p = 0.11$)		
Black	19	21
White	75	72
Other	6	7
Education*		
1–12 years	28	24
13+ years	73	76
Household income**		
<$20,000	40	44
$20,000 – $49,999	42	42
$50,000+	13	9
Unknown	5	5
Health Behaviors		
Smoking***		
Never	38	34
Former	31	26
Current	31	40
Alcohol abuse***	7	12
Proposed Mediators		
SF-36 mental health composite score (0–100), mean (SD)***	44.1 (12.4)	35.1 (12.0)
SF-36 physical health composite score (0–100), mean (SD)***	39.0 (12.6)	37.0 (11.7)
Any gynecological problem***	76	86
No close partner***	25	46

[a]Percentages may not add up to 100% because of rounding error.
***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

A substantially smaller effect was found with the other proposed mediators. The addition of physical component scale to model 2 resulted in a small (1%) reduction in the increased odds of sexual dissatisfaction relative to the increased odds of sexual dissatisfaction associated with model 2 (1.78 to 1.76). Endorsing any gynecological problem decreased the OR of sexual dissatisfaction from 1.78 to 1.72, a minimal change of 3% in the effect estimate. Addition of the close partner variable also attenuated the association between SAIM and sexual dissatisfaction, reducing the OR from 1.78 to 1.70, a 4% change in the overall adjusted effect estimate.

Table 4 shows the final adjusted model, with independent factors associated with sexual dissatisfaction. After controlling for demographics, health behaviors, and the proposed mediators, SAIM remained modestly associated with sexual dissatisfaction (OR 1.32, 95% CI 1.02, 1.72). The increased odds of sexual dissatisfaction associated with SAIM in the fully adjusted model dropped from 1.78 to 1.32, reflecting a clinically important 26% reduction in the effect estimate. Of note, age, marital status, race, and family income were not significantly associated with sexual satisfaction. Participants who reported lower educational attainment were less likely to report sexual dissatisfaction. Neither adverse health behavior

Table 2. Association of Proposed Mediators with Both Exposure (SAIM) and Outcome (Sexual Dissatisfaction) among Women Veteran Outpatients

	Age-adjusted OR (95% CI)*	
Proposed mediator	*SAIM*	*Sexual dissatisfaction*
SF-36 mental health composite score (per 1 SD decrease)	1.69 (1.54, 1.85)	2.08 (1.92, 2.27)
SF-36 physical health composite score (per 1 SD decrease)	1.28 (1.18, 1.41)	1.23 (1.14, 1.33)
Any gynecological problem	1.98 (1.57, 2.51)	1.77 (1.45, 2.15)
No close partner	1.40 (1.16, 1.68)	2.87 (2.42, 3.41)

*All $p < 0.001$.

Table 3. Odds Ratios of Sexual Dissatisfaction Given SAIM in Women Veteran Outpatients Nested Logistic Regression Models

	*OR**	*95% CI*
Model 1: Unadjusted	2.26	1.91, 2.68
Model 2: Demographics and health behaviors	1.78	1.42, 2.23
Model 2 plus SF-36 mental health composite score	1.42	1.11, 1.83
Model 2 plus SF-36 physical health composite score	1.76	1.39, 2.24
Model 2 plus any gynecological problem	1.72	1.37, 2.16
Model 2 plus no close partner	1.70	1.35, 2.14
Model 3: Model 2 plus all 4 proposed mediators	1.32	1.02, 1.72

*All $p < 0.05$.

of smoking or alcohol abuse was significantly associated with sexual dissatisfaction.

Each of the proposed mediators was independently associated with sexual dissatisfaction in the final model. A 1-SD decrement in mental health composite score was associated with 1.91 greater odds of sexual dissatisfaction and a 1-SD decrement in the physical health composite score was associated with 1.27 greater odds of sexual dissatisfaction. Endorsing a gynecological problem was associated with 1.39 greater odds, and lack of a close partner was associated with 2.51 greater odds of sexual dissatisfaction. Differences in the magnitudes of the effect estimates should be interpreted with caution in this model, which includes both measurement and categorical variables.

Discussion

In a sample of over 3000 female veterans, increased mental health-related quality of life, increased physical health-related quality of life, lack of gynecological morbidity, and having a close partner largely mediate the adverse effect of SAIM on sexual satisfaction. Decreased mental health-related quality of life was the most prominent mediator in the association between SAIM and sexual dissatisfaction.

The dominant mediator role of mental health-related quality of life has been suggested in studies of specific mental health conditions relevant for survivors of SAIM, such as posttraumatic stress disorder (PTSD) and depression. An association between PTSD and decreased sexual satisfaction has been directly described in male populations,[19,41–43] even among those with nonsexual trauma. Among women, the link between sexual dissatisfaction and PTSD is not as well described. However, sexual dysfunction is described among female populations with PTSD,[20] and sexual dissatisfaction as a manifestation of sexual dysfunction may be more prominent in survivors of sexual assault, with resultant PTSD. Avoidant behavior with respect to sexual activity and intrusive negative thoughts when attempting consensual sexual activity[44] may lead to sexual dissatisfaction. Sexual assault is more strongly associated than other types of physical assault with the development of PTSD in women.[13,15]

Depression likely contributes to lower mental health-related quality of life after SAIM.[21] Hypoactive sexual desire and anhedonia, which may manifest as decreased sexual satisfaction, are common features in depression.[45] Women with a history of violence victimization are more likely than their male counterparts to manifest depression.[46] The independent contributions of PTSD and depression as potential mental health-related mediators may be difficult to establish because of the high rate of cooccurrence of these psychiatric conditions among women veterans.[47] Thus, emotional health-related quality of life may be an ideal measure to capture the functional impact of these disorders.

Although we determined that decrements in physical health-related quality of life were statistically significant in our final adjusted model, this association showed a minimal clinical effect. Thus, our findings were concordant with the expected direction of association but not with the expected magnitude. This was unexpected because self-perceived health status has been associated with sexual satisfaction among women.[48] Moreover, sexuality can be adversely affected by a variety of chronic nongynecological medical conditions,[25,49,50] including respiratory disease,[51] renal disease requiring chronic dialysis,[52] neurological disease such as multiple sclerosis,[53,54] and endocrinopathies such as diabetes mellitus[55] and the metabolic syndrome,[56] and has been

Table 4. Odds Ratios for Sexual Dissatisfaction among Women Veteran Outpatients

Variable	*aOR*	*95% CI*	
SAIM**	**1.32**	**1.02**	**1.72**
Age (per 1 year increase)	1.00	0.99	1.01
Marital status			
Divorced	1.07	0.79	1.45
Never married	1.22	0.88	1.68
Widowed	0.97	0.54	1.73
Married	Ref.	—	—
Race			
Black	1.00	0.75	1.34
Other	0.88	0.55	1.40
White	Ref.	—	—
Education**			
Less than high school	**0.24**	**0.06**	**0.95**
High school only	**0.59**	**0.38**	**0.92**
Some college	0.84	0.57	1.24
Completed college	Ref.	—	—
Income			
<$20,000	1.05	0.70	1.57
$20,000 – $49,999	1.16	0.80	1.69
Don't Know	0.79	0.42	1.48
$50,000+	Ref.	—	—
Smoking			
Current	0.91	0.70	1.20
Former	0.94	0.96	1.26
Never	Ref.	—	—
Alcohol abuse	1.42	0.95	2.12
Proposed Mediators			
SF-36 mental health composite score (per 1 SD[a] decrease)**	**1.91**	**2.17**	**1.68**
SF-36 physical health composite score (per 1 SD[a] decrease)**	**1.27**	**1.42**	**1.13**
Any gynecological problem**	**1.39**	**1.01**	**1.90**
No close partner**	**2.51**	**1.92**	**3.29**

**$p < 0.05$.
[a]MCS, 1 SD = 13.0; PCS, 1 SD = 12.3.

associated with poorer self-reported health among women with heart disease.[57] Thus, we anticipated that women with greater decrements in physical health-related quality of life, associated with the experience of SAIM, would report decreased sexual satisfaction.

One explanation for the limited role of physical health-related quality of life as a mediator is that our sample was relatively homogeneous with respect to health status. Further, because physical health-related quality of life is less important for sexual functioning in women than in men,[58] women may be less likely to report decreased sexual satisfaction due to decreased health-related quality of life.

Endorsement of a gynecological problem was likewise associated with a modest increase in odds of sexual dissatisfaction. Although this finding is concordant with our expected direction of association, we anticipated a greater overall impact on our effect estimate, given the extensive literature describing the adverse and long-standing impact of sexual assault on gynecological and sexual functioning.[27–29]

Much of the literature on sexual functioning in survivors of sexual assault does not directly address the construct of sexual satisfaction. Thus, whereas gynecological problems may manifest as sexual dysfunction in assault survivors, this sexual dysfunction may not be reported as decreased sexual satisfaction but may be more appropriately measured as another form of sexual dysfunction, such as sexual pain disorders. Alternatively, the limited impact of gynecological problems on the association between SAIM and decreased sexual satisfaction may be explained by the relative homogeneity of our sample with respect to gynecological complaints. A substantial majority of both sexually satisfied and dissatisfied subjects endorsed any gynecological problem, suggesting that all women veterans may be at high risk for gynecological complaints. This finding deserves further study.

Not only having a close partner but having a long-standing stable partner is protective of sexual functioning.[24] Given the adverse impact of trauma on interpersonal relationships,[31] we anticipated that our close partner variable would be an important mediator of the association between SAIM and decreased sexual satisfaction. The minimal independent contribution of this variable to the association suggests that although it functions as an intermediary, the adverse mental health-related quality of life is the predominant mediator.

Of note, we found no independent effect of marital status after controlling for having a close partner. Lack of a close partner was more strongly associated with sexual dissatisfaction in our model than was marital status. This suggests that the stability of interpersonal relationships, rather than partner availability, more strongly affects sexual satisfaction in this population, a finding reported in other work.[59]

Our findings with respect to educational status were surprising. Lower education in our cohort was correlated with increased sexual satisfaction. Prior work has described increased interest in sex and decreased painful sex with higher educational attainment.[58] In international samples,[5,60] higher educational attainment was correlated with greater sexual satisfaction, possibly due to more permissive sexual attitudes. Conversely, among Finnish women,[5] higher educational attainment was associated with female orgasmic dysfunction. The study of orgasmic dysfunction as a possible mediator of decreased sexual satisfaction was beyond the scope of this work but presents a target for future research.

The remaining sociodemographic variables were not significant in our final model, suggesting limited independent contribution of these to sexual satisfaction in a model that includes mental health-related quality of life.

Our work has several important strengths. Sexual satisfaction, an important dimension of health-related quality of life, has been minimally studied in women veterans, and further research on this topic may improve care for a growing population of women veterans. Although prior studies have described an association between sexual trauma and sexual dissatisfaction, our study extends this line of inquiry by proposing and testing mediators of this association.

An important limitation of our study is that our data are cross-sectional. Thus, causality cannot be established, and the directionality of association can only be inferred. Some directions of associations proposed in our mediator model may logically be reversed. For example, sexual dissatisfaction may lead to a lower emotional health-related quality of life or may impact the ability to form close partnerships instead of the converse. This cannot be tested in cross-sectional data, which preclude the establishment of temporal associations. Additionally, the data were collected by self-report; thus, the medical conditions and other histories collected cannot be verified. Further, we did not collect data on satisfaction with sexual frequency. Relational and physical aspects of sexual satisfaction may logically be compromised by partner unavailability. Further, women voluntarily abstaining from either partnered or unpartnered sexual activity may report high sexual satisfaction.

The VA Women's Health Project did not collect data on premilitary sexual violence, including childhood sexual abuse, or ongoing interpersonal violence, including intimate partner violence (IPV). Thus, a potentially unmeasured confounder in these associations is nonmilitary violence. Sexual revictimization is common, affecting approximately two thirds of rape victims, and there is evidence that the adverse health effects of multiple victimizations are cumulative.[61] Thus, among subjects with a history of multiple victimizations, premilitary or ongoing interpersonal violence may account for these findings.

We cannot adequately assess the relative contribution of antidepressant medication on sexual satisfaction in these data. Antidepressant medications, especially selective serotonin reuptake inhibitor (SSRI) antidepressants, contribute to sexual dysfunction in women.[62] Given the burden of psychiatric illness in this population, it is possible that a large proportion of these women were prescribed antidepressant medications. Some women may have experienced adverse sexual side effects of SSRI medication and were dissatisfied because of this. Regardless, among women who are appropriately treated for psychiatric illnesses that have resulted from trauma, the association between SAIM, and decreased sexual satisfaction may be attenuated due to improved emotional health-related quality of life.

Finally, the sexual satisfaction construct is based on a single, face-valid question designed to capture all domains of sexual satisfaction. A single, face-valid question to measure overall sexual satisfaction has been related previously to functional, relational, and emotional aspects of sexuality.[2,34] A

multiple-question, psychometrically validated instrument designed to capture operationally different domains of sexual satisfaction was not available in these data. Indeed, the study of sexuality is often problematic because studies that enroll subjects willing to answer detailed multiquestion surveys addressing sexual issues may not represent the population under study.[63] Conversely, surveys designed to measure a broad range of topics, such as the one reported here, may be more representative of the study population but provide limited ability to elaborate details of the domains of sexual constructs. Future research should establish if the proposed mediators impact more strongly on operationally distinct domains of sexual satisfaction.

After controlling for mediators, a modest but significant association between SAIM and decreased sexual satisfaction remained. This suggests that the four proposed mediators do not fully describe the association between SAIM and sexual dissatisfaction. Elaboration of additional mediators presents an important target for further research.

Conclusions

Women veterans are exposed at high rates to sexual trauma, which adversely impacts sexual satisfaction, an important aspect of health-related quality of life. Much of the adverse impact of SAIM on sexual satisfaction in women veterans is explained by decrements in emotional health-related quality of life. The Department of Defense reported that 1620 service members were victims of SAIM during fiscal year 2007.[64] As the population of women veterans increases in both the VA and the community, the burden of SAIM may become more evident. Healthcare providers treating survivors of SAIM should focus on treatable sequelae of this adverse life event, including mental health consequences. Providers who recognize adverse mental health disorders in SAIM survivors should screen patients for adverse sexual effects in all domains of sexual satisfaction: emotional, physical including gynecological, and relational. In addition to primary prevention of SAIM, efforts to both screen for and then appropriately treat the sequelae of SAIM may improve patient satisfaction and overall well-being of women veterans.

Acknowledgments

Data collection for the VA Women's Health Project (Katherine Skinner, PI, deceased) was supported by VA HSR&D SDR 93–101. We gratefully honor the memory of Dr. Skinner and her contributions to the health of women veterans by continuing her research.

J.S.M.-H. was supported by a Department of Veterans Affairs Special Fellowship in the Health Issues of Women Veterans while this research was performed. J.M.L. was supported by National Institute on Drug Abuse K23 DA016665. A.S. was supported by a Department of Veterans Affairs Clinical Sciences R&D Merit Review. The conclusions are those of the authors and do not necessarily reflect the opinions of the Department of Veterans Affairs or the National Institute on Drug Abuse.

Preliminary results of this work were presented at the Society of General Internal Medicine 30th Annual Meeting, Toronto, Ontario, Canada, on April 26, 2007.

Disclosure Statement

The authors have no conflicts of interest to report.

References

1. Young M, Denny G, Young T, et al. Sexual satisfaction among married women age 50 and older. Psychol Rep 2000;86:1107–1122.
2. Bridges SK, Lease SH, Ellison CR. Predicting sexual satisfaction in women: Implications for counselor education and training. J Couns Dev 2004;82:158–166.
3. Byers ES, MacNeil S. Further validation of the interpersonal exchange model of sexual satisfaction. J Sex Marital Ther 2006;32:53–69.
4. Haning RV, O'Keefe SL, Randall EJ, et al. Intimacy, orgasm likelihood, and conflict predict sexual satisfaction in heterosexual male and female respondents. J Sex Marital Ther 2007;33:93–113.
5. Haavio-Mannila E, Kontula O. Correlates of increased sexual satisfaction. Arch Sex Behav 1997;26:399–419.
6. Lawrence K, Byers E. Sexual satisfaction in heterosexual long-term relationships: The interpersonal exchange model of sexual satisfaction. Pers Relatsh 2005:195;2:267–285.
7. Hays J, Ockene JK, Brunner RL, et al. Effects of estrogen plus progestin on health-related quality of life. N Engl J Med 2003;348:1839–1854.
8. Rosen RC, Bachmann GA. Sexual well-being, happiness, and satisfaction, in women: The case for a new conceptual paradigm. J Sex Marital Ther 2008;34:291–297.
9. Skinner KM, Furey J. The focus on women veterans who use Veterans Administration health care: The Veterans Administration Women's Health Project. Mil Med 1998;163:761–766.
10. Skinner KM, Kressin NR, Frayne S, et al. The prevalence of military sexual assault among female Veterans' Administration outpatients. J Interpers Violence 2000;15:291–310.
11. Frayne SM, Skinner KM, Sullivan LM, et al. Medical profile of women Veterans Administration outpatients who report a history of sexual assault occurring while in the military. J Womens Health Gend Based Med 1999;8:835–845.
12. Hankin CS, Skinner KM, Sullivan LM, Miller DR, Frayne S, Tripp TJ. Prevalence of depressive and alcohol abuse symptoms among women VA outpatients who report experiencing sexual assault while in the military. J Trauma Stress 1999;12:601–612.
13. Faravelli C, Giugni A, Salvatori S, Ricca V. Psychopathology after rape. Am J Psychiatry 2004;161:1483–1485.
14. Kang H, Dalager N, Mahan C, Ishii E. The role of sexual assault on the risk of PTSD among Gulf War veterans. Ann Epidemiol 2005;15:191–195.
15. McFarlane J, Malecha A, Watson K, et al. Intimate partner sexual assault against women: Frequency, health consequences, and treatment outcomes. Obstet Gynecol 2005;105:99–108.
16. Suris A, Lind L, Kashner TM, Borman PD, Petty F. Sexual assault in women veterans: An examination of PTSD risk, health care utilization, and cost of care. Psychosom Med 2004;66:749–756.
17. Yaeger D, Himmelfarb N, Cammack A, Mintz J. DSM-IV diagnosed posttraumatic stress disorder in women veterans with and without military sexual trauma. J Gen Intern Med 2006;21S:S65–69.

18. Fontana A, Rosenheck R. Duty-related and sexual stress in the etiology of PTSD among women veterans who seek treatment. Psychiatr Serv 1998;49:658–662.
19. Cosgrove DJ, Gordon Z, Bernie JE, et al. Sexual dysfunction in combat veterans with post-traumatic stress disorder. Urology 2002;60:881–884.
20. Letourneau EJ, Resnick HS, Kilpatrick DS, Best E. Comorbidity of sexual problems and posttraumatic stress disorder in female crime victims. Behav Ther 1996;27:321–336.
21. Baldwin DS. Depression and sexual dysfunction. Br Med Bull 2001;57:81–99.
22. Bartlik B, Kocsis JH, Legere R, Villaluz J, Kossoy A, Gelenberg AJ. Sexual dysfunction secondary to depressive disorders. J Gend Specif Med 1999;2:52–60.
23. Sadler AG, Booth BM, Nielson D, Doebbeling BN. Health-related consequences of physical and sexual violence: Women in the military. Obstet Gynecol 2000;96:473–480.
24. Laumann EO, Nicolosi A, Glasser DB, et al. Sexual problems among women and men aged 40–80 years: Prevalence and correlates identified in the Global Study of Sexual Attitudes and Behaviors. Int J Impot Res 2005;17:39–57.
25. Wilmoth MC. Sexuality: A critical component of quality of life in chronic disease. Nurs Clin North Am 2007;42: 507–514.
26. Lewis RW, Fugl-Meyer KS, Bosch R, et al. Epidemiology/risk factors of sexual dysfunction. J Sex Med 2004;1:35–39.
27. Golding JM. Sexual assault history and women's reproductive and sexual health. Psychol Women Q 1996;20:101–121.
28. Campbell R, Lichty LF, Sturza M, Raja S. Gynecological health impact of sexual assault. Res Nurs Health 2006;29: 399–413.
29. Letourneau EJ, Holmes M, Chasedunn-Roark J. Gynecologic health consequences to victims of interpersonal violence. Womens Health Issues 1999;9:115–120.
30. Reiter RC, Gambone JC. Demographic and historic variables in women with idiopathic chronic pelvic pain. Obstet Gynecol 1990;75:428–432.
31. Goff BSN, Crow JR, Reisbig AMJ, Hamilton S. The impact of individual trauma symptoms of deployed soldiers on relationship satisfaction. J Fam Psychol 2007;21:344–353.
32. Walker LEA. Battered woman syndrome: Empirical findings. Ann NY Acad Sci 2006;1087:142–157.
33. Yeh HC, Lorenz FO, Wickrama KA, et al. Relationships among sexual satisfaction, marital quality, and marital instability at midlife. J Fam Psychol 2006;20:339–343.
34. Davison SL, Bell RJ, LaChina M, et al. Sexual function in well women: Stratification by sexual satisfaction, hormone use, and menopause status. J Sex Med 2008;5:1214–1222.
35. McHorney CA, Ware JE Jr, Lu JF, et al. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994;32:40–66.
36. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–263.
37. Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–483.
38. Guerrero LK, Andersen PA, Afifi WA. Close encounters: Communication in Relationships, 2nd ed. Sage Publications, Thousand Oaks, CA, 2007.
39. Kaboli PJ, Doebbeling BN, Saag KG, Rosenthal GE. Use of complementary and alternative medicine by older patients with arthritis: A population-based study. Arthritis Rheum 2001;45:398–403.
40. Cohen J. Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, NJ: L. Erlbaum, 1988.
41. Kaplan PM. Post-traumatic stress syndrome and sexual dysfunction. J Sex Marital Ther 1989;15:74–77.
42. Kotler M, Cohen H, Aizenberg D, et al. Sexual dysfunction in male posttraumatic stress disorder patients. Psychother Psychosom 2000;69:309–315.
43. Letourneau EJ, Schewe PA, Frueh BC. Preliminary evaluation of sexual problems in combat veterans with PTSD. J Trauma Stress 1997;10:125–132.
44. van Berlo W, Ensink B. Problems with sexuality after sexual assault. Annu Rev Sex Res 2000;11:235–257.
45. Dobkin RD, Leiblum SR, Rosen RC, Menza M, Marin H. Depression and sexual functioning in minority women: Current status and future directions. J Sex Marital Ther 2006;32:23–36.
46. Porcerelli JH, Cogan R, West PP, et al. Violent victimization of women and men: Physical and psychiatric symptoms. J Am Board Fam Med 2003;16:32–39.
47. Frayne SM, Seaver MR, Loveland S, et al. Burden of medical illness in women with depression and posttraumatic stress disorder. Arch Intern Med 2004;164:1306–1312.
48. Walfisch S, Maoz B, Antonovsky H. Sexual satisfaction among middle-aged couples: Correlation with frequency of intercourse and health status. Maturitas 1984;6:285–296.
49. Kralik D, Koch T, Telford K. Constructions of sexuality for midlife women living with chronic illness. J Adv Nurs 2001; 35:180–187.
50. McInnes RA. Chronic illness and sexuality. Med J Aust 2003;179:263–266.
51. Goodell TT. Sexuality in chronic lung disease. Nurs Clin North Am 2007;42:631–638.
52. Coelho-Marques FZ, Wagner MB, Poli de Figueiredo CE, d'Avila DO. Quality of life and sexuality in chronic dialysis female patients. Int J Impot Res 2006;18:539–543.
53. Salonia A, Munarriz RM, Naspro R, et al. Women's sexual dysfunction: A pathophysiological review. BJU Int 2004;93:1156–1164.
54. Koch T, Kralik D, Eastwood S. Constructions of sexuality for women living with multiple sclerosis. J Adv Nurs 2002;39: 137–145.
55. Salonia A, Lanzi R, Scavini M, et al. Sexual function and endocrine profile in fertile women with type 1 diabetes. Diabetes Care 2006;29:312–316.
56. Esposito N. Women with a history of sexual assault. Health care visits can be reminders of a sexual assault. Am J Nurs 2006;106:69–71.
57. Addis IB, Ireland CC, Vittinghoff E, Lin F, Stuenkel CA, Hulley S. Sexual activity and function in postmenopausal women with heart disease. Obstet Gynecol 2005;106:121–127.
58. Laumann EO, Paik A, Rosen RC. Sexual dysfunction in the United States: Prevalence and predictors. JAMA 1999;281: 537–544.
59. Laumann EO, Youm Y. Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: A network explanation. Sex Transm Dis 1999;26:250–261.
60. Barrientos JE, Paez D. Psychosocial variables of sexual satisfaction in Chile. J Sex Marital Ther 2006;32:351–368.
61. Classen CC, Palesh OG, Aggarwal R. Sexual revictimization: A review of the empirical literature. Trauma Violence Abuse 2005;6:103–129.

62. Hensley PL, Nurnberg HG. SSRI sexual dysfunction: A female perspective. J Sex Marital Ther 2002;28 (Suppl 1): 143–153.
63. Hayes R, Dennerstein L. The impact of aging on sexual function and sexual dysfunction in women: A review of population-based studies. J Sex Med 2005;2:317–330.
64. Department of Defense, Sexual Assault Prevention and Response Office. FY07 Report on sexual assault in the military. United States Department of Defense, 2008.

Address reprint requests to:
Jennifer McCall-Hosenfeld, M.D., M.Sc.
Pennsylvania State University
Milton S. Hershey Medical Center
Division of General Internal Medicine
500 University Drive/MC: HU15
Hershey PA 17033

E-mail: jsm31@psu.edu

Differential diagnosis of chest pain in adults

Author James L Meisel, MD, FACP
Section Editor Mark D Aronson, MD
Deputy Editor Pracha Eamranond, MD, MPH

Last literature review for version 17.3: September 30, 2009 | **This topic last updated:** October 6, 2009

INTRODUCTION — The differential diagnosis of patients presenting with chest pain is extensive, ranging from benign musculoskeletal etiologies to life-threatening cardiac disease. Many of the diseases that cause chest pain are reviewed in detail elsewhere. This topic will discuss the differential diagnosis of chest pain in an approximate order of prevalence seen in primary care practice. Within each subsection, diseases that may pose an immediate life-threat are discussed first, followed by the more common etiologies, and then by other causes of chest pain. Details about the office evaluation of the patient with chest pain are found separately. (See "Diagnostic approach to chest pain in adults".)

CHEST WALL PAIN — Chest wall causes of pain are among the most common etiologies of chest pain seen by primary care clinicians, accounting for 36 percent of episodes in one report (table 1A-B) [1]. Chest wall tenderness may present concomitantly with myocardial ischemia; the latter should be considered first in any patient at risk by age, history, or associated symptoms [1]. Causes of true chest wall pain may be musculoskeletal or related to the skin and sensory nerves.

Musculoskeletal pain — Demographic features, characteristics of the chest pain, and associated symptoms may favor the diagnosis of musculoskeletal chest pain or suggest other causes of chest discomfort (table 2). As an example, the patient may describe a history of repetitive or unaccustomed activity involving the upper trunk or arms. Certain characteristics of the chest pain or associated symptoms may suggest a nonmusculoskeletal origin. (See "Clinical evaluation of musculoskeletal chest pain".)

Musculoskeletal chest pain is often insidious and persistent, lasting for hours to weeks. It is frequently sharp and localized to a specific area (such as the xiphoid, lower rib tips, or midsternum), but may be diffuse and poorly localized. The pain may be positional or exacerbated by deep breathing, turning, or arm movement; the first two, however, are also noted in a variety of visceral processes, particularly those involving the pleura and pericardium.

The proportion of patients with chest pain having a musculoskeletal source varies with the clinical setting. It is more common in ambulatory patients presenting to their primary care clinician than presenting to an emergency department (table 3). It also occurs more frequently among women than men. One study examined the incidence of musculoskeletal chest pain in 122 consecutive patients presenting to an emergency department with chest pain [2]. Of 36 patients diagnosed with costochondritis, 69 percent were women. By comparison, women represented only 31 percent of the presenting patients who did not have a subsequent diagnosis of costochondritis.

The differential diagnosis of musculoskeletal chest pain has been divided into three categories (see "Major causes of musculoskeletal chest pain":

- Isolated musculoskeletal chest pain syndromes (costosternal, posterior chest wall syndromes)
- Rheumatic diseases
- Nonrheumatic systemic diseases

Isolated musculoskeletal chest pain syndromes — There are a number of chest wall syndromes with chest pain associated with musculoskeletal inflammation (table 4).

- "Costochondritis" is one of the more common presentations of musculoskeletal chest pain. It is a diffuse pain syndrome, in which multiple areas of tenderness are found that reproduce the described pain. The upper costal cartilages at the costochondral or costosternal junctions are most frequently involved. The areas of tenderness are not accompanied by heat, erythema, or localized swelling.

- Chest wall pain occurring after coronary artery bypass surgery may be a result of incisional discomfort, of internal mammary artery grafting, or related to sternal wires [3].

- Costovertebral joint dysfunction syndrome is an uncommon condition that causes posterior chest wall pain and may mimic a pulmonary embolism. Thoracic disk herniation is another unusual cause of posterior chest pain; the pain is sometimes dermatomal and "band-like," and retrosternal or retrogastric pain has also been described [3,4].

Rheumatic diseases — Involvement of thoracic joints in rheumatic diseases can be associated with musculoskeletal chest wall pain (table 5). Examples include rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, and fibromyalgia. (See "Major causes of musculoskeletal chest pain".)

Nonrheumatic systemic diseases — A variety of systemic disorders can be characterized by bony involvement that can lead to chest wall pain (table 5). Examples include stress fractures [5], neoplasms including pathologic fractures [6], sickle cell anemia, and infections such as septic arthritis and osteomyelitis.

Skin and sensory nerves — Chest pain may be the presenting symptom of herpes zoster (shingles); it may precede the characteristic rash and, rarely, zoster may occur without a rash [7]. Dysesthesia is usually present in the affected dermatome. Postherpetic and postradiation neuralgia are other unusual causes of chest pain. (See "Postherpetic neuralgia".)

CARDIAC CAUSES OF CHEST PAIN — Cardiac causes of chest pain may be related to myocardial ischemia resulting from coronary heart disease, aortic dissection, valvular heart disease, inflammation of the myocardium or pericardium, or reversible left ventricular dysfunction due to emotional stress. In cardiac syndrome X ischemic type chest pain may occur in the absence of documented epicardial coronary artery disease. (See 'Cardiac syndrome X' below.)

Coronary heart disease — Ischemic chest pain due to coronary artery disease (CAD) encompasses a spectrum of presentations including stable angina pectoris, unstable angina, non-ST elevation myocardial infarction, and ST elevation myocardial infarction. (See "Classification of

unstable angina and non-ST elevation myocardial infarction".) Patients classically complain of chest heaviness, pressure, tightness or burning, but may vigorously deny "pain." (see "Pathophysiology and clinical presentation of ischemic chest pain". Other descriptions such as provocation with physical or emotional stress or cold, relief with rest, and radiation to the neck, jaw, and shoulder are common. Ischemic pain usually lasts more than 2 but less than 20 minutes, unless a myocardial infarction is occurring. Associated symptoms may include dyspnea, nausea and vomiting, diaphoresis, presyncope, or palpitations.

This classical description of pain due to CAD is most frequently seen in middle-aged men with risk factors for atherosclerosis; women, patients with diabetes, and the elderly often do not present with classical symptoms. As an example, one study of 94 patients found that 32 percent, especially women over age 65, presented with "atypical" symptoms; abdominal pain was most common, occurring in one-third of these patients, while paroxysmal dyspnea was the presenting symptom in 17 percent [8]. Such presentations do not necessarily suggest a better prognosis [9]. Even elderly women diagnosed with "nonspecific chest pain" may be at increased risk of cardiac morbidity [10]. The term "atypical" chest pain should be avoided; it increases the risk of misdiagnosing women, the elderly, patients with long standing diabetes, and those with a myocardial infarction who present with symptoms such as dyspnea or postprandial epigastric pain. (See "Clinical features and diagnosis of coronary heart disease in women".)

Patients with chest pain who use cocaine are at increased risk of acute coronary syndrome [11]. In one series, approximately one-fifth of patients with cocaine-associated chest pain experienced an acute coronary syndrome [12]. The differential diagnosis of chest pain in patients who have used cocaine is similar to that in the general population except that the likelihood of the patient having a serious event is increased [13]. (See "Cocaine: Acute intoxication".)

One interesting group of patients is those with variant angina, in which coronary vasospasm may result in classical anginal pain. It may be precipitated by hyperventilation and, occasionally, by exercise. Vasospastic angina may be associated with life-threatening arrhythmias. Patients are typically less than 60 years old and, other than cigarette or cocaine use, do not necessarily have traditional cardiovascular risk factors. The rest ECG may reveal transient ST segment elevation. (See "Variant angina".)

Spontaneous coronary artery dissection is a rare but often deadly cause of myocardial infarction that mainly affects otherwise healthy, young females [14].

Aortic dissection — Sudden, severe, and often migratory chest pain occurs in most but not all patients with dissection of the ascending or descending aorta, but its diagnosis often requires a high index of suspicion [15]. It is most common in men older than age 60. The pain typically is cataclysmic in onset, and is often described as a "ripping" or "tearing" sensation. Pain is commonly felt in the anterior or posterior chest, or in the neck, throat, or jaw. Hypertension is the most important risk factor; less common associations include Marfan's syndrome, congenital bicuspid and unicommisural aortic valves, aortic coarctation and, rarely, pregnancy [16,17]. Cocaine use may precipitate aortic dissection [18].

Aortic dissection should be considered in any patient who presents with a catastrophic illness associated with hypertension, an aortic murmur, and unexplained physical findings of vascular

origin. The presentation may be subtle, however; in one series, one-third of cases were initially misdiagnosed [17].

Symptoms associated with aortic dissection may be related to impaired blood flow to an organ or limb induced by the original dissection or by propagation of the dissection proximally or distally. Even in the absence of chest pain, findings that should prompt consideration of aortic dissection include, in approximately decreasing order of incidence:

- Abnormal aortic contour or widened mediastinum on chest radiograph, seen in most patients (picture 1)
- Congestive heart failure, which may be due to acute aortic insufficiency (movie 1)
- Neurologic deficits, including paraplegia, stroke, or decreased consciousness (due to spinal cord ischemia, dissection into the carotid arteries, or diminished carotid blood flow, respectively)
- Syncope, cardiac tamponade, and sudden death due to rupture of the aorta into the pericardial space
- Shock, hemothorax, and exsanguination, which may result if the dissection extends through the adventitia, with hemorrhage into the pleural space
- Acute lower extremity ischemia due to dissection into the iliac vessels
- Infrequently reported sequelae such as myocardial ischemia due to coronary occlusion, signs of mesenteric or renal ischemia, Horner syndrome due to compression of the superior cervical sympathetic ganglion, vocal cord paralysis due to compression of the left recurrent laryngeal nerve, and other rare physical findings

The initial evaluation to rule out dissection should involve confirming that chest pain is not typical for this disorder, that there are no new neurologic symptoms, checking that pulses and upper extremity blood pressures are symmetric and normal and that the murmur of aortic regurgitation is not present, excluding heart failure on exam, and determining that the aortic contour is not widened on chest radiograph. A normal electrocardiogram may raise concern that chest pain is not due to an acute myocardial infarction. Definitive diagnosis is made with aortography or noninvasive techniques including CT scanning, magnetic resonance imaging, and transesophageal echocardiography. (See "Clinical manifestations and diagnosis of aortic dissection".)

Valvular heart disease — Significant valvular abnormalities, particularly of the aortic or mitral valves, may present with chest pain.

- Aortic stenosis should be considered whenever a patient presents with progressive angina, dyspnea, and/or syncope. A detailed physical examination of the heart should be performed to exclude "weak and delayed" arterial pulses, a sustained apical impulse, and characteristic auscultatory findings (movie 2). (See "Auscultation of cardiac murmurs" and "Auscultation of heart sounds".) The ECG may reveal left ventricular hypertrophy. It is important to obtain an echocardiogram in patients with suspected aortic stenosis early in the evaluation since exercise stress testing may be contraindicated. (See "Pathophysiology and clinical features of valvular aortic stenosis in adults".)

- Patients with mitral stenosis infrequently experience chest pain. The pain often resembles angina and, although it is most commonly the result of pulmonary hypertension and right ventricular hypertrophy, may be due to underlying coronary artery disease or a coronary artery

embolism [19]. An atrial tachyarrhythmia with left atrial and pulmonary vascular distension is another cause of intermittent chest pain in mitral stenosis. (See "Overview of pulmonary hypertension".)

- Valvular pulmonic stenosis is a relatively common congenital defect, but a rare cause of chest pain in primary care practice. Pulmonic stenosis occurring later in life is associated with the carcinoid syndrome. (See "Clinical features of the carcinoid syndrome" and "Clinical manifestations and diagnosis of pulmonic stenosis".)

Pericarditis — The major clinical manifestations of acute pericarditis are chest pain (usually pleuritic), a pericardial friction rub, and widespread ST segment elevation on the electrocardiogram [20]. (See "Auscultation of heart sounds".) At least two of these features, with or without an accompanying pericardial effusion, are usually present.

The chest pain of acute pericarditis is typically of fairly sudden onset and occurs over the anterior chest. It is usually sharp and exacerbated by inspiration. However, dull, oppressive pain, difficult to distinguish from that of myocardial infarction, can also occur. The pain may decrease in intensity when the patient sits up and can radiate, especially to the trapezius ridge. (See "Evaluation and management of acute pericarditis".)

Myocarditis — Myocarditis may present with both cardiac and systemic symptoms. When chest pain occurs, it is usually associated with concomitant pericarditis although evidence of infarction may be seen. Systemic symptoms include fever, myalgias, and muscle tenderness. (See "Clinical manifestations and diagnosis of myocarditis in adults".)

Stress-induced cardiomyopathy — Emotional stress can precipitate severe, reversible left ventricular dysfunction in patients without coronary heart disease, probably related to exaggerated sympathetic stimulation. Patients most commonly present with acute substernal chest pain. (See "Stress-induced (takotsubo) cardiomyopathy".)

Cardiac syndrome X — Cardiac syndrome X is a syndrome of angina-like, non-gastrointestinal chest pain associated with normal coronary arteries; it is most commonly seen in premenopausal women [21]. The pain is typical of angina in approximately one-half of patients and may be precipitated by exertion, although it also occurs at rest. The pain also often has characteristics atypical for epicardial CAD; it is more severe, prolonged, and is variably relieved with nitrates [22,23]. An association with an underlying panic disorder has been described [23]. ST segment depression may be seen on treadmill exercise testing.

The diagnosis is one of exclusion, generally made if coronary angiography does not demonstrate evidence of CAD. In referral populations of women with chest pain (eg, hospitalization, undergoing coronary angiography), myocardial ischemia and/or coronary microvascular dysfunction are present in 20 to 50 percent of patients with normal coronary arteries. (See "Cardiac syndrome X: Angina pectoris with normal coronary arteries".)

Pheochromocytoma — Pheochromocytoma is a catecholamine-secreting tumor that rarely presents with chest pain [24]. It should particularly be considered in patients who lack other coronary risk factors; have other symptoms suggestive of pheochromocytoma (eg, severe or

paroxysmal hypertension, headache, or generalized sweating or palpitations); whose symptoms are paroxysmal or worsen after the administration of beta adrenergic blockage; or who appear to be experiencing demand ischemia [25]. (See "Clinical presentation and diagnosis of pheochromocytoma" and "Elevated cardiac troponin concentration in the absence of an acute coronary syndrome".)

GASTROINTESTINAL CAUSES OF CHEST PAIN — The heart and esophagus share similar neurologic innervation. Thus, it may be difficult to distinguish between chest pain due to myocardial ischemia and that originating from the esophagus based upon the history alone. Esophageal disease may cause symptoms thought "classical" for myocardial ischemia, including a sensation of chest pressure, provocation with exercise or emotion, palliation by rest or nitrates, or a crescendo pattern [26]. In one study of 28 patients referred to a cardiology clinic for chest pain, 36 percent ultimately had a diagnosis of reflux esophagitis [27]. After a history was taken independently by a cardiologist and a gastroenterologist, an accurate diagnosis was made in only 40 and 30 percent of patients with myocardial disease and reflux esophagitis, respectively.

Any patient at risk for CAD who presents with anginal-quality chest pain should have myocardial ischemia ruled out before being given a gastroenterologic diagnosis. Neither the clinical history nor the response of new chest pain to a "GI cocktail" (eg, viscous lidocaine and antacid) reliably differentiates the diagnoses, which often coexist [28]. There are, however, several clues that suggest an esophageal etiology (table 6) [26].

Gastroesophageal reflux disease — Chest pain due to gastroesophageal reflux disease (GERD) can mimic angina pectoris and may be described as squeezing or burning, located substernally and radiating to the back, neck, jaw or arms, lasting anywhere from minutes to hours, and resolving either spontaneously or with antacids. It may occur after meals, awaken patients from sleep, and be exacerbated by emotional stress. The preponderance of patients with reflux-induced chest pain also give a history of typical reflux symptoms (ie, dyspepsia, regurgitation, acid taste). (See "Clinical manifestations and diagnosis of gastroesophageal reflux in adults".) After cardiac disease has been ruled out, a trial of acid suppression may assist in the diagnosis of GERD [29-31]. (See "Chest pain of esophageal origin".)

Esophageal hyperalgesia — There are considerable experimental data to indicate that some patients with noncardiac chest pain have a lower threshold for esophageal pain than normal subjects. Studies utilizing intraesophageal balloon distension have shown that many patients with unexplained chest pain experience their pain at a lower volume of balloon inflation than that found in appropriate control subjects [32,33]. (See "Chest pain of esophageal origin", section on 'Esophageal hypersensitivity'.)

Some patients with cardiac syndrome X may have a similar problem in which there is altered sensory awareness to cardiac events, the so-called "sensitive heart" [34,35]. (See "Cardiac syndrome X: Angina pectoris with normal coronary arteries: Pathogenesis".)

Abnormal motility patterns and achalasia — The relatively uncommon diagnosis of a motility disorder or esophageal spasm should be entertained if chest pain is associated with dysphagia, and a barium swallow study does not reveal an anatomic abnormality of the esophagus. (See "Diffuse esophageal spasm and nutcracker esophagus".)

Esophageal rupture, mediastinitis, and foreign bodies — Spontaneous perforation of the esophagus most commonly results from a sudden increase in intraesophageal pressure combined with negative intrathoracic pressure caused by straining or vomiting (effort rupture of the esophagus or Boerhaave's syndrome) [36]. Other causes of perforation include caustic ingestion, pill esophagitis, Barrett's ulcer [37], infectious ulcers in patients with AIDS, and iatrogenic injury (typically from passing an instrument or tube through the esophagus or following dilation of esophageal strictures). (See "Complications of esophageal stricture dilation".)

The classical history in a patient with Boerhaave's syndrome is one of severe retching and vomiting followed by excruciating retrosternal chest and upper abdominal pain. Odynophagia, tachypnea, dyspnea, cyanosis, fever, and shock develop rapidly thereafter [9,36]. (See "Boerhaave's syndrome: Effort rupture of the esophagus".)

A patient with a foreign body impacted in the esophagus may present with chest pain. Esophageal foreign bodies are most commonly seen in children. Adult patients are often prisoners, edentulous, or are mentally retarded, psychiatrically ill, or dependent on alcohol [38]. (See "Foreign bodies in the esophagus in adults".)

Medication-induced esophagitis — Medications can induce esophageal abnormalities via both systemic and local actions. The types of medication causing direct esophageal injury can be roughly divided into antibiotics (most commonly doxycycline), antiinflammatory agents (especially aspirin), and others including bisphosphonates, potassium chloride, quinidine preparations, and iron compounds in the United States; emepronium, alprenolol, and pinaverium are common etiologies in other countries. The typical patient with medication-induced esophagitis does not have a history of prior esophageal disease. Patients will often present with the sudden onset of odynophagia and retrosternal pain; the pain may be so severe that swallowing saliva is difficult. Patients often relate (after careful questioning) the onset of symptoms to the swallowing of a pill without water, commonly at bedtime. (See "Medication-induced esophagitis".)

Other gastrointestinal causes of chest pain — The possibility of radiating or referred visceral pain due to peptic ulcer disease, cholecystitis or biliary colic [39], pancreatitis, kidney stones or even appendicitis [40] should be considered in any patient with unexplained chest pain.

PULMONARY CAUSES OF CHEST PAIN — The pulmonary causes of chest pain may be related to the pulmonary vessels, lung parenchyma, or pleural tissue.

Pulmonary vasculature — Chest pain caused by abnormalities of the pulmonary vessels may be due to an acute problem such as a pulmonary embolism or a chronic condition such as pulmonary hypertension.

Acute pulmonary thromboembolism — The diagnosis of acute pulmonary embolism often requires a high index of suspicion, especially since it is an uncommon cause of chest pain in the primary care setting [1,41,42]. It should be considered in any patient who presents with chest pain that is usually but not necessarily pleuritic in nature or dyspnea which is not fully explained by the clinical evaluation, chest radiograph, or electrocardiogram [43-45]. Individual symptoms and signs are not helpful diagnostically because their frequency is similar among patients with and without PE [46]. The Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) study

found that the most common symptoms of pulmonary embolism were dyspnea (73 percent), pleuritic chest pain (66 percent), cough (37 percent), and hemoptysis (13 percent) [45]. Ninety-seven percent of patients had pleuritic pain, dyspnea or tachypnea; 90 percent had dyspnea, tachypnea or signs of deep venous thrombosis; 84 percent had a chest x-ray abnormality; and 50 percent had nonspecific electrocardiographic abnormalities. Subsequent studies have corroborated these findings [47,48].

The PIOPED II study further elucidated that, if present, dyspnea may occur only on exertion. The onset of dyspnea is usually rapid. Orthopnea may occur. Occasionally, common symptoms may be absent or mild if present, even with severe pulmonary embolism. A low-probability objective clinical assessment does not exclude the diagnosis [46].

Most patients with pulmonary embolism have identifiable risk factors including immobilization, surgery within the last three months, stroke, history of venous thromboembolism, or malignancy (table 7) [49]. (See "Overview of the causes of venous thrombosis".)

Pulmonary hypertension and cor pulmonale — Patients with secondary pulmonary hypertension often have symptoms that reflect the underlying etiology (eg, chronic obstructive pulmonary disease, pulmonary embolic disease, collagen vascular disease). There are, however, symptoms directly attributable to secondary pulmonary hypertension including dyspnea on exertion, fatigue, lethargy, chest pain, and syncope with exertion. Typical exertional angina has been reported in patients with mitral stenosis or congenital heart disease and cor pulmonale even in the presence of normal coronary arteries [19]. The mechanism by which angina occurs is unclear; both pulmonary artery stretching and right ventricular ischemia have been proposed. (See "Overview of pulmonary hypertension".)

Idiopathic pulmonary arterial hypertension is a rare disease. Most patients present with exertional dyspnea, which is indicative of an inability to increase cardiac output with exercise. Exertional chest pain, syncope, and edema are indications of more severe pulmonary hypertension and impaired right heart function. (See "Overview of pulmonary hypertension".)

Lung parenchyma — Causes of chest pain related to the lung parenchyma include infection, cancer, or chronic diseases such as sarcoidosis.

Pneumonia — The patient with community acquired pneumonia (CAP) caused by pyogenic organisms classically presents with the sudden onset of rigors followed by fever, pleuritic chest pain, and cough productive of purulent sputum. Chest pain occurs in 30 percent of cases, chills in 40 to 50 percent, and rigors in 15 percent. Because of the rapid onset of symptoms, most individuals seek medical care within six days [50]. (See "Diagnostic approach to community-acquired pneumonia in adults".)

The Infectious Disease Society of America (IDSA) and American Thoracic Society (ATS) guidelines on community acquired pneumonia, as well as other ATS guidelines, can be accessed through the ATS web site at www.thoracic.org/sections/publications/statements/index.html.

Cancer — Isolated chest pain is a relatively rare presentation of lung cancer [1,41,42]. The chest pain experienced by 25 to 50 percent of lung cancer patients is usually in association with

cough, dyspnea, weight loss or hemoptysis. Some patients have a dull, intermittent pain on the side of the tumor; severe or persistent pain often indicates chest wall or mediastinal invasion. (See "Overview of the risk factors, pathology, and clinical manifestations of lung cancer" and "Evaluation of mediastinal masses".)

Sarcoidosis — Chest pain is a common manifestation of pulmonary sarcoidosis, although it rarely occurs in isolation; most commonly it is accompanied by cough and dyspnea. (See "Clinical manifestations and diagnosis of sarcoidosis".) Granulomatous involvement of the ventricular septum and conduction system of the heart can lead to a variety of arrhythmias (including heart block) and sudden death; such involvement may be heralded by chest pain, palpitations, syncope, or dizziness. (See "Cardiac sarcoidosis".)

Pleura and pleural space — Pleuritic chest pain is caused by irritation of nerve endings of pain fibers in the costal pleura. It often has a stabbing quality that worsens with inspiration. Pain referred from the pleura may be felt in the thoracic wall in the areas of skin innervated by the intercostal nerves [51].

Pneumothorax — A spontaneous pneumothorax (as well as an acute pulmonary embolus) should be considered in any patient who complains of the sudden onset of pleuritic chest pain and respiratory distress (picture 2). A primary spontaneous pneumothorax usually occurs without a precipitating event in a person (most commonly young, tall, adult male smoker) with no clinical lung disease; recurrence is common [52]. A secondary spontaneous pneumothorax occurs as a complication of underlying lung disease such as chronic obstructive pulmonary disease or pneumocystis pneumonia. (See "Primary spontaneous pneumothorax in adults" and "Secondary spontaneous pneumothorax in adults".)

A tension pneumothorax is rare, but potentially life-threatening unless treated emergently. It occurs when a tissue flap from the injured lung creates a one-way valve, progressively trapping air in the intrapleural space during inspiration. Respiratory failure occurs as the healthy lung is compressed. Physical findings include a unilateral loss of breath sounds with hypertympany, shift of the trachea away from the injured side, and jugular venous distension. The diagnosis is based upon a characteristic history and examination; it should be emergently treated prior to a confirmatory chest radiograph. Decompression is accomplished by inserting a large-bore needle into the second intercostal space in the midclavicular line on the affected side [53].

Pleuritis/serositis — Pleuritis is an inflammation of the parietal and serous pleura of the lung. Viral pleurisy is a common cause of pleuritic chest pain in young adults (table 8). Other causes include autoimmune diseases such as systemic lupus erythematosus or rheumatoid arthritis, and drugs that can cause a lupus-like syndrome including procainamide, hydralazine, isoniazid, and others. (See "Drug-induced lupus" and "Overview of lung disease associated with rheumatoid arthritis" and "Pulmonary manifestations of systemic lupus erythematosus in adults".)

Pleural effusion — A patient with a significant pleural effusion is more likely to present with dyspnea or vague chest discomfort than with typical "pleuritic" chest pain, except in the setting of pleuritis. (See "Diagnostic evaluation of a pleural effusion in adults".)

Mediastinal disease — Disease originating in the mediastinum is a rare cause of chest pain in

primary care practice. Such pain may be associated with signs and symptoms of involvement of different mediastinal or surrounding structures [54]. (See "Evaluation of mediastinal masses".)

PSYCHOGENIC/PSYCHOSOMATIC CAUSES OF CHEST PAIN — Chest pain may be a presenting symptom of panic disorder, depression, and hypochondriasis, as well as cardiac, cancer, or other phobias (table 9) [55-57]. (See "Overview of panic disorder".)

A report of patients evaluated in an emergency department for chest pain found that 20 percent had panic disorder as the etiology [58]. Reviews of the literature have estimated that approximately one-third of patients presenting to the emergency department for chest pain have a psychiatric disorder, while approximately one-half of patients with noncardiac chest pain have various psychiatric diagnoses [59,60]. Among patients with chest pain due to CAD, 20 to 30 percent also have a coexisting psychiatric disorder. Hyperventilation, which is associated with panic attacks, can also result in nonanginal chest pain and occasionally electrocardiographic changes, particularly nonspecific ST and T wave abnormalities [61,62].

Given its prevalence and long-term negative impact on function [63], panic disorder and other psychosocial pathology should be actively considered in evaluating the patient who complains of chest pain of uncertain etiology [64]. Vigilance is necessary since patients with psychiatric disorders may develop organic disease [28,60]. In addition, ischemia may occur during a panic attack in a patient with CAD [65]. Thus organic disease must be reasonably excluded before ascribing chest pain to a nonorganic origin.

Another psychological basis for chest pain is the Munchausen syndrome. This entity denotes a psychological disorder in which patients deliberately feign serious medical illness, inventing false symptoms and signs. In one literature review, 58 patients with cardiac Munchausen syndrome were identified [66]. Of these, 54 (95 percent) were male; the mean age was 44 years (range, 23 to 71). The most common presenting symptom was retrosternal chest pain (50 patients); other presenting complaints were syncope, dyspnea, and back pain. Patients typically gave a history of prior cardiac disease and often reported having "white collar" jobs; on investigation, these historical data proved to be untrue. (See "Factitious disorder and Munchausen syndrome".)

Acute myocardial infarction was the most common admitting diagnosis. All subjects had had numerous admissions and extensive cardiac testing, which were negative for cardiac disease but which the patients reported were positive. When confronted, most patients changed their history, became uncooperative, and refused psychiatric examination. The majority left hospitals against medical advice, and none reported for outpatient follow-up.

PAIN REFERRED TO THE CHEST — Referred pain may occur when the same spinal cord segments supplying dermatomal areas of the chest wall also innervate the very sensitive parietal pleura or peritoneum. As an example, irritation of the mediastinal pleura or of the central diaphragm due to gallbladder or liver disease may result in neck and shoulder pain, while more peripheral diaphragmatic irritation may result in inferior chest pain [51]. A herniated thoracic disc may cause "band-like" anterior chest pain [4].

INFORMATION FOR PATIENTS — Educational materials on this topic are available for patients. (See "Patient information: Chest pain".) We encourage you to print or e-mail this topic review, or

to refer patients to our public web site, www.uptodate.com/patients, which includes this and other topics.

SUMMARY — In the primary care setting, the etiology of nonemergent chest pain can be musculoskeletal (36 percent), cardiac (16 percent), gastrointestinal (19 percent), pulmonary (5 percent), or psychiatric (8 percent). (table 7).

- An approach to the diagnostic evaluation of the adult with chest pain is presented separately. (See "Diagnostic approach to chest pain in adults".)

- Chest wall pain can be categorized as musculoskeletal pain or disorders of the skin and sensory nerves. Musculoskeletal pain can be caused by costochondritis, pain related to surgery, costovertebral joint dysfunction, rheumatologic disorders, and non-rheumatologic disorders (see 'Chest wall pain' above.

- Cardiac causes of chest pain may be related to myocardial ischemia resulting from coronary heart disease, aortic dissection, valvular heart disease, inflammation of the myocardium or pericardium, reversible left ventricular dysfunction due to emotional stress, or cardiac syndrome X. Cardiac syndrome X is ischemic type chest pain that occurs in the absence of documented epicardial coronary artery disease. (See 'Cardiac causes of chest pain' above.)

- Gastrointestinal causes of chest pain may include gastroesophageal reflux, esophageal hyperalgesia, abnormal motility, esophageal disease, and medication use. It may be difficult to distinguish between chest pain due to myocardial ischemia and gastrointestinal causes based upon the history alone. Any patient at risk for CAD who presents with anginal-quality chest pain should have myocardial ischemia ruled out before being given a gastroenterologic diagnosis. (See 'Gastrointestinal causes of chest pain' above.)

- The pulmonary causes of chest pain may be related to the pulmonary vessels, lung parenchyma, pleural tissue, or mediastinal disease (see 'Pulmonary causes of chest pain' above.

- Chest pain may be a presenting symptom of panic disorder, depression, and hypochondriasis, as well as cardiac, cancer, or other phobias. (See 'Psychogenic/psychosomatic causes of chest pain' above.)

- Referred pain may occur when the same spinal cord segments supplying dermatomal areas of the chest wall also innervate other areas that are inflamed, including liver disease or a herniated thoracic disc (see 'Pain referred to the chest' above.

REFERENCES

1. Klinkman, MS, Stevens, D, Gorenflo, DW. Episodes of care for chest pain: A preliminary

report from MIRNET. Michigan Research Network. J Fam Pract 1994; 38:345.

2. Disla, E, Rhim, HR, Reddy, A, et al. Costochondritis. A prospective analysis in an emergency department setting. Arch Intern Med 1994; 154:2466.
3. Wise, CM. Chest wall syndromes (Editorial review). Curr Opin Rheum 1994; 6:197.
4. Brown, CW, Deffer, PA Jr, Akmakjian, J, et al. The natural history of thoracic disc herniation. Spine 1992; 17(6 Suppl):S97.
5. De Maeseneer, M, De Mey, J, Debaere, C, et al. Rib fractures induced by coughing: an unusual cause of acute chest pain. Am J Emerg Med 2000; 18:194.
6. Harbick, S, Frederick, R, Ward, M, Cummings, AJ. Unusual cause of chest pain in a 60-year-old woman. Ann Emerg Med 2000; 35:382.
7. Gilden, DH, Wright, RR, Schneck, SA, et al. Zoster sine herpete, a clinical variant. Ann Neurol 1994; 35:530.
8. Lusiani, L, Perrone, A, Pesavento, R, Conte, G. Prevalence, clinical features, and acute course of atypical myocardial infarction. Angiology 1994; 45:49.
9. Johnson, BD, Shaw, LJ, Buchthal, SD, et al. Prognosis in women with myocardial ischemia in the absence of obstructive coronary disease: results from the National Institutes of Health-National Heart, Lung, and Blood Institute-sponsored Women's Ischemia Syndrome Evaluation (WISE). Circulation 2004; 109:2993.
10. Robinson, JG, Wallace, R, Limacher, M, et al. Elderly women diagnosed with nonspecific chest pain may be at increased cardiovascular risk. J Womens Health (Larchmt) 2006; 15:1151.
11. Bamberg, F, Schlett, CL, Truong, QA, et al. Presence and extent of coronary artery disease by cardiac computed tomography and risk for acute coronary syndrome in cocaine users among patients with chest pain. Am J Cardiol 2009; 103:620.
12. Hollander, JE, Hoffman, RS, Gennis, P, et al. Prospective multicenter evaluation of cocaine-associated chest pain. Cocaine Associated Chest Pain (COCHPA) Study Group. Acad Emerg Med 1994; 1:330.
13. Hollander, JE, Henry, TD. Evaluation and management of the patient who has cocaine-associated chest pain. Cardiol Clin 2006; 24:103.
14. Lodha, A, Mirsakov, N, Malik, B, Shani, J. Spontaneous coronary artery dissection: case report and review of literature. South Med J 2009; 102:315.
15. Hagan, PG, Nienaber, CA, Isselbacher, EM, et al. The International Registry of Acute Aortic Dissection (IRAD): new insights into an old disease. JAMA 2000; 283:897.
16. DeSanctis, RW, Doroghazi, RM, Austen, WG, Buckley, MJ. Aortic dissection. N Engl J Med 1987; 317:1060.
17. Spittell, PC, Spittell, JA Jr, Joyce, JW, et al. Clinical features and differential diagnosis of aortic dissection: Experience with 236 cases (1980 through 1990). Mayo Clin Proc 1993; 68:642.
18. Nallamothu, BK, Saint, S, Kolias, TJ, Eagle, KA. Clinical problem-solving. Of nicks and time. N Engl J Med 2001; 345:359.
19. Stuckey, D. Cardiac pain in association with mitral stenosis and congenital heart disease. Br Heart J 1955; 17:397.
20. Lange, RA, Hillis, LD. Clinical practice. Acute pericarditis. N Engl J Med 2004; 351:2195.
21. Phan, A, Shufelt, C, Merz, CN. Persistent chest pain and no obstructive coronary artery

disease. JAMA 2009; 301:1468.

22. Rosen, SD, Uren, NG, Kaski, JC, et al. Coronary vasodilator reserve, pain perception and sensitivity in patients with syndrome X. Circulation 1994; 90:50.
23. Kaski, JC, Rosano, GM, Collins, P, et al. Cardiac syndrome X: clinical characteristics and left ventricular function. Long-term follow-up study. J Am Coll Cardiol 1995; 25:807.
24. Sohn, JE, Dunn, JP, Burford, H, et al. Think, diagnose, and rethink. Am J Med 2009; 122:823.
25. Thygesen, K, Alpert, JS, White, HD. Universal definition of myocardial infarction. J Am Coll Cardiol 2007; 50:2173.
26. Davies, HA, Jones, DB, Rhodes, J, Newcombe, RG. Angina-like esophageal pain: Differentiation from cardiac pain by history. J Clin Gastroenterol 1985; 7:477.
27. Voskuil, J, Cramer, MJ, Breumelhof, R, et al. Prevalence of esophageal disorders in patients with chest pain newly referred to the cardiologist. Chest 1996; 109:1210.
28. Ros, E, Armengol, X, Grande, L, et al. Chest pain at rest in patients with coronary artery disease. Myocardial ischemia, esophageal dysfunction, or panic disorder? Dig Dis Sci 1997; 42:1344.
29. Fass, R, Fennerty, MB, Ofman, JJ, et al. The clinical and economic value of a short course of omeprazole in patients with noncardiac chest pain. Gastroenterology 1998; 115:42.
30. Achem, SR, Kolts, BE, MacMath, T, et al. Effects of omeprazole versus placebo in treatment of noncardiac chest pain and gastroesophageal reflux. Dig Dis Sci 1997; 42:2138.
31. Castell, D, Katz, P. The acid suppression test for unexplained chest pain (editorial). Gastroenterology 1998; 115:222.
32. Richter, JE, Barish, CF, Castell, DO. Abnormal sensory perception in patients with esophageal chest pain. Gastroenterology 1986; 91:845.
33. Rao, SS, Gregersen, H, Hayek, B, et al. Unexplained chest pain: The hypersensitive, hyperreactive, and poorly compliant esophagus. Ann Intern Med 1996; 124:950.
34. Cannon, RO, Camici, PG, Epstein, SE. Pathophysiological dilemma of syndrome X. Circulation 1992; 85:883.
35. Borjesson, M, Albertson, P, Dellborg, M, et al. Esophageal dysfunction in syndrome X. Am J Cardiol 1998; 82:1187.
36. Pate, JW, Walker, WA, Hammond Cole, F, et al. Spontaneous rupture of the esophagus: A 30-year experience. Ann Thorac Surg 1989; 47:689.
37. Guenther, M, Wunderlich, C, Madisch, A, et al. Beans in the pericardium. Lancet 2009; 374:586.
38. Koyuncu, N, Yilmaz, S, Soysal, S. An unusual cause of chest pain: foreign body in the oesophagus. Emerg Med J 2007; 24:e1.
39. Thomas, LR, Baden, L, Zaleznik, DF. Clinical problem-solving. Chest pain with a surprising course. N Engl J Med 1999; 341:1134.
40. Majeski, J, Durst, GG. Geriatric acute perforated appendix: atypical symptoms lead to a difficult diagnosis. South Med J 1998; 91:669.
41. An exploratory report of chest pain in primary care. A report from ASPN. J Am Board Fam Pract 1990; 3:143.
42. Svavarsdottir, AE, Jonasson, MR, Gudmundsson, GH, Fjeldsted, K. Chest pain in family practice. Diagnosis and long-term outcome in a community setting. Can Fam Physician

1996; 42:1122.

43. Palla, A, Petruzzelli, S, Donnamaria, V, Giuntini, C. The role of suspicion in the diagnosis of pulmonary embolism. Chest 1995; 107(Suppl):21S.

44. Yu, DR, Miller, R, Bray, PF. Clinical problem-solving. Through thick and thin [see comments]. N Engl J Med 1998; 338:1684.

45. Stein, PD, Terrin, ML, Hales, CA, Palevsky, HI. Clinical, laboratory, roentgenographic, and electrocardiographic findings in patients with acute pulmonary embolism and no pre-existing cardiac or pulmonary disease. Chest 1991; 100:598.

46. PIOPED II.

47. Stein, PD, Beemath, A, Matta, F, et al. Clinical characteristics of patients with acute pulmonary embolism: data from PIOPED II. Am J Med 2007; 120:871.

48. Stein, PD, Fowler, SE, Goodman, LR, et al. Multidetector computed tomography for acute pulmonary embolism. N Engl J Med 2006; 354:2317.

49. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). The PIOPED Investigators. JAMA 1990; 263:2753.

50. Marrie, TJ. Community-acquired pneumonia. Clin Infect Dis 1994; 18:501.

51. Moore, KL. Clinically Oriented Anatomy, 2nd ed, Williams and Wilkins, Baltimore 1985. p.146.

52. Sadikot, RT, Greene, T, Meadows, K, Arnold, AG. Recurrence of primary spontaneous pneumothorax. Thorax 1997; 52:805.

53. Advanced Trauma Life Support Course for Physicians. Committee on Trauma, American College of Surgeons, Chicago 1993. p.143.

54. Gaissert, HA, Piyavisetpat, N, Mark, EJ. Case records of the Massachusetts General Hospital. Case 14-2009. A 36-year-old man with chest pain, dysphagia, and pleural and mediastinal calcifications. N Engl J Med 2009; 360:1886.

55. Alpha-1-antitrypsin deficiency. Usage of alpha-1-proteinase inhibitor concentrate in replacement therapy. A symposium. November 1, 1986, San Diego, California. Proceedings. Am J Med 1988; 84:1.

56. Eifert, GH, Hodson, SE, Tracey, DR, et al. Heart-focused anxiety, illness beliefs, and behavioral impairment: Comparing healthy heart-anxious patients with cardiac and surgical inpatients. J Behav Med 1996; 19:385.

57. Katon, W. Panic disorder and somatization: a review of 55 cases. Am J Med 1984; 77:101.

58. Worthington, JJ III, Pollack, MH, Otto, MW, et al. Panic disorder in emergency ward patients with chest pain. J Nerv Ment Dis 1997; 185:274.

59. Wulsin, LR, Yingling, K. Psychiatric aspects of chest pain in the emergency department. Med Clin North Am 1991; 75:1175.

60. Fleet, RP, Dupuis, G, Marchand, A, et al. Panic disorder, chest pain and coronary artery disease: Literature review. Can J Cardiol 1994; 10:827.

61. Bass, C, Chambers, JB, Kiff, P, et al. Panic anxiety and hyperventilation in patients with chest pain: a controlled study. Q J Med 1988; 69:949.

62. Evans, DW, Lum, LC. Hyperventilation: An important cause of pseudoangina. Lancet 1977; 1:155.

63. Lantinga, LJ, Sprafkin, RP, McCroskery, JH, et al. One-year psychosocial follow-up of

patients with chest pain and angiographically normal coronary arteries. Am J Cardiol 1988; 62:209.

64. Lynch, P, Galbraith, KM. Panic in the emergency room. Can J Psychiatry 2003; 48:361.
65. Ben Freedman, S, Tennant, CC. Panic disorder and coronary artery spasm. Med J Aust 1998; 168:376.
66. Mehta, NJ, Khan, IA. Cardiac Munchausen syndrome. Chest 2002; 122:1649.

Alternative diagnoses to cardiac ischemia for patients with chest pain

Non-ischemic cardiovascular	Pulmonary	Gastrointestinal
Aortic dissection*	Pleuritis	Biliary
Myocarditis	Pneumonia	Cholangitis
Pericarditis	Pulmonary embolus*	Cholecystitis
Chest wall	Tension pneumothorax*	Choledocholithiasis
Cervical disc disease	**Psychiatric**	Colic
Costochondritis	Affective disorders (eg, depression)	Esophageal
Fibrositis	Anxiety disorders	Esophagitis
Herpes zoster (before the rash)	Hyperventilation	Spasm
Neuropathic pain	Panic disorder	Reflux
Rib fracture	Primary anxiety	Rupture*
Sternoclavicular arthritis	Somatiform disorders	Pancreatitis
	Thought disorders (eg, fixed delusions)	Peptic ulcer disease
		Nonperforating
		Perforating*

* Potentially life-threatening conditions.

Adapted with permission from: ACC/AHA/ACP Guidelines for the Management of Patients with Chronic Stable Angina. J Am Coll Cardiol 1999; 33:2092. Copyright ©1999 American College of Cardiology.

Causes of nonemergent chest pain in MIRNET primary care practices

Cause	Prevalence, percent
Musculoskeletal, including costochondritis	36
Gastrointestinal	19
Cardiac	16*
Stable angina	10.5
Unstable angina or MI	1.5
Other cardiac	3.8
Psychiatric	8
Pulmonary	5
Other/unknown	16

MIRNET: Michigan Research Network.
* As high as 50 percent in older populations.
Adapted from Klinkman, MS, Stevens, D, Gorenflo, DW, J Fam Pract 1994; 38:345.

Important features of the history in musculoskeletal chest pain

Features suggestive of visceral causes
Middle aged or elderly patient
Risk factors for coronary disease
Exertional pain
Cough
Fever
Dyspnea
Atypical location
Features suggestive of musculoskeletal cause
Insidious onset
Recent repetitive unaccustomed activity (trunk and arms)
Pain may be localized or diffuse
Positional component
Persistent and prolonged (lasting hours-days)
Features suggestive of associated condition
Neck, thoracic, or shoulder pain (pain referred to chest)
Chronic low back pain, young patient (ankylosing spondylitis)
Ocular inflammation (ankylosing spondylitis or related disease)
Diffuse musculoskeletal pain/sleep disturbance (fibromyalgia)
Peripheral joint pain and swelling (rheumatoid arthritis)
Skin lesions (acne or psoriasis), (psoriatic arthritis, SCCH)

Prevalence of musculoskeletal pain

Population	Prevalence, percent
Non-emergent ambulatory care chest pain	36
All adult emergency room chest pain	10-15
Non-cardiac emergency room chest pain	26
Pediatric emergency room chest pain	20-25
Chest pain with negative coronary angiography (nonspecific tenderness in another 50-60 percent)	13-20

Prevalence of musculoskeletal chest pain in patients with chest pain in different clinical settings. See text for details.

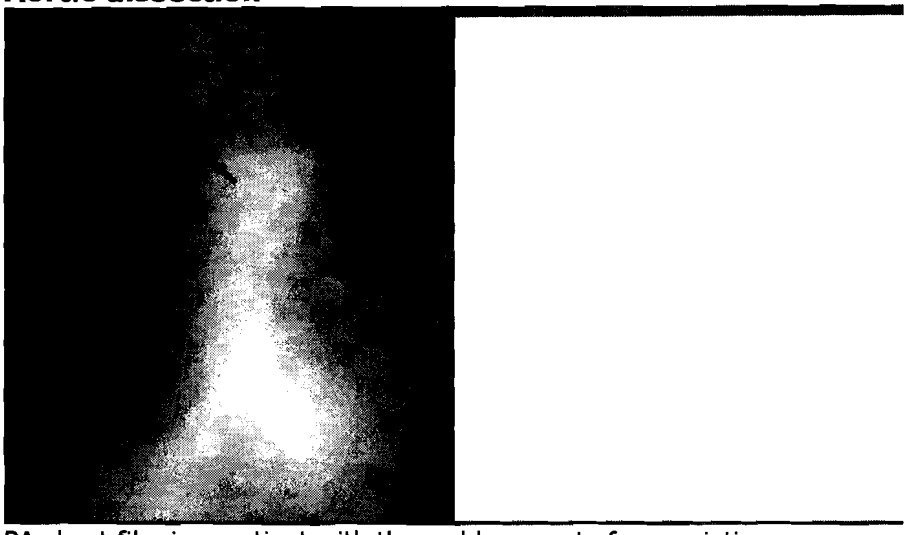
Characteristics of isolated musculoskeletal chest pain syndromes

Disorder	Clinical manifestations
Costosternal syndromes (costochondritis)	Multiple areas of tenderness that reproduce the described pain, usually in the upper costal cartilages at the costochondral or costosternal junctions; there is no swelling.
Tietze's syndrome	Painful, nonsuppurative localized swelling of the costosternal, sternoclavicular, or costochondral joints, most often involving one joint in the area of the second and third ribs; rare, primarily affects young adults.
Sternalis syndrome	Localized tenderness over the body of the sternum or overlying sternalis muscle; palpation often causes radiation of pain bilaterally.
Xiphoidalgia	Localized discomfort over the sternum at the xiphoid process.
Spontaneous sternoclavicular subluxation	Most often occurs in the dominant side, associated with moderate to heavy repetitive tasks; almost exclusively occurs in middle-aged women.
Lower rib pain syndromes	Pain in the lower chest or upper abdomen with a tender spot on the costal margin; pain can be reproduced by pressing on the spot.
Posterior chest wall syndromes	May be caused by herniated thoracic disc, leading to band-like chest pain that may have a unilateral dermatomal distribution. Also induced by costovertebral joint dysfunction; tenderness over the affected area, worse with coughing or deep breathing.

Rheumatic and systemic diseases associated with musculoskeletal chest wall pain

Systemic lupus erythematosus
Rheumatoid arthritis
Ankylosing spondylitis
Psoriatic arthritis
Sternocostoclavicular hyperostosis (SAPHO syndrome)
Fibromyalgia (fibrositis)
Infectious arthritis
Relapsing polychondritis
Other systemic conditions
Osteoporosis, osteomalacia
Tumors (benign, malignant, metastatic and primary)
Sickle cell disease

Aortic dissection



PA chest film in a patient with the sudden onset of excruciating interscapular pain and hypotension. The ascending aortic arch is dilated, displacing the trachea to the right (black arrow). A left lower lobe density is suggestive of a pleural effusion. Surgery revealed a dilated ascending aorta with a dissection approximately 3 cm distal to the aortic valve. *Courtesy of Robert A Novelline, MD.*

Features found more frequently in patients with esophageal chest pain

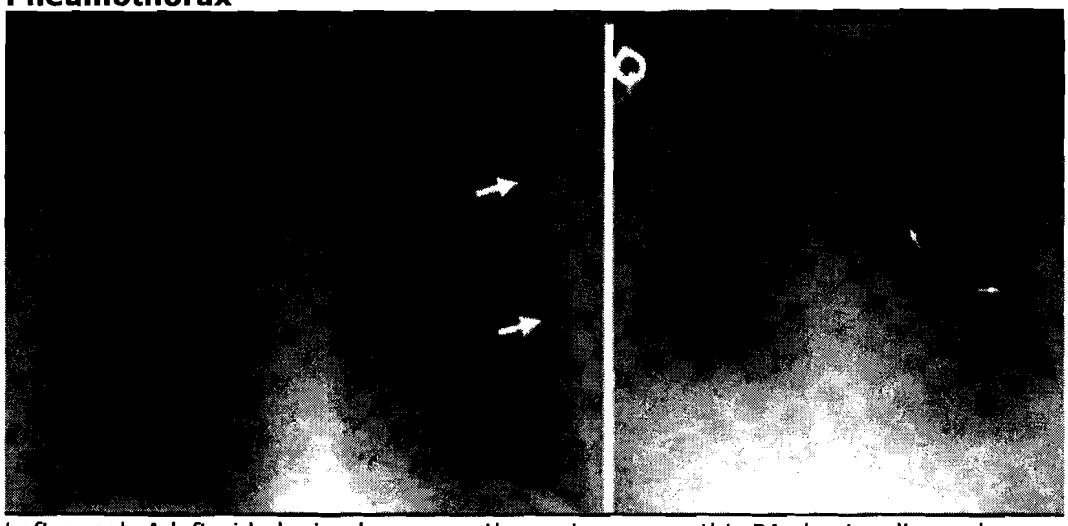
Pain provoked by swallowing
Pain provoked by postural changes
Pain palliated by antacids
An inconsistent relationship to exercise
Substernal chest pain that does not radiate
Frequent episodes of spontaneous pain
Nocturnal pain
Severe onset of pain, continuing as a background ache for several hours
Pain associated with heartburn and regurgitation of acid into the mouth

Data from Davies, HA, Jones, DB, Rhodes, J, Newcombe, RG, J Clin Gastroenterology 1985; 7:477.

Causes of venous thrombosis

Inherited thrombophilia
Factor V Leiden mutation
Prothrombin gene mutation
Protein S deficiency
Protein C deficiency
Antithrombin (AT) deficiency
Rare disorders
Dysfibrinogenemia
Acquired disorders
Malignancy
Presence of a central venous catheter
Surgery, especially orthopedic
Trauma
Pregnancy
Oral contraceptives
Hormone replacement therapy
Tamoxifen, Bevacizumab, Thalidomide, Lenalidomide
Immobilization
Congestive failure
Antiphospholipid antibody syndrome
Myeloproliferative disorders
Polycythemia vera
Essential thrombocythemia
Paroxysmal nocturnal hemoglobinuria
Inflammatory bowel disease
Nephrotic syndrome
Hyperviscosity
Waldenstrom's macroglobulinemia
Multiple myeloma
Marked leukocytosis in acute leukemia
Sickle cell anemia
HIV/AIDS

Pneumothorax



Left panel: A left-sided, simple pneumothorax is seen on this PA chest radiograph (large white arrows). Right panel: On the expiratory film, the pneumothorax is larger and more easily seen (small white arrows).
Courtesy of Robert A Novelline, MD.

Causes of pleuritic chest pain

Viral pleurisy
Pneumonia
Acute pulmonary embolus
Pneumothorax
Pericarditis
Collagen vascular diseases, including systemic lupus erythematosus, mixed connective tissue disease, and rheumatoid arthritis
Drug-induced lupus
Inflammatory bowel disease
Familial Mediterranean fever
Radiation pneumonitis
Pulmonary histoplasmosis, infection with the lung fluke Paragonimus

Diagnostic criteria for panic attack and panic disorder

Panic attack (summary of DSM-IV criteria)
A discrete period of intense fear or discomfort, in which four (or more) of the following symptoms develop abruptly and reach a peak within ten minutes:
Cardiopulmonary symptoms
Chest pain or discomfort
Sensations of shortness of breath or smothering
Palpitations, pounding heart, or accelerated heart rate
Neurological symptoms
Trembling or shaking
Parasthesias (numbness or tingling sensation)
Feeling dizzy, unsteady, light-headed or faint
Psychiatric symptoms
Derealization (feelings of unreality) or depersonalization (being detached from oneself)
Fear of losing control or going crazy
Fear of dying
Autonomic symptoms
Sweating
Chills or hot flushes
Gastrointestinal symptoms
Feeling of choking
Nausea or abdominal distress
Panic disorder (summary of DSM-IV criteria)
With agoraphobia
A. Recurrent, unexpected panic attacks.
B. At least one of the attacks has been followed by a month or more of: persistent concern about having additional attacks; worry about the implications of the attack or its consequences; a significant change in behavior related to the attacks.
C. The presence of agoraphobia, ie, anxiety about being in places or situations in which escape might be difficult (or embarrassing) or in which help might not be available in the event of having a panic attack.
Without agoraphobia
A. Both A and B above
B. Absence of agoraphobia

Adapted from Diagnostic and Statistical Manual of Mental Disorders, 4th Ed, Primary Care Version (DSM-IV-PC). American Psychiatric Association, Washington, DC 1995.

Annals of Internal Medicine

Established in 1927 by the American College of Physicians

Article: **The Net Clinical Benefit of Warfarin Anticoagulation in Atrial Fibrillation**

Daniel E. Singer,
Yuchiao Chang,
Margaret C. Fang,
Leila H. Borowsky,
Niela K. Pomernacki,
Natalia Udaltsova,
and Alan S. Go

Ann Intern Med September 1, 2009 151:297-305;

CHADS2 (85x2)

- **James L. Meisel, M.D., FACP**

Boston University School of Medicine

Singer et al have further validated the benefit of anticoagulation with adjusted-dose warfarin for populations at high risk for thromboembolic stroke due to nonvalular atrial fibrillation (1,2). Increased net clinical benefit may accrue to older patients, those with a history of ischemic stroke or TIA and those with CHADS2 scores of 2 or more (3).

The mnemonic "CHADS2 (85x2)" may help clinicians recall the thrust of the latest research (3,4,5). In this adaptation, age>=85, like a history of ischemic stroke or TIA, would be given a double point value. Thus age>=85 alone would raise a patient's CHADS2 score to 2, i.e., into the high risk zone. Age 75-84, like a history of congestive heart failure, hypertension and diabetes mellitus, would remain less potent but validated risk factors.

Weighing the probable benefit of stroke prevention against bleeding risk, especially in the elderly, remains an important part of

individualized decision-making. This is especially true since "CHADS2 (85x2)", while at the moment a logical tool to help physicians decide who should receive chronic anticoagulation, has not been independently validated and is partially dependent upon a large, well done but nonrandomized observational assessment (3).

References

1. Singer DE, Albers GW, Dalen JE, Fang MC, Go AS, Halperin JL, et al; American College of Chest Physicians. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest. 2008;133:546S-592S.

2. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA. 2001;285:2864-70.

3. Singer DE, Chang Y, Fang MC, Borowsky LH, Pomernacki NK, Udaltsova N, et al. The net clinical benefit of warfarin anticoagulation in atrial fibrillation. Ann Intern Med. 2009;151:297-305.

4. van Walraven C, Hart RG, Connolly S, Austin PC, Mant J, Hobbs FD, et al. Effect of age on stroke prevention therapy in patients with atrial fibrillation: the atrial fibrillation investigators. Stroke. 2009 Apr;40:1410-6.

5. Mant J, Hobbs FD, Fletcher K, Roalfe A, Fitzmaurice D, Lip GY, et al; BAFTA investigators. 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.

Conflict of Interest:

None declared

Atherosclerosis 203 (2009) 509–514

Contents lists available at ScienceDirect

Atherosclerosis

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

Cross-sectional relations of multiple inflammatory biomarkers to peripheral arterial disease: The Framingham Offspring Study☆

Joanne M. Murabito[a,b,*], Michelle J. Keyes[a,f], Chao-Yu Guo[g], John F. Keaney Jr.[h], Ramachandran S. Vasan[a,c,d,e,i], Ralph B. D'Agostino Sr.[a,f], Emelia J. Benjamin[a,c,d,e,i]

[a] National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA, United States
[b] Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, United States
[c] Department of Cardiology, Boston University School of Medicine, Boston, MA, United States
[d] Department of Preventive Medicine, Boston University School of Medicine, Boston, MA, United States
[e] Whitaker Cardiovascular Institute, Boston University School of Medicine, Boston, MA, United States
[f] Statistics and Consulting Unit, Department of Mathematics and Statistics, Boston University, Boston, MA, United States
[g] Children's Hospital and Harvard Medical School, Boston, MA, United States
[h] University of Massachusetts Medical School, Worcester, MA, United States
[i] Epidemiology Section, Boston University School of Public Health, Boston University, Boston, MA, United States

ARTICLE INFO

Article history:
Received 7 March 2008
Received in revised form 19 June 2008
Accepted 27 June 2008
Available online 9 July 2008

Keywords:
Peripheral arterial disease
Ankle-brachial index
Interleukin-6
Tumor necrosis factor receptor 2

ABSTRACT

Background: Emerging evidence suggests that different inflammatory biomarkers operate through distinct biologic mechanisms. We hypothesized that the relation to peripheral arterial disease (PAD) varies for individual markers.

Methods: In a community-based sample we measured 12 biomarkers including *plasma* CD40 ligand, fibrinogen, lipoprotein-associated phospholipase-A2 mass and activity, osteoprotegerin, P-selectin, and tumor necrosis factor receptor 2 (TNFR2); and *serum* C-reactive protein, intracellular adhesion molecule-1, interleukin-6, monocyte chemoattractant protein-1, and myeloperoxidase in Framingham Offspring Study participants ($n = 2800$, 53% women, mean age 61 years). We examined the cross-sectional relation of the biomarker panel to PAD using (1) a global test of significance to determine whether at least one of 12 biomarkers was related to PAD using the TEST statement in the LOGISTIC procedure in SAS and (2) stepwise multivariable logistic regression with forward selection of markers with separate models for (1) ankle-brachial index (ABI) category (<0.9, 0.9–1.0, >1.0) and (2) presence of clinical PAD (intermittent claudication or lower extremity revascularization).

Results: The group of inflammatory biomarkers were significantly related to both ABI and clinical PAD ($p = 0.01$ and $p = 0.02$, respectively, multi-marker adjusted global significance test). Multivariable forward elimination regression retained interleukin-6 and TNFR2 as significantly associated with PAD. For one standard deviation change in interleukin-6 and TNFR2 concentrations, there was a 1.21 ($p = 0.005$) and 1.19 ($p = 0.009$) increased odds of a change in ABI level respectively. Similar results were observed for clinical PAD.

Conclusion: Interleukin-6 and TNFR2 were significantly associated with PAD independent of established risk factors and each other, suggesting that each marker represents a distinct biologic pathway.

☆ The Framingham Heart Study is supported by NIH/NHLBI contract N01-HC-25195. Additional funding for this work supported by NIH grants RO1 HL064753, HL076784, AG028321 (E.J.B.); HL 080124, HL 077477, HL71039, K24 HL 4334 (R.S.V.); tumor necrosis factor-alpha concentrations measured via American Diabetes Association Career Development Award and NCRR GCRC M01-RR-01066 James B. Meigs, MD, MPH. Lipoprotein-associated phospholipase-A2 Lp-PLA2 activity measurements were provided by GlaxoSmithKline at no cost to the FHS.

* Corresponding author at: The Framingham Heart Study, 73 Mount Wayte Avenue, Suite 2, Framingham, MA 01702-5827, United States. Tel.: +1 508 935 3461; fax: +1 508 626 1262.
E-mail address: murabito@bu.edu (J.M. Murabito).

0021-9150/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.atherosclerosis.2008.06.031

1. Introduction

Peripheral arterial disease (PAD) affects approximately five to eight million Americans [1] and is a powerful predictor of incident coronary heart disease, stroke, and mortality [2–4]. The ankle-brachial index (ABI), a subclinical measure of PAD, is considered a marker of generalized atherosclerosis. It is now well established that inflammation plays a central role in the pathogenesis of atherosclerosis and, further, that various inflammatory markers predict incident cardiovascular disease (CVD) events [5,6]. However, risk for CVD associated with inflammatory markers is often attenuated with adjustment for traditional risk factors.

The relation between inflammatory markers and PAD is not fully characterized. C-reactive protein predicts risk of symptomatic PAD [7,8], and also is associated with atherosclerosis in the aorta [9] and femoral artery [10], but the associations are attenuated after accounting for established risk factors. Adjusted inverse relations between the ABI and C-reactive protein have been reported in men [11], in ever smokers [12], and in persons with prevalent CVD [13]. Higher C-reactive protein concentrations are associated with progression of aortic, iliac, and lower extremity atherosclerosis [14], and in one small study both a low ABI and a high C-reactive protein identified persons at greatest risk for clinical events and death [15]. Reports of the relations between other inflammatory markers and PAD are limited, often focus on a single marker, or on small hospital-based or referral-based samples, and demonstrate conflicting results [10,16–18]. Emerging evidence suggests that different inflammatory markers operate through distinct biologic mechanisms, and thus the relative importance to the atherosclerotic process and PAD may differ for individual markers.

We examined the cross-sectional relations of a panel of 12 inflammatory biomarkers and PAD in a large community-based sample. We selected the inflammatory and oxidative stress markers to represent various stages and pathways in the inflammatory process, including chemokines (monocyte chemoattractant-1), cytokines (interleukin-6, tumor necrosis factor-alpha and tumor necrosis receptor-2; selectins [P-selectin and CF40 ligand] cell adhesion molecules [intercellular adhesion molecule-1]) acute phase reactants (C-reactive protein, fibrinogen), and an oxidative stress marker (myeloperoxidase). We hypothesized that different inflammatory markers represent distinct biologic pathways, and thus not all markers would be related to PAD. Moreover, we postulated that the strength of the relation would vary for different biomarkers operating through unique pathways. To our knowledge, no other study has evaluated the relation between multiple biomarkers from potentially diverse biologic pathways and PAD conjointly.

2. Methods

2.1. Study sample

The Framingham Offspring Study was initiated in 1971 when 5124 adult children (and offspring spouses) of the Original cohort were enrolled in the Framingham Heart Study. Offspring participants have been examined approximately every 4–8 years since the study's inception. Written informed consent was obtained at each examination and the Institutional Review Board of Boston University Medical Center approved the examination content.

Offspring participants who attended the seventh examination cycle (1998–2001) were eligible for this study. The examination included a standardized medical history and physical examination, electrocardiogram, noninvasive cardiovascular testing, and measurement of fasting lipids, glucose, and a panel of inflammatory biomarkers. Of the 3539 participants attending the examination, 205 were examined off-site and did not have ABI testing, 92 participants had incomplete ABI data, and 12 participants were excluded because of an ABI > 1.4. We further excluded participants missing all biomarker data ($n = 289$) and participants with incomplete risk factor data ($n = 141$). Thus, our study sample included 2800 Offspring participants with data available for all 12 biomarkers and complete risk factor data.

2.2. Measurement of ankle-brachial index

Ankle-brachial systolic blood pressure measurements were obtained using a standard protocol by trained technicians and the details previously published [19]. An 8 MHz Doppler pen probe and an ultrasonic Doppler flow detector (Parks Medical Electronics, Inc.) were used to measure the systolic blood pressure in each limb. All limb blood pressures were repeated, and if the initial and repeat blood pressures differed by more than 10 mmHg at any one site, a third measurement was obtained. Measurements were obtained from the dorsalis pedis artery only if the posterior tibial pulse could not be located by palpation or with Doppler probe. For this study, the ABI was defined as the ratio of the average systolic blood pressure in the ankle divided by the average systolic blood pressure in the higher arm. The lower ABI was used for analysis. Based on prior epidemiologic studies, we analyzed ABI < 0.9 as indicative of PAD.

2.3. Intermittent claudication and lower extremity revascularization

Intermittent claudication was assessed using a standardized physician-administered questionnaire that inquired about the presence of exertional calf discomfort related to walking uphill or walking rapidly and was relieved with rest. Two physicians independently interviewed all participants suspected to have intermittent claudication. An endpoint panel, comprised of three senior investigators, examined all medical evidence and made the final diagnosis of the presence of intermittent claudication. Participants were also queried about revascularization procedures including lower extremity bypass surgery and percutaneous transluminal angioplasty. The endpoint panel reviewed hospital records for all cardiovascular procedures.

2.4. Inflammatory biomarker measurement

At examination cycle seven, 12 biomarkers were measured including *plasma* CD40 ligand, fibrinogen, lipoprotein-associated phospholipase A2 mass and activity, osteoprotegerin, P-selectin, and tumor necrosis factor receptor 2 (TNFR2); and *serum* C-reactive protein, intracellular adhesion molecule-1, interleukin-6, monocyte chemoattractant protein-1, and myeloperoxidase. Specimens were collected from fasting participants and plasma and serum aliquots were stored at −80 °C until analysis. Biomarkers, except C-reactive protein, were measured in duplicate with commercially available ELISA kits from R&D Systems (intracellular adhesion molecule-1, interleukin-6, monocyte chemoattractant protein-1, P-selectin, TNFR2), Bender MedSystems (CD40 ligand), Diagnostica (fibrinogen), Oxis (myeloperoxidase) and ALPCO (osteoprotegerin). The Dade Behring BN100 nephelometer was used to measure high sensitivity C-reactive protein. Lipoprotein-associated phospholipase A2 activity was measured by GlaxoSmithKline, and mass was measured by DiaDexus. Details for assays have been previously published [20]. The intra-assay coefficients of variation for the biomarkers were as follows: CD40 ligand 4.4 ± 3.4%, fibrinogen 1.1 ± 1.1%, intracellular adhesion molecule-1 3.7 ± 2.4%, interleukin-6 3.1 ± 2.1%, lipoprotein-associated phospholipase A2 activity 7.0% (low) and 5.9% (high) and mass (based on 24% dupli-

cate readings) 6% (low) and 8% (high) concentrations, monocyte chemoattractant protein-1 3.8 ± 3.3%, myeloperoxidase 3.0 ± 2.5%, osteoprotegerin 3.7 ± 2.9%, P-selectin 3.0 ± 2.2%, TNFR2 2.2 ± 1.6%. The kappa statistic based on 146 C-reactive protein samples was 0.95. Additionally, plasma tumor necrosis factor alpha (R&D Systems, CV 7.6% low, 5.6% high control) and urinary isoprostanes, 8-Epi-$PGF_{2\alpha}$ (Cayman, Ann Arbor, MI; CV 9.1 ± 5.8%), indexed to urinary creatinine were measured on a subset of participants.

2.5. Clinical covariate assessment

Covariates were defined at the time of examination cycle seven. Medication use and current smoking within the year preceding the exam were self-reported. Resting blood pressure was measured twice by the examining physician. Hypertension was defined as an average blood pressure of systolic ≥140 or diastolic ≥90 mmHg or use of anti-hypertensive medication. Body mass index was calculated as weight in kilograms divided by the height in meters squared. Diabetes was defined by fasting blood glucose of ≥126 mg/dL, or use of insulin or oral hypoglycemic agents. CVD was defined as coronary heart disease, stroke or transient ischemic attack, and heart failure. An endpoint adjudication panel made the final diagnostic determination using previously reported criteria [21].

2.6. Statistical analysis

Sex-specific standardization of biomarkers was performed (i.e., within each sex, biomarkers were standardized to have a mean of 0 and a standard deviation of 1). Due to skewed distributions, biomarker concentrations were natural logarithmically transformed for analysis. Our primary analysis was the simultaneous consideration of multiple biomarkers (independent variables) in relation to PAD defined as two separate variables: (1) ABI category (ABI: <0.9, 0.9–1.0, >1.0) and (2) presence of clinically overt PAD defined as intermittent claudication or lower extremity revascularization. Separate logistic regression models were run for ABI category and presence of clinically overt PAD. First, we performed a global test of significance to determine whether at least one of 12 biomarkers was related to the PAD dependent variables using the TEST statement in the LOGISTIC procedure in SAS. The analysis was adjusted for age, sex, and the following 13 clinical covariates previously reported to be correlated with biomarkers and or PAD [19,22,23]: current cigarette smoking, number of pack-years of cigarette smoking, diabetes, fasting glucose, body mass index, waist circumference, total to HDL cholesterol ratio, fasting triglyceride, lipid lowering treatment, hypertension, aspirin use, prevalent CVD (myocardial infarction, coronary insufficiency, angina pectoris, stroke, or transient ischemic attack), and use of hormone replacement therapy. Second, we conducted a stepwise multivariable ordinal logistic regression with PAD as the dependent variable, with forward selection of biomarkers using a $p < 0.05$ adjusting for age, sex, and forcing the 13 clinical covariates into the model. For biomarkers identified to be related to PAD in the second step of the analysis we calculated point estimates of the odds ratio (or i.e., the relative change in odds of PAD), with 95% confidence intervals, per standard deviation increase of the biomarker examined.

We conducted several secondary analyses. We examined effect modification by age (<60, ≥60 years) and sex for significant biomarker—PAD relations. We repeated the analysis in persons free of CVD. Finally because multiple reports have used different markers or sets of markers, we analyzed the multivariable-adjusted linear relations of each log-transformed marker (dependent variable), one marker at a time, to the independent PAD measures using PROC GLM in SAS. Tumor necrosis factor-alpha was measured on a subset of participants attending examination cycle seven ($n = 2129$) and was included in the secondary analysis. SAS version 8.1 was used to perform all analyses [24].

3. Results

3.1. Participant characteristics and biomarker concentrations

Clinical characteristics of the study sample by presence of clinically overt PAD, and by ABI category are shown in Table 1. Participants with PAD, defined by symptoms or an ABI < 0.9, were older than participants without PAD. The untransformed median for the 12 biomarkers and tumor necrosis factor-alpha (available on a subset) by ABI level are shown in Table 2. A graded increase in marker concentrations across decreasing ABI levels was present for all markers except lipoprotein-associated phospholipase A2 mass and activity whereas an inverse relation was seen for CD40 ligand.

3.2. Global relations of multiple biomarkers and measures of PAD

The inflammatory markers as a group were significantly related to both ABI and clinically detected PAD ($p = 0.01$ and $p = 0.02$ respectively for the multi-marker adjusted global test of significance) as shown in Table 3. The forward elimination regression retained interleukin-6 and TNFR2 in the final models as significantly associated with ABI and with intermittent claudication or lower extremity revascularization. The odds of a one category reduction in ABI

Table 1
Clinical characteristics of the study sample

Variable mean (S.D.) or %	Intermittent claudication		Ankle-brachial index		
	Yes or vascular intervention, $N = 90$	No, $N = 2710$	<0.9, $N = 111$	0.9–1.0, $N = 225$	>1.0–1.4, $N = 2464$
Age, years	67 (9)	61 (9)	70 (8)	65 (10)	60 (9)
Women, %	43	54	50	71	52
Current smoking, %	27	13	30	26	12
Pack years among ever smokers, mean	63 (25)	43 (23)	65 (25)	47 (23)	41 (22)
Diabetes, %	29	12	26	20	11
Body mass index, kg/m^2	29.7 (6.1)	28.0 (5.1)	28.2 (5.5)	28.5 (6.6)	28.0 (4.9)
Total/HDL cholesterol ratio	4.5 (1.4)	4.0 (1.3)	4.4 (1.5)	4.0 (1.5)	4.1 (1.3)
Lipid lowering treatment, %	49	19	40	28	19
Hypertension, %	74	44	76	62	42
Hypertension treatment, %	68	32	66	49	31
Aspirin use[a], %	57	30	49	34	30
Prevalent cardiovascular disease[b], %	52	10	33	15	9
Hormone replacement among women, %	38	30	16	31	31

[a] Aspirin use is defined as three or more tablets per week.
[b] Cardiovascular disease did not include intermittent claudication.

Table 2
Unadjusted inflammatory marker data by ankle-brachial index level

Marker, units	Ankle-brachial index		
	<0.9, *N* = 111	0.9–1.0, *N* = 225	>1.0–1.4, *N* = 2464
	Untransformed marker concentrations, median (lower, upper quartile)		
CD40 ligand, ng/mL	0.78 (0.46, 2.31)	0.96 (0.53, 3.04)	1.27 (0.56, 4.07)
C-reactive protein, mg/L	3.77 (1.99, 8.89)	3.71 (1.53, 7.08)	2.02 (0.94, 4.76)
Fibrinogen, mg/dL	425 (384, 481)	388 (349, 446)	368 (326, 418)
Intercellular adhesion molecule-1, ng/mL	283 (243, 323)	255 (223, 292)	239 (209, 279)
Interleukin-6, pg/mL	4.82 (2.85, 7.84)	3.46 (2.21, 5.66)	2.58 (1.75, 4.09)
LpPLA2, mass, nmol/(mL min)	284 (231, 374)	293 (231, 367)	288 (229, 360)
LpPLA2, activity, ng/mL	143 (120, 173)	134 (116, 162)	141 (119, 165)
Monocyte chemoattractant protein-1, pg/mL	346 (283, 412)	326 (269, 409)	310 (252, 378)
Myeloperoxidase, ng/mL	43.4 (29.5, 61.7)	40.3 (26.9, 58.2)	39.9 (27.8, 59.6)
Osteoprotegerin, pmol/L	6.71 (5.35, 8.36)	5.85 (4.88, 7.10)	5.30 (4.39, 6.34)
P-selectin, pg/mL	39.7 (31.4, 54.0)	38.5 (31.4, 48.7)	36.0 (28.2, 45.1)
Tumor necrosis factor receptor 2, pg/mL	2407 (2024, 3187)	2154 (1718, 2856)	1945 (1642, 2340)
Tumor necrosis factor alpha[a], pg/mL	1.51 (1.15, 1.85)	1.37 (0.99, 1.92)	1.18 (0.92, 1.58)
Urine[a] 8-epi-$PGF_{2\alpha}$, ng/mmol	162 (103, 247)	151 (100, 234)	131 (88, 192)

[a] TNF-α data is available on a subset of 2129 participants, urine 8-epi-PGF_2 available on 2404 participants LpPLA2 = lipoprotein-associated phospholipase A2.

Table 3
Joint consideration of biomarkers in relation to the ankle-brachial index and clinical peripheral arterial disease

	Global P[a]	Stepwise selection biomarker[b]	Odds ratio (95% confidence interval)[c]	*p*-Value
ABI[d]	0.01	Interleukin-6	1.21 (1.06, 1.38)	0.005
		TNFR2	1.19 (1.05, 1.36)	0.009
Intermittent claudication or lower extremity revascularization	0.02	Interleukin-6	1.36 (1.06, 1.74)	0.02
		TNFR2	1.31 (1.04, 1.64)	0.02

[a] A simultaneous test of whether at least one of the 12 biomarkers were related to PAD (PAD is the dependent variable). Covariates in multivariable model include age, sex, current cigarette smoking, number of pack-years of cigarette smoking, diabetes, fasting glucose, body mass index, waist circumference, total to HDL cholesterol ratio, fasting triglyceride, lipid lowering treatment, hypertension, aspirin use (≥3 per week), prevalent cardiovascular disease (excluding intermittent claudication), and hormone replacement therapy use (women only).

[b] Individual biomarkers significantly related to PAD after forward stepwise selection (PAD is the dependent variable) are displayed.

[c] Point estimate indicates relative change in odds of PAD (ABI level or presence versus absence of intermittent claudication or lower extremity revascularization) per 1-standard deviation increment in log-marker (1-standard deviation increment is 0.71 for log Interleukin-6 and 0.30 for log TNFR2).

[d] The ABI was categorized as follows: <0.9, 0.9–1.0, >1.0.

level increased by 21 and 19% per a 1-standard deviation increase in interleukin-6 and TNFR2, respectively. Similar results were observed for intermittent claudication or revascularization.

3.3. Secondary analyses

In analyses of ABI, excluding participants with prevalent CVD ($n = 2496$), the global test examining whether the markers as a group were related to ABI was not significant ($p = 0.34$). The forward stepwise selection regression retained only interleukin-6 with a nearly identical point estimate (estimate 1.21, 95% confidence interval 1.05, 1.39, $p = 0.01$). The analysis of clinical PAD was not run in participants free of prevalent CVD due to small numbers ($n = 43$). No significant interactions were noted for sex and age with regard to the association between biomarkers and ABI.

In adjusted regression models examining each marker separately, C-reactive protein, interleukin-6, fibrinogen, tumor necrosis factor alpha, and TNFR2 were significantly inversely related to ABI level (p-values ranging from <0.0001 to 0.02). For each biomarker, we exponentiated the adjusted mean log-transformed biomarker and its 95% confidence interval to obtain its adjusted geometric mean and corresponding 95% confidence interval (Table 4). Similar markers were associated with clinical PAD (C-reactive protein, interleukin-6, and TNFR2; p-values ranging from <0.0001 to 0.01) with the following exceptions: fibrinogen and tumor necrosis factor alpha were not significantly associated (data not shown).

4. Discussion

4.1. Principal findings

In our cross-sectional community-based study, we examined the relations of a panel of 12 inflammatory biomarkers to PAD assessed by ABI, and by clinically defined intermittent claudication and/or lower extremity revascularization. Interleukin-6 and TNFR2 were significantly related to both measures of PAD. In secondary analyses, examining the relation of each marker separately to ABI, we observed additional significant inverse relations for C-reactive protein, fibrinogen, and tumor necrosis factor alpha after adjusting for known risk factors.

4.2. Interleukin-6 and PAD

Interleukin-6 is known to play a critical role in the inflammatory process with both pro-inflammatory and anti-inflammatory effects that include the stimulation of C-reactive protein, fibrinogen and other acute phase reactants and increased endothelial cell adhesiveness. In accordance with our results, in a small study of patients with intermittent claudication, interleukin-6 concentrations were higher in patients compared to healthy controls both at rest and after treadmill exercise ($p < 0.001$) suggesting that this marker is associated with peripheral atherosclerosis [25]. In a hospital-based investigation of the interleukin-6 G (-174) C geno-

Table 4
Secondary analyses: multivariable-adjusted regression of individual biomarkers on ankle-brachial index[a]

Biomarker	Geometric means and 95% confidence intervals[b]			
	Ankle-brachial index level			
	<0.9	0.9–1.0	>1.0–1.4	*p*-Value
C-reactive protein	2.91 (2.42, 3.51)	2.44 (2.15, 2.77)	2.19 (2.11, 2.28)	0.007
Interleukin-6	3.64 (3.21, 4.13)	3.09 (2.84, 3.37)	2.86 (2.78, 2.93)	0.0005
Fibrinogen	394 (381, 407)	372 (363, 381)	372 (369, 374)	0.005
Tumor necrosis factor alpha	1.43 (1.29, 1.58)	1.32 (1.24, 1.42)	1.25 (1.22, 1.27)	0.02
TNFR2	2258 (2142, 2379)	2087 (2013, 2164)	2009 (1987, 2030)	<0.0001

[a] Biomarkers with $p < 0.05$ displayed.
[b] For each biomarker (dependent variable), we exponentiated the adjusted mean log-transformed biomarker and its 95% confidence interval to obtain its adjusted geometric mean and corresponding 95% confidence interval. Covariates in multivariable model include covariates listed in Table 3 legend.

types, in patients with type II diabetes with and without PAD, the GG genotype and higher plasma concentrations of interleukin-6 and other inflammatory markers were more common in PAD patients [26]. The investigators of that report hypothesize that the GG genotype promotes PAD in patients with diabetes by inducing release of interleukin-6 which in turn results in increased concentrations of other biomarkers such as C-reactive protein. In the Edinburgh Artery Study, inflammatory marker concentrations, including interleukin-6, were significantly elevated at baseline in participants who developed symptomatic PAD during follow-up [27]. In that study, interleukin-6 was a predictor of incident PAD but the association was attenuated with adjustment for CVD risk factors.

Elevated concentrations of interleukin-6 have been noted in a community-based sample of older participants with a low ABI [28], a finding similar to our study. Furthermore, interleukin-6 was predictive of PAD progression defined by declining ABI over 12 years of follow-up even after adjusting for traditional risk factors and other inflammatory markers (C-reactive protein, intracellular adhesion molecule-1, vascular adhesion molecule-1, and E-selectin) [16], and hemostatic factors [17]. Moreover, interleukin-6 was the only inflammatory marker independently associated with ABI decline in persons free of baseline PAD. The independent predictive value of interleukin-6 in relation to PAD progression may reflect its role in both inflammatory and hemostatic processes. Additionally, interleukin-6 predicts the development of type II diabetes [29] and hypertension [30], both significant predictors of PAD. Finally, interleukin-6 predicts risk for incident CVD events [5] and persons with coronary disease have nearly a threefold risk of intermittent claudication. Hence, the association between interleukin-6 and PAD is likely mediated through a variety of complex interrelated biologic pathways and appears to extend to early peripheral atherosclerosis, atherosclerosis progression, and incident symptomatic disease.

4.3. TNFR2 and PAD

Tumor necrosis factor alpha is a pro-inflammatory cytokine that affects vascular tissues including endothelial cells. Tumor necrosis factor alpha exerts its biologic effects through two cell surface receptors, TNFR1 and TNFR2. However, the role of TNFR2 in the regulation of inflammatory responses in endothelial cells is unclear. In mice, the proatherogenic effect of tumor necrosis factor alpha was mediated primarily through TNFR2 [31]. Further, in mice endothelial TNFR2 is essential for tumor necrosis factor alpha induced leukocyte–endothelial-cell interaction which mediates several important steps of the inflammatory response including leukocyte rolling, adhesion, and transmigration [32]. A potential mechanism for TNFR2 mediated endothelial dysfunction is the down-regulation of lysyl oxidase, a key enzyme in extracellular matrix maturation. TNFR2 has been shown to be involved in lysyl oxidase down-regulation, which in turn is associated with endothelial dysfunction [33]. To our knowledge there is only one small study of patients with intermittent claudication and critical limb ischemia demonstrating elevated tumor necrosis factor receptor concentrations compared with controls [34].

4.4. Other markers and PAD

In the Physician's Health Study, C-reactive protein was the strongest nonlipid predictor of the development of symptomatic PAD [8]. In that report both C-reactive protein and fibrinogen improved risk prediction for PAD. However, the two markers were correlated and C-reactive protein was the stronger predictor of risk. The associations between C-reactive protein and fibrinogen and incident PAD were confirmed by the Edinburgh Artery Study and persisted after accounting for risk factors and prevalent CVD [27]. Additional associations between C-reactive protein and ABI, PAD progression, and risk for adverse CVD events among individuals with PAD have been reported [11,13,28]. However, these prior reports were limited as only a few other biomarkers were examined. If we considered each marker separately, both C-reactive protein and fibrinogen were associated with PAD. But in our global model that considered all 12 biomarkers conjointly neither C-reactive protein nor fibrinogen was significantly associated with PAD. One possible explanation may be that the effect of C-reactive protein and fibrinogen may be mediated through interleukin-6 and TNFR2. It is known that interleukin-6 up-regulates both C-reactive protein and fibrinogen and that all three biomarkers are correlated.

4.5. Strengths and limitations

The strengths of the present study include the community-based sample, the simultaneous measurement of a panel of biomarkers, and the direct measurement of clinical factors previously reported to be correlated with PAD and or the inflammatory markers. Several limitations merit comment. The study is cross-sectional and thus we cannot infer that the associations between PAD and the inflammatory markers are causal. We suspect, but cannot establish with the current study design that the relations are bidirectional, with inflammation contributing to PAD and PAD exacerbating systemic inflammation. Conversely, we note that we may have failed to detect small to modest associations. Medication usage (aspirin and lipid lowering treatments) may have altered some inflammatory marker concentrations. Since medication usage was higher in those with PAD, our results may have been biased toward a null result. In addition, the estimated effect sizes of the observed associations were modest; we acknowledge that statistical significance is not synonymous with clinical significance. We acknowledge that walk test data would have enhanced

the accuracy of a PAD diagnosis. Lastly, our sample is primarily white, limiting the ability to generalize our results to other racial and ethnic groups.

5. Conclusions

In a community-based sample interleukin-6 and TNFR2 were significantly associated with PAD accounting for established risk factors. Their effects appear to be independent of each other suggesting that each marker represents a distinct biologic pathway mediating the complex process of vascular inflammation in peripheral atherosclerosis. Further research is needed to establish the role of these markers in predicting incident clinical PAD events and disease progression and to determine whether therapies targeting these markers alter prognosis in patients with PAD.

Conflict of interest

None.

Acknowledgement

The authors appreciate the statistical insights provided by Joseph M. Massaro, PhD.

References

[1] 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(6):738–43.
[2] 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(3):388–93.
[3] Resnick HE, Lindsay RS, McDermott MM, Devereux RB, Jones KL, Fabsitz RR, et al. Relationship of high and low ankle brachial index to all-cause and cardiovascular disease mortality: the Strong Heart Study. Circulation 2004;109(6):733–9.
[4] Murabito JM, Evans JC, Larson MG, Nieto K, Levy D, Wilson PW. The ankle-brachial index in the elderly and risk of stroke, coronary disease, and death: the Framingham Study. Arch Intern Med 2003;163(16):1939–42.
[5] Cesari M, Penninx BW, Newman AB, Kritchevsky SB, Nicklas BJ, Sutton-Tyrrell K, et al. Inflammatory markers and onset of cardiovascular events: results from the Health ABC study. Circulation 2003;108(19):2317–22.
[6] Tuomisto K, Jousilahti P, Sundvall J, Pajunen P, Salomaa V. C-reactive protein, interleukin-6 and tumor necrosis factor alpha as predictors of incident coronary and cardiovascular events and total mortality. A population-based, prospective study. Thromb Haemost 2006;95(3):511–8.
[7] Ridker PM, Stampfer MJ, Rifai N. Novel risk factors for systemic atherosclerosis: a comparison of C-reactive protein, fibrinogen, homocysteine, lipoprotein(a), and standard cholesterol screening as predictors of peripheral arterial disease. JAMA 2001;285(19):2481–5.
[8] Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Plasma concentration of C-reactive protein and risk of developing peripheral vascular disease. Circulation 1998;97(5):425–8.
[9] Khera A, de Lemos JA, Peshock RM, Lo HS, Stanek HG, Murphy SA, et al. Relationship between C-reactive protein and subclinical atherosclerosis: the Dallas Heart Study. Circulation 2006;113(1):38–43.
[10] Hulthe J, Wikstrand J, Fagerberg B. Relationship between C-reactive protein and intima-media thickness in the carotid and femoral arteries and to antibodies against oxidized low-density lipoprotein in healthy men: the Atherosclerosis and Insulin Resistance (AIR) study. Clin Sci (Lond) 2001;100(4):371–8.
[11] Folsom AR, Pankow JS, Tracy RP, Arnett DK, Peacock JM, Hong Y, et al. Association of C-reactive protein with markers of prevalent atherosclerotic disease. Am J Cardiol 2001;88(2):112–7.
[12] Tracy RP, Psaty BM, Macy E, Bovill EG, Cushman M, Cornell ES, et al. Lifetime smoking exposure affects the association of C-reactive protein with cardiovascular disease risk factors and subclinical disease in healthy elderly subjects. Arterioscler Thromb Vasc Biol 1997;17(10):2167–76.
[13] McDermott MM, Green D, Greenland P, Liu K, Criqui MH, Chan C, et al. Relation of levels of hemostatic factors and inflammatory markers to the ankle brachial index. Am J Cardiol 2003;92(2):194–9.
[14] van dM I, De Maat MP, Hak AE, Kiliaan AJ, del Sol AI, van der Kuip DA, et al. C-reactive protein predicts progression of atherosclerosis measured at various sites in the arterial tree: the Rotterdam Study. Stroke 2002;33(12):2750–5.
[15] Beckman JA, Preis O, Ridker PM, Gerhard-Herman M. Comparison of usefulness of inflammatory markers in patients with versus without peripheral arterial disease in predicting adverse cardiovascular outcomes (myocardial infarction, stroke, and death). Am J Cardiol 2005;96(10):1374–8.
[16] Tzoulaki I, Murray GD, Lee AJ, Rumley A, Lowe GD, Fowkes FG. C-reactive protein, interleukin-6, and soluble adhesion molecules as predictors of progressive peripheral atherosclerosis in the general population: Edinburgh Artery Study. Circulation 2005;112(7):976–83.
[17] Tzoulaki I, Murray GD, Price JF, Smith FB, Lee AJ, Rumley A, et al. Hemostatic factors, inflammatory markers, and progressive peripheral atherosclerosis: the Edinburgh Artery Study. Am J Epidemiol 2006;163(4):334–41.
[18] Hoogeveen RC, Morrison A, Boerwinkle E, Miles JS, Rhodes CE, Sharrett AR, et al. Plasma MCP-1 level and risk for peripheral arterial disease and incident coronary heart disease: Atherosclerosis Risk in Communities study. Atherosclerosis 2005;183(2):301–7.
[19] 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(6):961–5.
[20] Jefferson AL, Massaro JM, Wolf PA, Seshadri S, Au R, Vasan RS, et al. Inflammatory biomarkers are associated with total brain volume: the Framingham Heart Study. Neurology 2007;68(13):1032–8.
[21] Abbott RD, McGee DL. The Framingham Study: An Epidemiologic Investigation of Cardiovascular Disease. Section 37. The Probability of Developing Certain Cardiovascular Diseases in Eight Years at Specified Values of Some Characteristics. Bethesda, MD: National Heart, Lung and Blood Institute; 1987.
[22] Benjamin EJ, Dupuis J, Larson MG, Lunetta KL, Booth SL, Govindaraju DR, et al. Genome-wide association with select biomarker traits in the Framingham Heart Study. BMC Med Genet 2007;8(Suppl 1):S11.
[23] Schnabel R, Larson MG, Dupuis J, Lunetta KL, Lipinska I, Meigs JB, et al. Relations of inflammatory biomarkers and common genetic variants with arterial stiffness and wave reflection. Hypertension 2008.
[24] SAS Institute Inc., version 8. Cary, NC: SAS Institute Inc.; 1999.
[25] Signorelli SS, Mazzarino MC, Di Pino L, Malaponte G, Porto C, Pennisi G, et al. High circulating levels of cytokines (IL-6 and TNFalpha), adhesion molecules (VCAM-1 and ICAM-1) and selectins in patients with peripheral arterial disease at rest and after a treadmill test. Vasc Med 2003;8(1):15–9.
[26] Libra M, Signorelli SS, Bevelacqua Y, Navolanic PM, Bevelacqua V, Polesel J, et al. Analysis of G(-174)C IL-6 polymorphism and plasma concentrations of inflammatory markers in patients with type 2 diabetes and peripheral arterial disease. J Clin Pathol 2006;59(2):211–5.
[27] Tzoulaki I, Murray GD, Lee AJ, Rumley A, Lowe GD, Fowkes FG. Inflammatory, haemostatic, and rheological markers for incident peripheral arterial disease: Edinburgh Artery Study. Eur Heart J 2007;28(3):354–62.
[28] McDermott MM, Guralnik JM, Corsi A, Albay M, Macchi C, Bandinelli S, et al. Patterns of inflammation associated with peripheral arterial disease: the InCHIANTI study. Am Heart J 2005;150(2):276–81.
[29] Spranger J, Kroke A, Mohlig M, Hoffmann K, Bergmann MM, Ristow M, et al. Inflammatory cytokines and the risk to develop type 2 diabetes: results of the prospective population-based European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Diabetes 2003;52(3):812–7.
[30] Sesso HD, Wang L, Buring JE, Ridker PM, Gaziano JM. Comparison of interleukin-6 and C-reactive protein for the risk of developing hypertension in women. Hypertension 2007;49(2):304–10.
[31] Chandrasekharan UM, Mavrakis L, Bonfield TL, Smith JD, DiCorleto PE. Decreased atherosclerosis in mice deficient in tumor necrosis factor-alpha receptor-II (p75). Arterioscler Thromb Vasc Biol 2007;27(3):e16–7.
[32] Chandrasekharan UM, Siemionow M, Unsal M, Yang L, Poptic E, Bohn J, et al. Tumor necrosis factor alpha (TNF-alpha) receptor-II is required for TNF-alpha-induced leukocyte–endothelial interaction in vivo. Blood 2007;109(5):1938–44.
[33] Rodriguez C, Alcudia JF, Martinez-Gonzalez J, Raposo B, Navarro MA, Badimon L. Lysyl oxidase (LOX) down-regulation by TNFalpha: a new mechanism underlying TNFalpha-induced endothelial dysfunction. Atherosclerosis 2007.
[34] Fiotti N, Giansante C, Ponte E, Delbello C, Calabrese S, Zacchi T, et al. Atherosclerosis and inflammation. Patterns of cytokine regulation in patients with peripheral arterial disease. Atherosclerosis 1999;145(1):51–60.

General Internal Medicine
Boston University School of Medicine
2009 Publications – N-Z

1. **Nader CM**, Tsevat J, Justice AC, Mrus JM, Levin F, Kozal MJ, Mattocks K, Farber S, Rogers M, Erdos J, Brandt C, Kudel I, Braithwaite R. Development of an electronic medical record-based clinical decision support tool to improve HIV symptom management. AIDS Patient Care and STDs. 2009;23(7):521-28.

2. **Naimi TS**, Nelson DE, Brewer RD. Driving after binge drinking. American Journal of Preventive Medicine 2009;37(4):314-20.

3. Nelson DE, **Naimi TS**, Brewer RD, Nelson HA. State alcohol use estimates among youth and adults, 1993-2005. American Journal of Preventive Medicine. 2009;36(3):218-24.

4. Oates DJ, **Paasche-Orlow MK**. Health literacy: communication strategies to improve patient comprehension of cardiovascular health. Circulation. Feb 2009;119(7):1049-51.

5. **Paasche-Orlow MK**, Jacob DM, Hochhauser M, Parker RM. National survey of patients' bill of rights statutes. J Gen Intern Med. Apr 2009;24(4):489-94.

6. **Paasche-Orlow MK**, McCaffery K, Wolf MS. Bridging the international divide for health literacy research. Patient Educ Couns. Jun 2009;75(3):293-4.

7. Parikh NI, Gona P, Larson MG, Fox CS, Benjamin EJ, **Murabito JM**, O'Donnell CJ, Vasan RS, Levy D. Long-term trends in myocardial infarction incidence and case fatality in the National Heart, Lung, and Blood Institute's Framingham Heart study. Circulation. Mar 2009;119(9):1203-10.

8. Parikh NI, Hwang SJ, Ingelsson E, Benjamin EJ, Fox CS, Vasan RS, **Murabito JM**. Breastfeeding in infancy and adult cardiovascular disease risk factors. Am J Med. Jul 2009;122(7):656-63.

9. Perry JRB, Stolk L, Franceschini N, Lunetta KL, Zhai G, McArdle PF, Smith AV, Aspelund T, Bandinelli S, Boerwinkle E, Cherkas L, Eiriksdottir G, Estrada K, Ferrucci L, Folsom AR, Garcia M, Gudnason V, Hofman A, Karasik D, Kiel DP, Launer LJ, van Meurs J, Nalls MA, Rivadeneira F, Shuldiner AR, Singleton A, Soranzo N, Tanaka T, Visser JA, Weedon MN, Wilson SG, Zhuang V, Streeten EA, Harris TB, Murray A, Spector TD, Demerath EW, Uitterlinden AG, **Murabito JM**. Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. Nat Genet. Jun 2009;41(6):648-50.

10. Pololi L, Conrad P, Knight S, **Carr P**. A study of the relational aspects of the culture of academic medicine. Acad Med. 2009;84(1):106-14.

11. Pololi L, Kern DE, **Carr P**, Conrad P, Knight S. The culture of academic medicine: Faculty perceptions of the lack of alignment between individual and institutional values. JGIM. 2009;24(12):1289-95.

12. Pregler J, **Freund KM**, Kleinman M, Phipps MG, et al. The heart truth professional education campaign on women and heart disease: Needs assessment evaluation results. Journal of Women's Health. 2009;18(10):1541-7.

13. **Raj A**, **Cheng DM**, Krupitsky EM, Coleman S, **Bridden C**, **Samet JH**. Correlates of any condom use among Russian narcology patients reporting recent unprotected sex. AIDS and Behav. 2009;13:310-17.

14. **Raj A**, **Cheng DM**, Krupitsky EM, Levenson S, Egorova VY, **Meli S**, Zvartau EE, **Samet JH**. Binge drinking and unsafe sex: A study of narcology hospital patients from St. Petersburg, Russia. Subst Abus. 2009;30(3):213-22.

15. **Raj A**, Reed E, Santana MC, **Walley AY**, Welles SL, Horsburgh CR, Flores SA, Silverman JG. The associations of binge alcohol use with HIV/STI risk and diagnosis among heterosexual African American men. Drug and Alcohol Dependence. 2009;101:101-6.

16. Ramsey S, Whitley E, Mears VW, McKoy JM, Everhart RM, Caswell RJ, Fiscella K, Hurd TC, **Battaglia T**, Mandelblatt J. Evaluating the cost effectiveness of cancer patient navigation programs: Conceptual and practical issues. Cancer. Dec 2009;115(23):5394-403. doi: 10.1002/cncr.24603.

17. Rongey CA, Kanwal F, Hoang T, **Gifford AL**, Asch SM. Viral RNA testing in hepatitis C antibody-positive veterans. Am J Prev Med. Mar 2009;36(3):235-8.

18. **Rose AJ**, **Berlowitz DR**, Frayne SM, **Hykek EM**. Measuring quality of oral anticoagulation care: Extending quality measurements to a new field. Jt Comm J Qual Patient Saf. 2009;35(3):146-55.

19. **Rose AJ**, **Berlowitz DR**, **Manze M**, Orner MB, **Kressin NR**.Comparing methods of measuring treatment intensity in hypertension care. Circulation: Quality and Outcomes. 2009;2:385-91.

20. **Rose AJ**, **Berlowitz DR**, **Manze M**, Orner MB, **Kressin NR**. Intensifying therapy for hypertension despite suboptimal adherence. Hypertension 2009;54(3):524-29.

21. **Rose AJ**, **Hermos JA**, Frayne SM, Pogach LM, **Berlowitz DR**, Miller DR. Does opioid therapy affect quality of care for Diabetes Mellitus? American Journal of Managed Care. 2009;15(4):217-24.

22. **Rose AJ**, Ozonoff A, **Berlowitz DR**, **Henault LE**, **Hylek EM**. Warfarin dose management affects INR control. J Thromb Haemost 2009;7:94-101.

23. **Rose AJ**, Ozonoff A, Grant RW, **Henault LE**, **Hylek EM**. The epidemiology of sub-therapeutic anticoagulation in the United States. Circ Cardiovasc Qual Outcomes 2009;2:591-97.

24. Russell SL, Katz RV, **Kressin NR**, Green BL, Wang MQ, Claudio C, Tzvetkova K. Beliefs of women's risk as research subjects: A four-city study examining differences by sex and by race/ethnicity. Journal of Women's Health. 2009:18(2):235-43.

25. **Saitz R**, Palfai TP, **Cheng DM**, **Horton NJ**, Dukes K, Kraemer KL, Roberts MS, Guerriero RT, **Samet JH**. Some medical inpatients with unhealthy alcohol use may benefit from brief intervention. J Stud alc Drugs. 2009;70:426-35.

26. 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. American College of Cardiology Foundation. 2009;2(11):1271-81.

27. Schroy PC 3rd, Glick JT, Robinson P, Lydotes MA, Heeren TC, Prout M, **Davidson P**, Wong JB. A cost-effectiveness analysis of subject recruitment strategies in the HIPAA era: Results from a colorectal cancer screening adherence trial. Clin Trials. Dec 2009;6(6):597-609.

28. Sebastiani P, Hadley EC, Province M, Christensen K, Rossi W, Perls TT, **Ash AS**. A Family Longevity Selection Score (FLoSS): Ranking sibships by their longevity, size and availability for study. American Journal of Epidemiology. 2009;170(12):1555-62.

29. Singer SJ, Hartmann CW, **Hanchate A**, Zhao S, Meterko M, Shokeen P, Lin S, Gaba DM, Rosen AK. Comparing safety climate between two populations of hospitals in the United States. Health Research and Educational Trust. Oct 2009;44(5 pt 1):1563-83. doi: 10.1111/j.1475-6773.2009.00994.x.

30. **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):783-88.

31. Sobel RM, **Paasche-Orlow MK**, Waite KR, Rittner SS, Wilson EA, Wolf MS. Asthma 1-2-3: a low literacy multimedia tool to educate African American adults about asthma. J Community Health. Aug 2009;34(4):321-7.

32. Stahre MA, Brewer RD, Fonseca VP, **Naimi TS**. Binge drinking among U.S. active duty military personnel. American Journal of Preventive Medicine. 2009;36(3):208-17.

33. **Stillman M**. A foundation of failure. J Gen Intern Med. 2009;24(4):538-9.

34. Tokuda Y, Doba N, Butler JP, **Paasche-Orlow MK**. Health literacy and physical and psychological wellbeing in Japanese adults. Patient Educ Couns. Jun 2009;75(3):411-7.

35. **Tsui JI**, Maselli J, Gonzales R. Sociodemographic trends in national ambulatory care visits for hepatitis C virus infection. Dig Dis Sci. Dec 2009;54(12):2694-8.

36. **Tsui J**, Vittinghoff E, Anastos K, Augenbraun M, Young M, Nowicki M, Cohen MH, Peters MG, Golub ET, Szczech L. Hepatitis C seropositivity and kidney function decline among women with HIV: Data from the Women's Interagency HIV Study. Am Journal of Kidney Diseases. 2009;54(1):43-50.

37. **Tsui J**, Vittinghoff E, Hahn JA, Evans JL, Davidson PJ, Page K. Risk behaviors after hepatitis C virus seroconversion in young injection drug users in San Francisco. Drug and Alcohol Dependence. 2009;105:160-63.

38. **Tsui J**, Whooley MA, Monto A, Seal K, Tien PC, Shlipak M. Associations of hepatitis C virus seropositivity with inflammatory markers and heart failure in persons with coronary heart disease: Data from the heart and soul. J of Cardiac Failure. 2009;15(5):451-56.

39. Volandes AE, Mitchell SL, Gillick MR, Chang Y, **Paasche-Orlow MK**. Using video images to improve the accuracy of surrogate decision-making: a randomized controlled trial. J Am Med Dir Assoc. Oct 2009;10(8):575-80.

40. Volandes AE, **Paasche-Orlow MK**, Barry MJ, Gillick MR, Minaker KL, Chang Y, Cook EF, Abbo ED, El-Jawahri A, Mitchell SL. Video decision support tool for advance care planning in dementia: Randomized controlled trial. BMJ. 2009 May 28;338:b2159.

41. **Walley AY**, Farrar D, **Cheng DM**, **Alford DP**, **Samet JH**. Are opioid dependence and methadone maintenance treatment (MMT) documented in the medical record? A patient safety issue. J Gen Intern Med. 2009; 24(9):1007–11.

42. Wee CC, Pratt JS, Fanelli R, Samour PQ, Trainor LS, **Paasche-Orlow MK**. Best practice updates for informed consent and patient education in weight loss surgery. Obesity (Silver Spring). May 2009;17(5):885-8.

43. Wharam JF, **Paasche-Orlow MK**, Farber NJ, Sinsky C, Rucker L, Rask KJ, Figaro MK, Braddock C 3rd, Barry MJ, Sulmasy DP. High quality care and ethical pay-for-performance: a Society of General Internal Medicine policy analysis. J Gen Intern Med. Jul 2009;24(7):854-9.

44. Wilper AP, Woolhandler SJ, Boyd JW, **Lasser KE**, McCormick D, Bor DH, Himmelstein DU. The health and health care of U.S. prisoners: A nationwide survey. American Journal of Public Health. 2009;99(4):666-72.

45. Wilper AP, Woolhandler SJ, **Lasser KE**, McCormick D, Bor DH, Himmelstein DU. Health insurance and mortality. American Journal of Public Health. 2009;99(12):2289-95.

46. Witt DW, 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(5):952-6.

47. Witt DW, Delate T, McCool KH, Dowd MB, Clark NP, Crowther MA, Garcia DA, Ageno W, Dentali F, **Hylek EM**, Rector WG. Incidence and predictors of bleeding following polypectomy in patients receiving and not receiving anticoagulation therapy. J Thromb Haemost. Dec 2009;7(12):1982-89.

48. Wright WG, Jones JA, Spiro A III, Rich SE, **Kressin NR.** Use of patient self-report oral health outcome measures in assessment of dental treatment outcomes. Journal of Public Health Dentistry. 2009:69(2):95-103.

AIDS PATIENT CARE and STDs
Volume 23, Number 7, 2009

DOI: 10.1089/apc.2008.0209

ORIGINAL ARTICLE

Development of an Electronic Medical Record-Based Clinical Decision Support Tool to Improve HIV Symptom Management

Claudia M. Nader, M.D.,[1] Joel Tsevat, M.D., M.P.H.,[2] Amy C. Justice, M.D., Ph.D.,[3,4] Joseph M. Mrus, M.D., M.Sc.,[5] Forrest Levin, M.S.,[3,4] Michael J. Kozal, M.D.,[4,6] Kristin Mattocks, M.D., M.P.H.,[3,4] Steven Farber, PA-C., J.D.,[4,6] Michelle Rogers, Ph.D.,[7] Joseph Erdos, M.D.,[4] Cynthia Brandt, M.D.,[4,8] Ian Kudel, Ph.D.,[7] and Ronald Braithwaite, M.D., M.Sc.[3,4]

Abstract

Common symptoms associated with HIV disease and its management are often underrecognized and undertreated. A clinical decision support tool for symptom management was developed within the Veterans Health Administration electronic medical record (EMR), aiming at increasing provider awareness of and response to common HIV symptoms. Its feasibility was studied in March to May 2007 by implementing it within a weekly HIV clinic, comparing a 4-week intervention period with a 4-week control period. Fifty-six patients and their providers participated in the study. Patients' perceptions of providers' awareness of their symptoms, proportion of progress notes mentioning any symptom(s) and proportion of care plans mentioning any symptom(s) were measured. The clinical decision support tool used portable electronic "tablets" to elicit symptom information at the time of check-in, filtered, and organized that information into a concise and clinically relevant EMR note available at the point of care, and facilitated clinical responses to that information. It appeared to be well accepted by patients and providers and did not substantially impact workflow. Although this pilot study was not powered to detect effectiveness, 25 (93%) patients in the intervention group reported that their providers were very aware of their symptoms versuas 27 (75%) control patients ($p = 0.07$). The proportion of providers' notes listing symptoms was similar in both periods; however, there was a trend toward including a greater number of symptoms in intervention period progress notes. The symptom support tool seemed to be useful in clinical HIV care. The Veterans Health Administration EMR may be an effective "laboratory" for developing and testing decision supports.

Introduction

Many of the symptoms related to HIV disease, its complications, and/or its management (e.g., fatigue, pain, diarrhea, sleep disturbances) are underrecognized and therefore undertreated in many care settings.[1–5] Although antiretroviral therapy (ART) has greatly increased life expectancy, it may precipitate side effects that substantially decrease quality of life[1] and may create a barrier to the high adherence levels necessary for maximum benefit.[2] Survey instruments that detect symptoms common in HIV care have been developed to facilitate effective symptom management[6–8] but those instruments have not been regularly incorporated into clinical practice.

We postulated that providers underrecognize and undertreat common symptoms because of the substantial time burden required to ask about the many individual symptoms, and because of providers' lack of comfort regarding appropriate management strategies. At the same time, we observed that the growing sophistication of clinical decision support tools may alleviate such barriers,[9–12] and that the advanced electronic medical record (EMR) system of the Veterans

[1]Department of Medicine, Boston University School of Medicine, Boston, Massachusetts.
[2]Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio.
[3]General Internal Medicine, [6]Section of Infectious Diseases, [8]Section of Medical Informatics, Yale University School of Medicine, New Haven, Connecticut.
[4]Connecticut VA Healthcare System, West Haven, Connecticut.
[5]Tibotec Therapeutics, Bridgewater, New Jersey.
[7]Veterans Affairs Medical Center, Cincinnati, Ohio.

Health Administration (VHA) may serve as a "laboratory" to test the feasibility of new types of clinical decision supports. With this in mind, we constructed an EMR-based clinical decision support tool to increase providers' awareness of and responses to common symptoms, and tested the feasibility of incorporating our tool into routine care.

Methods

We sought to construct a clinical decision support tool that would elicit information about symptoms at the time of check-in for a routine clinic visit, organize that information to emphasize what is most useful for clinical care, present that information in an easy-to-use form at the point-of-care, and recommend clinical responses based on that information. We chose these design factors because they encompass a broad range of information management necessary for clinical care. In addition, many of those design factors have been shown to help integrate computerized systems into clinical workflow.[9] Because our ultimate objective was to improve symptom management in HIV care, we refer to our decision support tool as the Tool to Enhance Management of Symptoms (TEMS). Because clinical guidelines suggest assessing and reinforcing adherence at each visit[13] and because adherence rates may be related to the prevalence of side effects, we augmented the symptom information with a briefer query regarding medication adherence.

Clinical setting

We aimed to implement our TEMS within a weekly half-day HIV clinic at our urban Veterans Affairs (VA) Medical Center. Approximately 30 patients are seen in a typical clinic session, which is staffed by two to four attending physicians (from General Internal Medicine and Infectious Diseases), one physician assistant, one clinic coordinator, and several rotating trainees (fellows and residents).

Design of clinical decision support tool

We sought to base recommended response strategies for symptoms on evidence-based guidelines when available, and to supplement those with expert opinion from our senior clinic personnel. The site principal investigator (R.B.), along with the clinic director and infectious diseases trainee, constructed response strategies based on their expert opinion, in conjunction with HIV textbooks,[14,15] recent Department of Health and Human Services guidelines,[13] well-regarded HIV care websites,[16–18] and other sources. The informatics infrastructure underlying the TEMS was developed by a senior programmer having 14 years of experience with VHA data systems (F.L.), working in conjunction with the Chief Information Officer of the Connecticut VA Healthcare System (J.E.) and the director of the Informatics Fellowship program at our institution (C.B.). The senior programmer required approximately 200 hours of programming time to implement the intervention.

We describe our design aims by discussing each of the four main design factors that we sought to incorporate into TEMS: elicitation, organization, presentation, and recommendation (Fig. 1 and Appendix).

Elicitation. We strove to elicit information in a manner that would minimize respondent burden, would not interrupt the clinical workflow, and would direct information transfer into the EMR without manual transcription. For those reasons, we sought to identify surveys of HIV symptoms that were validated yet comparatively brief, and could be administered by using portable devices that could interface with the EMR.

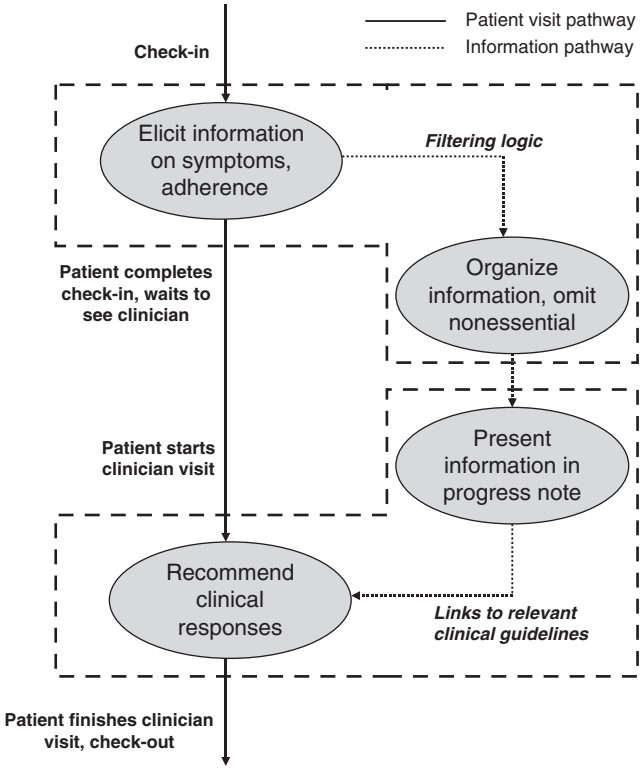


FIG. 1. Schematic diagram of design factors for Tool to Enhance Management of Symptoms (TEMS). These factors encompass a wide spectrum of information management necessary for clinical care, and are synchronized with the workflow of a typical clinic visit. Information is elicited at the same time that vital signs are measured, and the computer-generated progress note can be viewed at the same time that other patient information is used for decision making.

Among several HIV symptom indices that we considered as candidates for inclusion in TEMS, only 2 have been validated in the ART era.[6–8] We chose the 20-item HIV Symptom Index[6] because it has been widely used in clinical studies of HIV/AIDS, including Adult AIDS Clinical Trials Group (AACTG) studies and the Veterans Aging Cohort Study (VACS).[19] An example of an item in the index is, "How much have you been bothered by fatigue or loss of energy?" with possible responses, "I do not have this symptom." "I have it and it doesn't bother me." "I have it and it bothers me a little." "I have it and it bothers me." and "I have it and it bothers me a lot."

With the intent of minimizing additional respondent burden, we chose to gather information on medication adherence by using a single question, which we based on a patient adherence instrument developed by the Outcomes Committee of the AACTG.[20] The question asks, "When was the last time you missed one or more doses of your HIV medications?" with possible responses of "today," "yesterday," "within the last week," "within the last month," and "more than 1 month ago."

Organization. We strove to organize the information to deemphasize any that was not clinically important or actionable, as evidence suggests that providers often ignore clinical decision support information that is insufficiently specific.[12] We sought to use two distinct "filters" on the survey information, the first excluding information pertaining to symptoms that were not sufficiently bothersome to prioritize, and the second limiting the detail of information when a patient appeared to endorse a great number of unrelated symptoms simultaneously (such patients would be unlikely to respond to any 1 or 2 interventions to alleviate symptoms, and may have a more global problem such as depression).

Presentation. We strove to ensure that symptom information was presented through an interface that was simple and clear and had the capacity to be linked to response strategies. For that reason, we collected the symptom information on a "tablet" thin-client computer and aimed to present the information within an EMR progress note that could be accessed during the patient encounter. We aimed to construct flexible templates for notes (i.e., notes that could have alternative structures depending upon the level of information detail), so that notes could generally remain as brief as possible yet "telescope" into longer notes when it was essential to communicate additional information.

Recommendation. We sought to provide clinical recommendations by developing an interface that would anticipate clinicians' information needs and address them rapidly, in real time.[12] Our goals were (1) to fit into the user's workflow[12] (e.g., allowing the clinician to place an order to address the symptom while viewing the note that alerted her to that symptom); (2) to minimize medical knowledge barriers (e.g., permitting the clinician to retrieve the preferred approach for addressing a symptom in case she does not already know it); and (3) to minimize institution-specific knowledge barriers (e.g., allowing the clinician to learn how a particular drug or laboratory test is listed in the EMR order menu in case she does not already know it). Consequently, we sought to embed hyperlinks in the note to clinical and diagnostic algorithms, experts' recommendations, and relevant orders (including medications, tests, and consultations).

Study design

Four clinic sessions comprised the control phase, and were followed by four clinic sessions comprising the intervention phase. During the control phase (March 2007), TEMS was not activated and therefore participants did not complete the HIV symptom survey. During the intervention phase (mid-April through mid-May 2007), TEMS was activated, and all participants were asked to complete the HIV symptom survey. Therefore, only in the intervention phase did providers receive the computer-generated progress notes and have access to the other functionalities of TEMS. In both the control phase and the intervention phase, all participants received a one-item postvisit survey in which they were queried about their perception of their providers' level of awareness about their symptoms. The item asked, "How aware of your symptoms do you think your health care provider was?" with possible responses, "not at all aware," "a little bit aware," "somewhat aware," "quite a bit aware," and "very much aware."[21,22]

To maximize the generalizability of the study, we did not impose any inclusion criteria other than having a clinic appointment during the period of the study (March 2007–May 2007), having HIV infection, and providing informed consent. Patients were excluded only if they were making their first visit to the clinic because a thorough symptom review would likely be performed as part of routine care at initial visits. Each individual could participate in each phase of the study no more than once (i.e., a patient could participate once in both phases of the study, but could not participate in either phase twice); however, we did not require that patients who participated in one phase also participate in the other phase. This study was approved by the Institutional Review Board at the Connecticut VA Healthcare System and the Human Investigations Committee at Yale University. The sponsor did not have any role in the collection, analysis, or interpretation of data; in the decision to submit study results for publication; or in the drafting or revision of the manuscript.

Outcomes

In a subset of participants, we assessed the acceptability of the system and the information provided to physicians by having a cognitive engineer (M.R.) review the entire process. By using a human factors approach, the cognitive engineer analyzed barriers to: (1) entering symptoms into the tablet personal computer; (2) transferring the symptoms electronically into the VHA EMR; and (3) having providers act on the symptom information. In addition, the study coordinator (K.M.) independently noted which barriers seemed to be most prevalent for participants.

Because this was a pilot study and its primary outcome was to assess the feasibility of TEMS, the study was not powered to detect clinically significant changes in the effectiveness of symptom recognition and/or management. However, this study did incorporate two prospectively defined effectiveness measures as secondary outcomes. First, we evaluated responses on the postvisit survey, comparing the proportion of intervention versus control participants who thought that their providers were "very aware" of their symptoms by using the χ^2 test for proportions. Second, we performed chart reviews to assess the proportion of progress notes that included at least one symptom addressed by TEMS. The reviews were completed in a blinded fashion and in duplicate (C.N. and S.F.); the agreement between the blinded reviewers was high (κ score $= 0.862$ for having at least one symptom mentioned in progress note; κ score $= 1.00$ for having at least 1 symptom mentioned in the treatment plan). Because only a minority ($N = 8$) of patients participated in both the control and intervention phases and therefore could serve as their own control, each progress note during the intervention period was compared with the most recent progress note that preceded the intervention period, even if it pertained to a visit prior to the control phase.

Results

Of 60 clinic patients invited to participate, 56 (93%) agreed to participate, and 55 (98%) completed all parts of the study (1 patient left before completing the postvisit survey). Eight patients (14%) participated in both the intervention and control phases, 28 patients (50%) participated in the control phase only, and 20 patients (36%) participated in the intervention

Table 1. Baseline Characteristics of Study Participants

Characteristics	*Intervention group* (n = *28*)	*Control group* (n = *36*)	*All* (n = *56*)[a]
Mean (SD) age, years	53.6 (11.2)	55.7 (8.4)	54.4 (9.5)
Race			
Black, *n* (%)	21 (75.0)	24 (66.7)	38 (67.9)
White, *n* (%)	5 (17.9)	10 (27.8)	15 (26.8)
Hispanic, *n* (%)	2 (7.1)	2 (5.6)	3 (5.4)
Male gender, *n* (%)	27 (96.4)	35 (97.2)	55 (9.8)
$CD4^+$ count (cells/mm^3)			
Mean (SD)	410.7 (241.6)	388.0 (218.5)	409.5 (225.6)
Median	410	383	394
HIV RNA ≤400 copies/mL, *n* (%)	18 (64.3)	24 (66.7)	38 (67.9)
Type of current ARV therapy			
NNRTI-based regimen, *n* (%)	4 (19.0)	6 (19.3)	10[b] (20.8)
PI-based regimen, *n* (%)	2 (9.5)	3 (9.7)	3[b] (6.2)
Boosted PI-based regimen, *n* (%)	13 (61.9)	13 (41.9)	24[b] (50.0)
Other, *n* (%)	2 (9.5)	9 (29.0)	11[b] (22.9)
Coinfection with HCV, *n* (%)	17 (60.7)	22 (61.1)	33 (58.9)

[a]Eight patients participated in both the intervention group and the control group.
[b]Some patients were not taking any ARV medication (total of 8).
ARV, antiretroviral; NNRTI, non-nucleoside reverse transcriptase inhibitor; PI, protease inhibitor; HCV, hepatitis C virus; SD, standard deviation.

phase only (Table 1). The mean (standard deviation [SD]) age of the patients was 54.4 (9.5); 40 (73%) were minorities and 33 (59%) had hepatitis C coinfection. Approximately two thirds had plasma HIV RNA levels below 400 copies per milliliter, and the median CD4 count was approximately 400 cells per milliliter. Most (79%) reported last missing a dose of antiretroviral medications more than 1 month ago, but a substantial minority (21%) reported more recent nonadherence (4% within the last week to month, 14% within the last day to week, and 4% on the day of the survey).

Feasibility

The vast majority of patients required fewer than 5 minutes to complete the survey, and none required more than 10 minutes. Approximately half of the patients were able to use

Table 2. Endorsement of Symptoms by Participants in the Intervention Phase

	Any symptom (Score ≥1)[a]	*Bothersome symptoms (Score ≥ 3)*[a]	*Of patients with bothersome symptoms, patients attributing them to HIV medications*
Fatigue or loss of energy, *n* (%)	19 (67.9)	9 (32.1)	0 (0.0)
Fever, chills, or sweats, *n* (%)	10 (35.7)	5 (17.9)	1 (20.0)
Feeling dizzy or lightheaded, *n* (%)	10 (35.7)	3 (10.7)	0 (0.0)
Pain, numbness, or tingling in hands or feet, *n* (%)	17 (60.7)	14 (50.0)[b]	1 (7.1)
Trouble remembering, *n* (%)	18 (64.3)	6 (21.4)	0 (0.0)
Nausea or vomiting, *n* (%)	4 (14.3)	3 (10.7)	0 (0.0)
Diarrhea or loose bowel movements, *n* (%)	15 (53.6)	3 (10.7)	3 (100.0)
Sad, down, or depressed, *n* (%)	13 (46.4)	4 (14.3)	1 (25.0)
Felt nervous or anxious, *n* (%)	17 (60.7)	6 (21.4)	2 (33.3)
Skin problems,[c] *n* (%)	15 (53.6)	8 (28.6)	1 (12.5)
Cough or trouble catching your breath, *n* (%)	12 (42.9)	6 (21.4)	0 (0.0)
Headache, *n* (%)	11 (39.3)	2 (7.1)	0 (0.0)
Loss of appetite or change in the taste of food, *n* (%)	16 (57.1)	6 (21.4)	2 (33.3)
Bloating, pain, or gas in your stomach, *n* (%)	14 (50.0)	5 (17.9)	1 (20.0)
Muscle aches or joint pain, *n* (%)	16 (57.1)	10 (35.7)	0 (0.0)
Problems with having sex,[d] *n* (%)	14 (50.0)	9 (32.1)	3 (33.3)
Changes in the way your body looks,[e] *n* (%)	11 (39.3)	5 (17.9)	1 (20.0)
Weight loss or wasting, *n* (%)	14 (50.0)	8 (28.6)	2 (25.0)
Hair loss or changes in the way your hair looks, *n* (%)	6 (21.4)	2 (7.1)	1 (50.0)

[a]The following scoring system was used:
score 0—"I do not have this symptom"; score 1—"I have this symptom and it doesn't bother me"; score 2—"I have this symptom and it bothers me a little"; score 3—"I have this symptom and it bothers me"; and score 4—"I have this symptom and it bothers me a lot."
[b]Only *"pain, numbness, or tingling in hands or feet"* was rated as bothersome or very bothersome by most patients.
[c]Skin problems, such as rash, dryness, or itching.
[d]Problems with having sex, such as loss of interest or lack of satisfaction.
[e]Changes in the way your body looks, such as fat deposits or weight gain.

Table 3. Post-visit Survey: Patients' Perceptions of Providers' Awareness of their Symptoms

	Control group (n = *36*)	*Intervention group* (n = *27*)[a]
Not at all aware, *n* (%)	0 (0.0)	0 (0.0)
A little bit aware, *n* (%)	1 (2.8)	0 (0.0)
Somewhat aware, *n* (%)	3 (8.3)	1 (3.7)
Quite a bit aware, *n* (%)	5 (13.9)	1 (3.7)
Very much aware, *n* (%)	27 (75.0)[b,c]	25 (92.6)[b,c]

[a]Only 27 of the 28 patients enrolled in the intervention phase completed the post-visit survey.
[b]$p = 0.07$ based on differences between "very much aware" versus all other response categories.
[c]Excluding patients who participated in both phases, 90% of intervention patients' physicians were thought to be "very much aware" vs. 75% of control subjects' physicians ($p = 0.02$).

the electronic tablet on their own, whereas the other half needed assistance. Based on observations made by the cognitive engineer, the most common reasons for needing assistance were inability to read the text because of small font; difficulty handling the stylus because of arthritis or other dexterity impairments; or difficulty understanding particular words (for example, asking what a "provider" was). Those factors were also independently endorsed as particularly important by our study coordinator. Although all patients completed the symptom survey, some patients were concerned about the confidentiality of their information, and others were concerned about its usefulness ("doctors never check the computer anyway"). Additionally, one patient with a superficial skin infection was concerned that the tablet and stylus may serve as a vector for communicable disease.

While this pilot study did not include quantitative end points for provider feasibility, the cognitive engineer qualitatively assessed provider feasibility by reviewing the electronic medical records and meeting with the providers individually at the end of the study period. She noted that most of the physicians acted upon the new information in the computer-generated progress note (e.g., some physicians copied and pasted the information from the symptom index-generated note into their own progress note) and that the workflow was only slightly affected.

Patients seemed to be very comfortable using TEMS to endorse symptoms. Twelve of the 20 symptoms queried were endorsed by most participants (Table 2). However, because TEMS used "bothersome" as the minimum severity threshold for inclusion in the computer-generated progress note, most endorsed symptoms were not included in the notes. Indeed, participants were able to use the tool to discriminate among varying severities of symptoms; only 1 symptom ("pain, numbness, or tingling in hands or feet") was rated as "bothersome" by most participants.

Effectiveness

In the intervention phase, 25 (93%) participants thought that their clinicians were "very aware" of their symptoms, whereas during the control phase, only 27 (75%) participants thought their clinicians were very aware ($p = 0.07$; Table 3).

Additionally, although the proportion of providers' notes listing symptoms was identical in the control and intervention phases (Table 4), there was a trend toward including a greater number of symptoms in intervention phase progress notes (mean [SD] number of symptoms mentioned in the progress note but not in the problem list: 3.6 [3.2] in the intervention phase versus 2.7 [2.3] in control phase, $p = 0.07$; mean [SD] number of symptoms mentioned in the problem list: 1.9 [1.5] in the intervention phase versus 1.6 [1.3] in control phase, $p = 0.22$).

Discussion

We have developed a clinical decision support tool (TEMS) that focuses on symptom management in HIV care. TEMS encompasses a wide breadth of information management in clinical care, from eliciting information through recommending clinical approaches to that information. Its design is particularly noteworthy because it processes and filters elicited information in order to emphasize that which is most clinically relevant, and therefore minimizes additional time burden on clinicians. In that respect, TEMS is innovative because informatics interventions generally do not place a great emphasis on minimizing demands on clinicians' time and attention, and therefore are at risk of inducing "alert fatigue" and subsequent reductions in effectiveness.

Our pilot study suggests that TEMS was accepted by clinicians and did not substantially impede workflow, and therefore it was successful in this initial feasibility test. We have learned that TEMS acceptance by patients could be improved by increasing font size, increasing stylus size, simplifying language, reassuring patients about confidentiality, and cleaning the keyboard/stylus with disinfectant in between uses. Nevertheless, available resources did not allow us to assess many important aspects of its feasibility. Future work

Table 4. Symptoms Mentioned in Progress Notes and Treatment Plans[a]

	Control Phase[b] (n = *28*)	*Intervention Phase*[b] (n = *28*)
Progress notes mentioning symptoms, *n* (%)	22 (78.6)	22 (78.6)
Number of symptoms mentioned:		
Mean (SD)	2.7 (2.3)	3.6 (3.2)
Median	2	3
Treatment plans listing symptoms, *n* (%)	23 (82.1)	23 (82.1)
Number of symptoms listed:		
Mean (SD)	1.6 (1.3)	1.9 (1.5)
Median	2	2

[a]Progress notes were defined exclusive of treatment plans.
[b]Because not all patients in the intervention phase also participated in the control phase, we used their most recent prior visit as a surrogate.
SD, standard deviation.

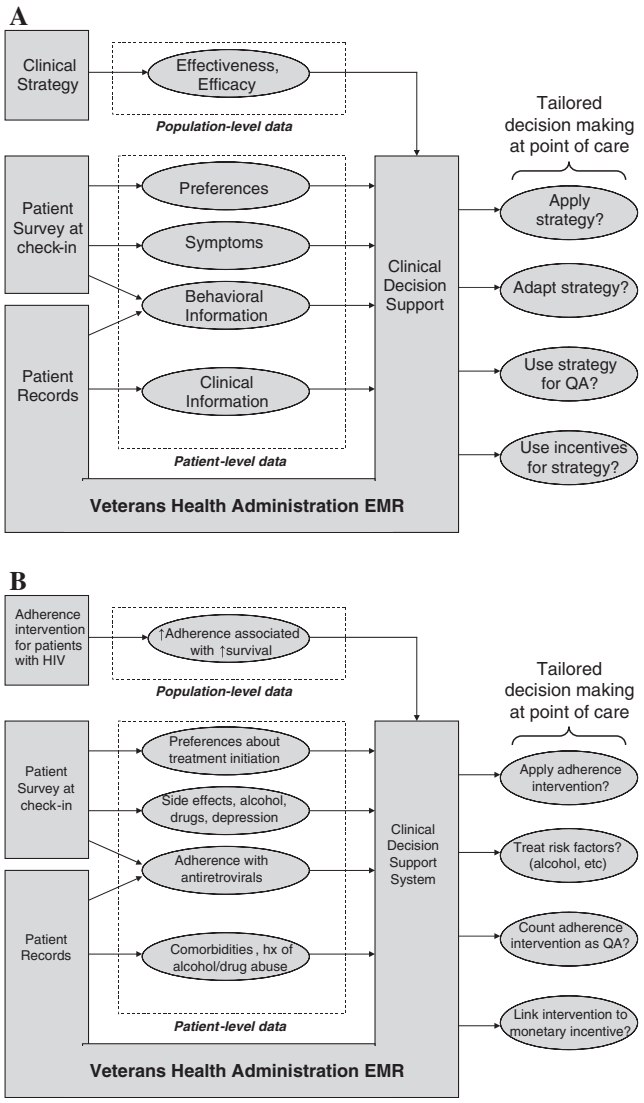


FIG. 2. General framework for using clinical decision support tools within the Veterans Health Administration electronic medical record (EMR) system to individualize care. The clinical decision support tool described in this report is only one example of a nearly infinite variety of ways in which the EMR may be used to adapt clinical strategies to the preferences, symptoms, behaviors, and clinical histories of individual patients (**A**). For example, a more comprehensive version of the clinical decision support tool described in this report could aim to improve medication adherence based on patient-level behavioral risk factors, symptoms, and comorbidities (**B**). EMR, electronic medical record; QA, quality assurance; ↑ increased; Hx, history.

should better evaluate the impact of TEMS on providers by measuring the additional time required for clinicians to view notes and to respond to them, assessing more precisely patient time requirements and staffing requirements generated by TEMS, and surveying providers regarding their satisfaction with the tool. Additionally, the favorable trends that were observed regarding effectiveness (i.e., perception of providers' symptom awareness) need to be confirmed in future studies having greater statistical power.

There are many possible ways to use clinical decision support systems to individualize care in the VHA, as EMR information may be combined with self-reported patient information in a wide range of domains (i.e., preferences, symptoms, behaviors, clinical data), and the resulting information may be used to tailor clinical strategies for individual patients (Fig. 2A). For example, an expanded version of TEMS could be developed with the aim of improving adherence to ART (Fig. 2B) by combining pharmacy refill data with symptom survey data to identify patients with probable adherence difficulties, and to individualize approaches to improve adherence based on relevant behavioral risk factors, for example, depression,[2,23,24] alcohol abuse,[25–28] or other substance abuse.[29–32]

TEMS has important limitations, some of which were not anticipated during its design. Even though it uses portable electronic tablets that have the capacity to transmit information wirelessly, evolving VA security standards have forced us to use hardwired network connections, thereby removing some of the tablet's flexibility. We envisioned a wide network of hyperlinks to help clinicians respond to symptom information, but constructing those links was not possible within the time and budgetary constraints of this pilot project. More generally, clinical decision support tools require regular updates of clinical knowledge and technical support in order to impact care in a sustainable manner.[10–12] Nonetheless, we have accomplished our main objectives and have designed a clinical decision support tool having a structure that can be generalized to other diseases and clinical management questions. Because informatics expertise and EMRs are becoming increasingly sophisticated, we may be entering a promising era for clinical decision support tools that aim to individualize care.

Acknowledgments

This work was funded by Department of Veterans Affairs, Health Services Research and Development grant PCC 05-069-1, Implementing Symptoms Assessment into Clinical HIV Care.

Author Disclosure Statement

No competing financial interests exist.

References

1. Braithwaite RS, Goulet J, Kudel I, Tsevat J, Justice AC. Quantifying the decrement in utility from perceived side effects of combination antiretroviral therapies in patients with HIV. Value Health 2008;11:975–979.
2. Paterson DL, Swindells S, Mohr J, et al. Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. Ann Intern Med 2000;133:21–30.
3. Justice AC, Rabeneck L, Hays RD, Wu AW, Bozzette SA. Sensitivity, specificity, reliability, and clinical validity of provider-reported symptoms: A comparison with self-reported symptoms. Outcomes Committee of the AIDS Clinical Trials Group. J Acquir Immune Defic Syndr 1999;21:126–133.
4. Justice AC, Chang CH, Rabeneck L, Zackin R. Clinical importance of provider-reported HIV symptoms compared with patient-report. Med Care 2001;39:397408.

5. Fontaine A, Larue F, Lassauniere JM. Physicians' recognition of the symptoms experienced by HIV patients: How reliable? J Pain Symptom Manage 1999;18:263–270.
6. Justice AC, Holmes W, Gifford AL, et al. Development and validation of a self-completed HIV symptom index. J Clin Epidemiol 2001;54(Suppl 1):S77–90.
7. Holzemer WL, Hudson A, Kirksey KM, Hamilton MJ, Bakken S. The revised Sign and Symptom Check-List for HIV (SSC-HIVrev). J Assoc Nurses AIDS Care 2001;12:60–70.
8. Holzemer WL, Henry SB, Nokes KM, et al. Validation of the Sign and Symptom Check-List for Persons with HIV Disease (SSC-HIV). J Adv Nurs 1999;30:1041–1049.
9. Shiffman RN, Liaw Y, Brandt CA, Corb GJ. Computer-based guideline implementation systems: A systematic review of functionality and effectiveness. J Am Med Inform Assoc 1999;6:104–114.
10. Sim I, Gorman P, Greenes RA, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc 2001;8:527–534.
11. Bates DW, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: Making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10:523–530.
12. Osheroff JA, Teigh JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc 2007;14:141–145.
13. Guidelines for the use of antiretroviral agents in HIV-1–infected adults and adolescents. October 10, 2006. The Panel on Clinical Practices for the Treatment of HIV Infection convened by the Department of Health and Human Services. www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf. (Last accessed November 2006).
14. Wormser G. AIDS and Other Manifestations of HIV Infection, Fourth edition. San Diego: Elsevier Academic Press; 2004.
15. Sande M, Eliopoulos G, Moellering R, et al. The Sanford Guide to HIV/AIDS Therapy. Antimicrobial Therapy, Inc., 2006.
16. Currently Approved Drugs for HIV: A Comparative Chart. www.aidsmeds.com/lessons/drugchart.htm (Last accessed December 2008).
17. HIV Drug Interactions. www.hiv-druginteractions.org (Last accessed December 2008).
18. HIV InSite. University of California, San Francisco. http://hivinsite.ucsf.edu (Last accessed December 2008).
19. Justice AC, Dombrowski E, Conigliaro J, et al. Veterans Aging Cohort Study (VACS): Overview and description. Med Care 2006;44:S13–S24.
20. Chesney MA, Ickovics JR, Chambers DB, et al. Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: The AACTG adherence instruments. Patient Care Committee & Adherence Working Group of the Outcomes Committee of the Adult AIDS Clinical Trials Group (AACTG). AIDS Care 2000;12:255–266.
21. Detmar SB, Aaronson NK. Quality of life assessment in daily clinical oncology practice: A feasibility study. Eur J Cancer 1998;34:1181–1186.
22. Detmar SB, Muller MJ, Schomagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: A randomized controlled trial. JAMA 2002;288:3027–3034.
23. Tucker JS, Burnam MA, Sherbourne CD, Kung FY, Gifford AL. Substance use and mental health correlates of nonadherence to antiretroviral medications in a sample of patients with human immunodeficiency virus infection. Am J Med 2003;114:573–580.
24. Kleeberger CA, Buechner J, Palella F, et al. Changes in adherence to highly active antiretroviral therapy medications in the Multicenter AIDS Cohort Study. AIDS 2004;18:683–688.
25. Cook RL, Sereika SM, Hunt SC, Woodward WC, Erlen JA, Conigliaro J. Problem drinking and medication adherence among persons with HIV infection. J Gen Intern Med 2001;16:83–88.
26. 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;28:572–577.
27. Braithwaite RS, McGinnis KA, Conigliaro J, et al. The temporal association between alcohol consumption and adherence in a cohort of HIV+ and matched HIV− patients. Alcohol Clin Exp Res 2005;29:1190–1197.
28. Braithwaite RS, Conigliaro J, Roberts MS, et al. Estimating the impact of alcohol consumption on survival for HIV+ individuals. AIDS Care 2007;19:459–466.
29. Lucas GM, Gebo KA, Chaisson RE, Moore RD. Longitudinal assessment of the effects of drug and alcohol abuse on HIV-1 treatment outcomes in an urban clinic. AIDS 2002;16:767–774.
30. 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.
31. Tucker JS, Burnam MA, Sherbourne CD, Kung FY, Gifford AL. Substance use and mental health correlates of nonadherence to antiretroviral medications in a sample of patients with human immunodeficiency virus infection. Am J Med 2003;114:573–580.
32. Kleeberger CA, Buechner J, Palella F, et al. Changes in adherence to highly active antiretroviral therapy medications in the Multicenter AIDS Cohort Study. AIDS 2004;18:683–688.
33. Lucas GM, Mullen A, Weidle PJ, Hader S, McCaul ME, Moore RD. Directly administered antiretroviral therapy in methadone clinics is associated with improved HIV treatment outcomes, compared with outcomes among comparison groups. Clin Infect Dis 2006;42:1628–1635.

Address correspondence to:
Claudia M. Nader
Department of Medicine
Boston University School of Medicine
Crosstown Building- 2nd Floor
801 Massachusetts Ave
Boston, MA 02118

E-mail: clnader@bu.edu

Appendix. Detailed Description of Incorporating Design Factors into TEMS

This appendix discusses in more detail how we incorporated each of the four main design factors into TEMS: elicitation, organization, presentation, and recommendation.

Elicitation

TEMS collects information on a portable electronic "tablet" (Panasonic model CF-08) that enables respondents to answer questions by using a hand-held stylus similar to a pen, or alternatively by touching the screen with a finger. After a patient has his or her vital signs measured, the medical assistant registers the patient in the tablet-based survey application and remains with the patient in order to help with any technical difficulties that may arise as the patient uses the tablet. The tablet queries patients about their symptoms by using the 20-question HIV Symptom Index. Because the HIV Symptom Index does not fit onto 1 screen, patients must proceed through 5 screens in order to answer all questions. The implementation was programmed so that respondents are not able to "skip" a question.

We developed TEMS in conjunction with the clinicians who would implement it, and strove to incorporate their suggestions. Clinicians thought that the usefulness of symptom information would be greatly enhanced if the duration of each symptom could be reported, along with patients' judgment about whether that symptom was due to a medication side effect. For that reason, we augmented the HIV Symptom Index by asking additional questions regarding duration of symptoms ("How long has this symptom bothered you?" with possible responses of "less than 1 week," "between 1 week and 1 month," "between 1 month and 1 year," and "longer than 1 year") and attributability of symptoms to medications using the 1-question item adapted from the AACTG questionnaire ("Do you think that this symptom is caused by drugs that you take to treat your HIV infection?" with possible responses of "yes," "unsure," and "no"). Also, in response to ideas from providers, we ensured that each patient's vital signs would be included with the symptom information. To avoid errors that might result from manual entry, TEMS extracts vital sign information automatically from the electronic medical record (EMR).

We had hoped to use the wireless capacities of the tablet to transmit symptom information into the EMR in order to minimize staff burden (i.e., they would have the flexibility to elicit symptom information at a variety of places and times). However, shortly before beta-testing TEMS, the VHA issued an embargo of wireless data-encoding algorithms. Therefore, we beta-tested TEMS by using a wired connection to transmit the data, and will try wireless transmission pending the approval of a VHA-compliant data-coding algorithm.

Organization

We maximized the specificity of information to be presented to the provider by using two distinct information filters, each of which reduces the level of information detail when that information is unlikely to lead to an effective clinical response. The first filter excludes symptom information when a symptom is sufficiently minor so that it is unlikely to warrant a clinical response (i.e., if the symptom is not bothersome or very bothersome). The second filter reduces the detail of symptom information when a patient endorses multiple unrelated symptoms, because clinical approaches directed at the various symptoms themselves may be less effective than clinical approaches geared to a latent underlying condition (e.g., depression). Thus, we defined a specifiable "threshold" for the number of symptoms: if a patient endorsed more than the threshold number of symptoms, then the note informed the provider about which symptoms were endorsed but did not include duration and attributability; if a patient endorsed the threshold number of symptoms or fewer, then the note included the full detail of symptom duration and attributability. We have currently set the threshold at three symptoms, but we have not yet determined the optimal threshold value.

Presentation

TEMS generates progress notes by using one of two templates (Fig. 1). If a patient endorses a number of symptoms that falls below the numerical threshold, then the note gives full detail about each symptom. Appendix Figure 1A shows the note that was generated for a patient who endorsed two symptoms (headache, which has lasted less than 1 week, with the patient being unsure of its relationship to medications; and loss of appetite, which has lasted for 1 week to 1 month, attributed by the patient to his HIV medications). However, if a patient endorses a number of symptoms that exceeds the specified threshold, then the note limits the level of detail. Appendix Figure 1B shows the note that was generated when a patient endorsed 13 separate symptoms. Although the patient reported severity and symptom attribution, that information is not represented in the note. Both note formats contain the result of the one-item adherence query and the vital signs.

In our original concept of the tool, we had planned for physicians to see a "pop-up" alert on their computer screen when the computer-generated progress note was created (we thought that this would be desirable because many physicians review the patient record before patients enter the examination room). However, the pop-up presented an unforeseen programming burden because of technical features of the VHA's EMR architecture, and therefore we had to substitute a low-tech approach in which patients carried a bright yellow "alert" card into their examination room, serving as a prompt for the physician to look at the results of the symptom index.

Recommendation

Each symptom in each progress note is linked to a file that recommends a corresponding clinical response strategy (Fig. 2), either diagnostic considerations or symptom treatments. Because diagnostic considerations are often distinct for individuals with HIV, TEMS emphasizes those considerations that are particularly applicable to individuals with HIV infection. Additionally, because HIV-infected individuals who are severely immunosuppressed have particular diagnostic considerations (i.e., opportunistic infections, etc.), TEMS stratifies strategies by CD4 count when indicated ($<$200 cells per milliliter versus $\geq$200 cells per milliliter). When diagnoses would be facilitated by using another screening instrument (e.g., the Alcohol Use Disorders Identification Test to screen for hazardous alcohol consumption), that instrument along with its scoring algorithm is included in the response strategy.

To mitigate barriers to provider adherence, each recommendation is expressed by using the terminology and care options particular to our institution (e.g., rather than recommending "refer for alcohol treatment," TEMS will recommend "Substance Abuse Clinic consult"), and we designed our therapeutic strategies to include only generic drugs that were on formulary. The full text of all response strategies is available at www.vacohort.org.

We had intended to link each symptom in the computer-generated progress note to its response strategy by means of a hyper-link, but technical barriers discouraged such an approach. As an alternative, we placed a Uniform Resource Locator (URL) next to each symptom so that the URL could be copied and pasted into a Microsoft Word document and "pointed" to a document that contains the corresponding response strategy. Similarly, we intended to embed in each response strategy hyper-links to relevant orders, but that also proved to be prohibitively burdensome. As an alternative, we carefully edited each response strategy to ensure concordance between an order's specification in our response strategy and its representation in the EMR.

Driving After Binge Drinking

Timothy S. Naimi, MD, MPH, David E. Nelson, MD, MPH, Robert D. Brewer, MD, MSPH

Background: Although binge drinking is strongly associated with alcohol-impaired driving, little is known about the prevalence of or risk factors for driving after binge drinking.

Purpose: The purpose of this study was to assess the prevalence of, and risk factors for, driving during or shortly after a specific binge drinking episode.

Methods: The data were analyzed in 2007 and 2008 from 14,085 adults from 13 states in 2003 and 14 states in 2004 who reported binge drinking and answered an additional series of questions about binge drinking behaviors as part of the Behavioral Risk Factor Surveillance System survey. Binge drinking was defined as the consumption of five or more drinks during a drinking occasion.

Results: Overall, 11.9% of binge drinkers drove during or within 2 hours of their most recent binge drinking episode. Those drinking in licensed establishments (bars, clubs, and restaurants) accounted for 54.3% of these driving episodes. Significant independent risk factors for driving after binge drinking included male gender (AOR=1.75); being aged 35–54 or ≥55 years compared to 18–34 years (AOR=1.58 and 2.37, respectively); and drinking in bars or clubs compared to drinking in the respondent's home (AOR=7.81). Drivers who drank most of their alcohol in licensed establishments consumed an average of 8.1 drinks, and 25.7% of them consumed ≥10 drinks.

Conclusions: Because binge drinking and subsequent driving were common in establishments licensed to sell alcohol, and because licensing is conditional on responsible beverage service practices (i.e., not selling to intoxicated people), efforts to prevent impaired driving should focus on enforcing responsible beverage service in licensed establishments.

(Am J Prev Med 2009;37(4):314–320) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Excessive drinking is the third leading actual cause of death in the U.S.,[1] is responsible for approximately 79,000 deaths annually, and shortens the lives of those who die by approximately 30 years.[2,3] Binge drinking, defined as the consumption of five or more drinks on an occasion, is responsible for more than half of these deaths[3] and contributed to the 13,000 deaths from alcohol-related motor vehicle crashes in 2006.[4] In the U.S., approximately 15% of all adults report one or more episodes of binge drinking in the past month, resulting in 1.5 billion binge drinking episodes annually (or approximately seven episodes per adult per year).[5] Survey research has shown that binge drinking is strongly associated with alcohol-impaired driving. For example, a recently published study found that 88% of self-reported episodes of alcohol-impaired driving involved adults who reported past-month binge drinking.[6] Another study found that 12% of binge drinkers report that they drove after having "perhaps too much to drink" in the past month, although the amount of alcohol consumed was not quantified.[7] In addition, the increase in binge drinking episodes observed among U.S. adults from 1993 to 2001[5] paralleled a similar increase in alcohol-impaired driving episodes during this time period.[7]

However, little is known about the likelihood of, or risk factors for, driving after a specific binge drinking episode. Although most impaired-driving countermeasures focus on reducing driving among individuals who are already impaired, little attention has been focused on the role played by demographic or environmental factors that may be risk factors for this impairment and/or propensity to drive. It was hypothesized that establishments licensed to sell alcohol were the site of a large proportion of binge drinking and subsequent driving episodes, but it was not clear whether that association was mostly accounted for by the characteristics of those consuming alcohol in licensed establishments. Therefore, the purpose of this study was to assess the prevalence of, and risk factors for, driving during or shortly after a specific binge drinking episode.

From the Alcohol Team, Emerging Investigations and Analytic Methods Branch, Division of Adult and Community Health, National Centers for Chronic Disease Prevention and Health Promotion, CDC, Atlanta, Georgia

Address correspondence and reprint requests to: Timothy S. Naimi, MD, MPH, c/o ZPHS Hospital, P.O. Box 467, Zuni NM 87327. E-mail: tnaimi@post.harvard.edu.

0749-3797/09/$–see front matter
doi:10.1016/j.amepre.2009.06.013

Methods

Data for this study came from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) survey, a random-digit telephone survey of U.S. adults aged ≥18 years in all 50 states, the District of Columbia, and the territories of Guam, Puerto Rico, and the U.S. Virgin Islands. The survey includes questions on a variety of health risk behaviors, including alcohol consumption. Details about the BRFSS are available at http://www.cdc.gov/brfss/technical_infodata/index.htm.

A binge drinker was defined as someone who consumed alcohol in the past 30 days and 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 5 or more drinks on an occasion?* In 2003 and 2004, the BRFSS survey offered states a new, optional module of six additional questions to be asked of those who reported binge drinking; all questions were about a respondent's most recent binge drinking episode. Driving after binge drinking was assessed by the question: *Did you drive a motor vehicle, such as a car, truck, or motorcycle during or within a couple of hours after this occasion?* A yes answer to this question was the numerator for determining the prevalence of driving after binge drinking. Because each respondent was providing information about a single binge drinking event, prevalence information was combined with the number of binge drinkers in particular strata to determine the number of episodes of driving after binge drinking. Other questions in the module elicited information about the number and type of alcohol-containing beverages (beer, wine, or liquor) consumed during their most recent binge drinking episode and the physical location where most binge drinks were consumed.

Analyses were limited to the 18 states that used this set of binge drinking questions in both 2003 and 2004 (nine states); 2003 only (four states); or 2004 only (five states). States using the module in both years were California, Maine, Michigan, Minnesota, Montana, Nevada, New Hampshire, Wisconsin, and Wyoming; states using it 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. 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 among respondents in the states and years included in the study was 16.3%, which is approximately 1% higher than that for the U.S. during 2003–2004. The weighted prevalence of driving after binge drinking was very similar in 2003 (11.9%) and 2004 (11.6%).

The study was restricted to those who reported one or more episodes of binge drinking in the past 30 days. Data were weighted by age, gender, and race or ethnicity to be representative of the adult population for each state and year analyzed; weights were divided by 2 for states with 2 years of data. After excluding binge drinkers with missing or incomplete information from the binge drinking module, data from 14,085 respondents were analyzed, including 1848 respondents who reported driving during or within 2 hours of binge drinking.

All data analyses were conducted using SAS, version 9.0, and SUDAAN, version 9.0. Analyses were conducted for three types of variables: demographics, alcohol-specific measures, and binge drinking location. Demographic characteristics included age group (18–24, 25–34, 35–44, 45–54, ≥55 years), which was collapsed into three groups (18–34, 35–54, and ≥55 years) for regression analysis based on strata size and similarity with respect to driving characteristics; gender; race or ethnicity (white, non-Hispanic; black, non-Hispanic; other, non-Hispanic; and Hispanic), which was collapsed to white non-Hispanic versus other for regression analysis; education level (less than high school, high school graduate, some college, and college graduate), which was collapsed to some college or more versus high school graduate or less for regression analysis; income level (<$25,000, $25,000–$49,999, and ≥$50,000), which was collapsed to ≥$50,000 versus ≤$49,000 for regression analysis; marital status (married, never married, unmarried couple, divorced, or separated), which was collapsed to married versus nonmarried for regression analysis; and employment status (employed, unemployed, student, homemaker, or retired), which was collapsed to employed versus nonemployed for regression analysis. Alcohol stratification variables included the number of binge episodes during the past 30 days (one to two, three to four, and five or more), which was collapsed to one to two versus three or more for regression analysis; and the total number of drinks consumed during the last binge episode (five to six, seven to nine, and ten or more), which was collapsed to five to six versus seven or more for regression analysis. The physical location where most of the drinks were consumed (own home, another person's home, bar or club, restaurant, other public place, or other) was also analyzed.

Results

Overall, 75.1% of binge drinkers were men, 75.1% were aged ≤44 years, 49.7% consumed seven or more drinks during their most recent binge drinking episode, and 58.3% drank in a private residence (their home or someone else's home; Table 1).

After weighting, 11.9% of binge drinkers reported driving during or within 2 hours of their most recent binge drinking episode (Table 1). Men were more likely than women to drive after binge drinking (13.2% vs 8.1%), and men accounted for 82.9% of all recent binge drinking and driving episodes. The prevalence of driving after binge drinking increased slightly with age; 50.6% of binge drinking and driving episodes involved those aged ≥35 years. Among underage adults aged 18–20 years, 10.3% drove after binge drinking and they accounted for 6.3% of driving episodes (data not shown in Table 1). More than 90% of most recent binge drinking and driving episodes were accounted for by high school graduates or those with at least some college education, and 78.6% of driving episodes were reported by binge drinkers who were employed. Although those who reported five or more binge drinking episodes in the past 30 days were more likely to drive after their most recent binge episode than those who reported binge drinking once or twice, approximately half (48.7%) of driving episodes involved those who reported binge drinking only once or twice in the past

Table 1. Number and weighted percentage of binge drinkers, prevalence of driving after binge drinking[a] among binge drinkers, and proportion of driving episodes among those driving after binge drinking, by selected characteristics

Characteristic	No. of binge drinkers (weighted %)[c]	Percentage of binge drinkers who drove after binge drinking[a,b] (n=14085)	Proportion of driving episodes among binge drinkers[a,b] (n=1848)
All	14,085 (100.0)	11.9 (10.9, 13.0)	100.0
Gender			
Male	9,611 (75.1)	13.2 (11.9, 14.5)	82.9 (79.9, 85.6)
Female	4,474 (24.9)	8.1 (6.9, 9.6)	17.1 (14.4, 20.1)
Age (years)			
18–24	2,076 (24.3)	10.9 (8.6, 13.8)	22.2 (17.9, 27.3)
25–34	3,617 (28.5)	11.3 (9.6, 13.3)	27.1 (23.3, 31.4)
35–44	3,653 (22.3)	11.2 (9.6, 13.1)	21.1 (18.0, 24.5)
45–54	2,841 (15.1)	13.9 (11.4, 16.9)	17.7 (14.5, 21.4)
≥55	1,898 (9.7)	14.5 (11.9, 17.6)	11.8 (9.7, 14.4)
Race			
White, non-Hispanic	12,064 (73.2)	12.3 (11.3, 13.3)	75.4 (69.8, 80.2)
Black, non-Hispanic	373 (4.1)	14.3 (9.6, 20.7)	4.9 (3.3, 7.3)
Other, non-Hispanic	701 (5.4)	16.1 (10.2, 24.4)	7.2 (4.5, 11.4)
Hispanic	877 (17.3)	8.6 (5.8, 12.7)	12.5 (8.5, 17.9)
Education			
<High school	903 (9.5)	7.5 (6.4, 13.5)	7.5 (5.2, 10.8)
High school grad	4,455 (31.0)	12.2 (10.4, 14.3)	31.9 (27.7, 36.4)
Some college	4,318 (30.4)	12.2 (10.3, 14.3)	31.2 (26.9, 35.8)
College grad	4,400 (29.0)	12.1 (10.4, 13.9)	29.4 (25.6, 33.5)
Income ($)			
<25K	2,958 (23.7)	10.3 (7.9, 13.3)	20.0 (15.7, 25.1)
25K–50K	4,490 (30.3)	14.2 (12.3, 16.3)	35.2 (30.9, 39.8)
>50K	5,761 (46.0)	11.9 (10.5, 13.4)	44.8 (40.2, 49.5)
Marital status			
Married	6,801 (47.5)	10.3 (9.0, 11.7)	41.1 (36.7, 45.6)
Previously married[d]	2,791 (12.2)	17.0 (14.4, 19.9)	17.3 (14.6, 20.4)
Unmarried couple	795 (7.9)	12.9 (8.2, 19.5)	8.5 (5.4, 13.2)
Never married, single	3,683 (32.5)	12.1 (10.4, 14.1)	33.1 (28.9, 37.6)
Employment			
Employed	11,198 (76.1)	12.3 (11.2, 13.5)	78.6 (73.8, 82.7)
Unemployed[e]	1,027 (9.3)	9.6 (6.8, 13.5)	7.6 (5.3, 10.6)
Homemaker	301 (2.1)	7.2 (3.7, 13.5)	1.2 (0.6, 2.4)
Student	695 (8.2)	11.5 (7.0, 18.4)	7.9 (4.7, 12.8)
Retired	853 (4.3)	13.1 (9.9, 17.1)	4.7 (3.5, 6.3)
No. of binge episodes, past 30 days			
1–2	8,248 (56.5)	10.3 (9.1, 11.6)	48.7 (44.1, 53.4)
3–4	2,632 (18.8)	13.2 (10.7, 16.3)	20.9 (17.0, 25.3)
≥5	3,205 (24.7)	14.6 (12.5, 17.0)	30.4 (26.4, 34.7)
No. of drinks, most recent binge episode			
5–6	7,730 (50.3)	10.8 (9.6, 12.1)	45.7 (41.2, 50.3)
7–9	3,512 (25.8)	12.7 (10.7, 14.9)	27.4 (23.5, 31.7)
≥10	2,843 (23.9)	13.4 (10.9, 16.3)	26.9 (22.4, 31.8)
Location, most recent binge episode			
Home	5,264 (38.3)	4.0 (3.2, 5.0)	12.8 (10.3, 15.9)
Another's home	2,327 (20.0)	13.6 (11.3, 16.3)	22.9 (19.1, 27.1)
Restaurant	1,001 (7.4)	16.3 (12.1, 21.6)	10.1 (7.4, 13.7)
Bar/club	4,090 (25.3)	20.8 (18.3, 23.6)	44.2 (39.6, 48.8)
All other[f]	1,380 (9.0)	13.2 (10.2, 17.0)	10.0 (7.7, 12.9)

[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. Driving after binge drinking refers to those who reported driving during or within 2 hours of their most recent binge drinking episode.

[b]Results were weighted to be representative of states and years included in this study; BRFSS data are weighted by gender, age, and race or ethnicity.

[c]The sum of strata for selected variables may not be 14,062 or 100% because of nonresponse to that variable or rounding error, in the case of the weighted percentages. BRFSS data are weighted by gender, age, and race or ethnicity.

[d]Previously married included those who were divorced, separated, or widowed.

[e]Unemployed included those who were unemployed for less than 1 year, more than 1 year, or who reported they were unable to work.

[f]*Other* refers to places such as parks, sporting events, concerts, or other locations.

BRFSS, Behavioral Risk Factor Surveillance System

30 days. Similarly, although those who consumed more drinks per binge (seven or more drinks) were nonsignificantly more likely to report driving after their most recent binge episode, almost half of driving episodes involved those who consumed five to six drinks. The prevalence of driving after binge drinking varied considerably based on drinking location, ranging from 4.0% for those who reported binge drinking at home to 20.8% of those who drank in bars or clubs. Those who reported binge drinking in bars or clubs accounted for 44.2% of driving episodes; licensed establishments (bars, clubs, and restaurants) accounted for more than half (54.3%) of all driving episodes.

Binge drinkers who drank most of their alcohol in licensed establishments (bars, clubs, and restaurants) and who subsequently drove consumed an average of 8.1 (95% CI=7.45, 8.65) drinks; 53.5% consumed seven or more drinks; and 25.7% consumed ten or more drinks. Among binge drinkers who drove, a greater proportion of those drinking in bars or clubs consumed ten or more drinks compared with those drinking in restaurants (28.7% vs 13.6%, data not shown).

In stratified analysis, differences observed in the prevalence of driving based on the location of binge drinking were generally consistent across various demographic characteristics, the frequency of binge drinking, and the number of drinks consumed during the binge drinking episode (Table 2). Specifically, driving after binge drinking at home was reported by ≤6% of all subgroups. In contrast, driving after binge drinking in bars or clubs was reported by more than 20% of those in most subgroups. The prevalence of driving after binge drinking in bars increased significantly with age, ranging from 17.0% among those aged 18–34 years to 36.7% among those aged ≥55 years. However, the prevalence of driving after binge drinking in bars or clubs did not differ substantially when evaluated by

Table 2. Prevalence of driving after binge drinking,[a] by location of binge drinking episode and selected characteristics

	Location of binge drinking episode (% [CIs])				
	Home	**Other's home**	**Restaurant**	**Bar/club**	**Other locations[b]**
Overall	4.0 (3.2, 5.0)	13.6 (11.3, 16.3)	16.3 (12.1, 21.6)	20.8 (18.3, 23.6)	13.2 (10.2, 17.0)
Gender					
Male	4.0 (3.1, 5.1)	15.2 (12.3, 18.7)	19.7 (14.1, 26.8)	24.5 (21.1, 28.2)	15.3 (11.6, 20.0)
Female	4.2 (2.5, 6.8)	8.5 (6.0, 12.1)	7.9 (4.4, 13.7)	12.5 (9.9, 15.7)	6.5 (3.2, 12.7)
Age (years)					
18–34	4.0 (2.6, 5.9)	11.8 (9.0, 15.2)	11.7 (7.5, 17.9)	17.0 (13.7, 20.9)	12.3 (8.1, 18.4)
35–54	3.5 (2.5, 4.7)	17.5 (13.1, 22.9)	—[e]	26.5 (22.9, 30.5)	12.9 (8.7, 18.9)
≥55	6.0 (3.8, 9.5)	13.8 (8.1, 22.6)	—[e]	36.7 (28.2, 46.1)	—[e]
Race					
White, non-Hispanic	4.1 (3.2, 5.2)	14.1 (11.7, 17.0)	16.8 (12.8, 21.6)	20.9 (18.7, 23.4)	12.3 (9.4, 16.1)
Other race or ethnicity	3.9 (2.3, 6.3)	12.1 (7.3, 19.3)	—[e]	—[e]	—[e]
Education					
≤High school	4.0 (3.2, 5.0)	14.8 (11.0, 19.6)	15.7 (9.6, 24.6)	21.5 (16.9, 26.8)	13.9 (9.4, 20.1)
>High school	4.0 (3.0, 5.4)	12.6 (9.9, 16.0)	16.6 (11.3, 23.6)	20.5 (17.6, 23.7)	13.0 (9.2, 18.0)
Income ($)					
≤50K	3.9 (2.9, 5.4)	12.4 (9.2, 16.5)	—[e]	22.6 (18.9, 26.8)	14.7 (10.2, 20.6)
>50K	4.1 (2.8, 5.7)	15.7 (12.0, 20.2)	15.7 (11.0, 21.9)	19.7 (16.5, 23.3)	13.3 (9.0, 19.7)
Marital status					
Married	3.0 (2.2, 4.2)	12.2 (8.9, 16.4)	15.3 (9.9., 22.9)	23.5 (19.7, 27.7)	11.4 (8.0, 16.1)
Not married[c]	5.2 (3.9, 7.1)	14.6 (11.5, 18.3)	17.8 (12.0, 25.7)	19.5 (16.3, 23.1)	15.7 (10.8, 22.3)
Employment					
Employed	4.2 (3.2, 5.5)	14.2 (11.5, 17.4)	16.5 (11.7, 22.7)	21.8 (19.2, 24.5)	12.6 (9.4, 16.6)
Not employed[d]	3.3 (2.2, 5.0)	11.9 (7.7, 17.8)	—[e]	17.8 (12.0, 25.7)	—[e]
No. of binge episodes, past 30 days					
1–2	2.8 (1.9, 4.2)	11.6 (8.7, 15.3)	13.4 (9.6, 18.3)	17.0 (14.4, 20.0)	13.8 (9.7, 19.2)
≥3	5.2 (4.0, 6.8)	16.8 (13.2, 21.0)	—[e]	25.4 (21.1, 30.3)	12.4 (8.6, 17.6)
No. of drinks consumed, binge episode					
5–6	3.3 (2.3, 4.6)	12.4 (9.5, 16.0)	13.8 (9.9, 19.0)	19.9 (16.9, 23.3)	15.1 (10.5, 21.2)
≥7	4.6 (3.1, 6.8)	15.8 (11.0, 22.3)	—[e]	22.2(16.0, 29.9)	12.3 (7.1, 20.4)

[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. Driving after binge drinking refers to those who reported driving during or within 2 hours of their most recent binge drinking episode.
[b]Other locations is a combination of two response categories: (1) other public place, which referred to parks, sporting events, or concerts and (2) other location, which was a response option for those who did not answer "don't know/not sure" to drinking location but who did not consume most of their alcohol in their or another person's home, a restaurant, a bar, a club, or a public place.
[c]Not-married people included those who were never married or were single, in an unmarried couple, widowed, divorced, or separated.
[d]Not-employed people included those who were unemployed, were unable to work, or were retired, homemakers, or students.
[e]Point estimates and CIs not reported because of CIs spanning 20% or sample sizes <50.

various strata of education, income, marital status, and employment.

In multivariate logistic regression analysis, significantly increased odds of driving after binge drinking were observed for men compared with women, those aged ≥35 years compared to younger adults, unmarried people compared to married people, those reporting three or more compared to one or two binge drinking episodes in the past 30 days, and those drinking outside of their own home (Table 3). Compared to those binge drinking in their own home, the AOR for driving after binge drinking in bars or clubs was 7.81 and 5.90 for drinking in restaurants. The AOR of driving for people binge drinking in any licensed establishment (bars or clubs and restaurants) was also significantly increased (AOR 3.4, 95% CI=2.7, 4.4) compared with driving after drinking in any private residence (a respondents' home or someone else's home). Race/ethnicity, education, income, and the number of binge drinks consumed were not significantly associated with subsequent driving.

Table 3. Prevalence and AORs[a] for driving after binge drinking[b] among binge drinkers

Characteristic	Percentage who drove after binge drinking[b]	AOR[a] (95% CI)
Gender		
Female	8.1	1.00 (ref)
Male	13.2	1.75 (1.37, 2.33)
Age		
18–34	11.2	1.00 (ref)
35–54	11.9	1.58 (1.25, 2.01)
≥55	14.3	2.37 (1.69, 3.34)
Race or ethnicity		
Other than white, non-Hispanic	10.8	1.00 (ref)
White, non-Hispanic	12.2	0.98 (0.72, 1.34)
Education level		
Greater than high school	12.1	1.00 (ref)
High school or less	11.2	1.01 (0.80, 1.27)
Income level ($)		
≥50K	11.9	1.00 (ref)
<50K	12.2	1.05 (0.84, 1.32)
Marital status		
Married	10.2	1.00 (ref)
Not married[c]	13.2	1.32 (1.04, 1.68)
Employment status		
Not employed[d]	11.1	1.00 (ref)
Employed	12.0	1.25 (0.91, 1.71)
No. of binge episodes, past 30 days		
1–2	10.3	1.00 (ref)
≥3	13.7	1.52 (1.21, 1.89)
No. of drinks consumed, most recent binge episode		
5–6	10.6	1.00 (ref)
≥7	12.9	1.06 (0.84, 1.32)
Drinking location, most recent binge episode		
Own home	4.0	1.00 (ref)
At another person's home	13.6	4.61 (3.25, 6.53)
Restaurant	16.3	5.90 (3.77, 9.22)
Bar or club	20.8	7.81 (5.69, 10.73)
Other location[e]	13.2	4.31 (2.88, 6.45)

[a]AORs for driving after a respondent's most recent binge drinking episode were determined by logistic regression and were adjusted for the covariates listed in this table.
[b]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. Driving after binge drinking refers to those who reported driving during or within 2 hours of their most recent binge drinking episode.
[c]Unmarried people included those who were never married/single, in an unmarried couple, widowed, divorced, or separated.
[d]Not employed people included those who were unemployed, unable to work, retired, homemakers, or students.
[e]Other location refers to places such as parks, sporting events, concerts, or other locations.

Discussion

To our knowledge, this is the first U.S. study to examine the likelihood of driving following a specific binge drinking event, and the first to assess personal and contextual risk factors (e.g., location of alcohol consumption) affecting these associations. Overall, almost one in eight binge drinkers drove during or within 2 hours of their most recent binge drinking episode. Of these people, more than half consumed most of their alcohol in establishments licensed to sell alcohol. Bars and clubs accounted for 43% of binge drinking and driving episodes; 25% of those who drove after binge drinking in any establishment licensed to sell alcohol (bars, clubs, and restaurants) consumed ten or more drinks. These findings emphasize the need to implement effective measures to reduce binge drinking, including the implementation of policies to prevent overservice in licensed establishments where selling alcohol to intoxicated people is generally illegal.[8]

Although a strength of this study was that it established a temporal relationship between binge drinking and subsequent driving, the current findings are consistent with the strong cross-sectional relationship between binge drinking and impaired driving that has been described in previous studies.[6,7] A population-based study of California drinkers observed that driving while intoxicated was influenced by age, gender, individual drinking patterns, increased alcohol outlet density, and drinking in bars and restaurants.[9] The importance of drinking location and subsequent impaired driving was further illustrated by a New Mexico study of 5000 people convicted of driving-while-intoxicated (DWI), in which 45% of those convicted were drinking in bars or lounges prior to their arrest.[10] Another study of 16,000 DWI arrestees in Ventura County CA found that 44% had consumed their last drink in a bar, club, or restaurant and that those drinking in these establishments had significantly higher blood

alcohol concentrations than those drinking in other locations.[11]

Studies from other developed countries (e.g., Canada and Australia) have also found that drinking in bars and night clubs is strongly associated with drinking more than usual and drinking to the point of intoxication.[12–18] Further, a disproportionate number of drinkers who experienced alcohol-related harms or engaged in closely related risk behaviors, including driving while impaired, had been drinking in establishments licensed to sell alcohol.[19,20]

This study reaffirms the predominant role played by men in alcohol-impaired driving,[6] as men were more likely than women to binge drink and were also more likely to drive after doing so. While those in relatively younger age groups accounted for most binge drinking and driving episodes, the AOR of driving after a single binge drinking episode was progressively higher among those in older age groups. This finding may represent a cohort effect, as younger people have grown up in an era of enhanced awareness of, and social stigma associated with, impaired driving. It is also possible that a higher proportion of older binge drinkers were alcohol dependent, but alcohol dependence could not be assessed using this data source. Finally, drinking more drinks was not associated with a higher likelihood of driving. It is possible that a progressive loss of judgment after consuming more drinks was counteracted by a greater awareness of impairment, because the sensation of impairment is generally attained at or above the number of drinks used to define binge drinking.[21] It is also possible that those drinking more drinks were less likely to report subsequent driving.

This study is subject to several limitations. First, data were from self-report, and survey respondents may under-report how much they drink and whether they engage in impaired driving. Therefore, the number of drinks reported by respondents was likely conservative, as was the proportion of those who reported driving after binge drinking. Similarly, the median response rate among states included in this study was 55%, and nonrespondents may be more likely to binge drink and engage in impaired driving, although it is unclear how the proportion of binge drinkers who drive would vary between respondents and nonrespondents. Second, this study assessed the location where most drinks were consumed, and some binge drinkers may have consumed alcohol at more than one location. Third, the study did not include information about all U.S. states, and therefore may not be representative of the U.S. as a whole. And fourth, the BRFSS used a five-drink threshold to define binge drinking among women; the National Institute of Alcohol Abuse and Alcoholism recommends using a four-drink threshold, and BRFSS survey adopted this threshold in 2006. It is unknown what proportion of women drinking exactly four drinks subsequently drove a motor vehicle, but it is likely that the women would have accounted for a somewhat larger proportion of binge drinking and driving episodes had the four-drink threshold been used.

It was not possible to determine whether all people who reported driving during or within 2 hours of a binge drinking episode were legally intoxicated. However, consuming five drinks for men or four drinks for women at a typical rate (i.e., within 2 hours) results in blood alcohol concentrations of 0.08 mg/dL, the legal limit for defining alcohol-impaired driving in all states in the U.S.[22] Further, half of drivers in this study consumed seven or more drinks during their most recent binge episode, suggesting that many of these people may have had blood alcohol concentrations well in excess of 0.08 mg/dL. And finally, it should be noted that the risk of a motor vehicle crash increases at blood alcohol concentrations in excess of 0.03 mg/dL.[23]

Given the frequency with which binge drinkers subsequently drive a motor vehicle, population-based strategies to reduce both binge drinking and impaired driving are required to reduce alcohol-related motor vehicle crashes. Such strategies are the cornerstone of prevention, because there are few targeted interventions available to address the likelihood of excessive drinking or subsequent driving among high-risk demographic groups (e.g., men). Effective population-based strategies to reduce binge drinking include increasing alcohol excise taxes, limiting alcohol outlet density and hours of sale, enhanced enforcement of the age-21 minimum legal drinking age, and limiting days of alcohol sales.[24-26] Effective strategies to prevent alcohol-impaired driving include implementation of the age-21 minimum legal drinking age, 0.08 laws, sobriety checkpoints, lower blood alcohol concentration laws for young and inexperienced drivers, immediate driver's license revocation for those arrested for driving while intoxicated, sobriety checkpoints, server training programs, mass media campaigns intended to reduce impaired driving, ignition interlocks, multi-component impaired-driving interventions with community mobilization, and school-based instructional programs.[27]

Because driving after binge drinking in licensed establishments accounted for more than half of such episodes, implementing and strengthening existing interventions to prevent on-premise binge drinking in retail alcohol outlets are warranted.[28] Effective interventions to improve responsible beverage service include limits on drink discounting, "dram shop" liability laws, mandatory server training programs, and enhanced enforcement of laws prohibiting sales to intoxicated patrons.[27,29,30] However, a number of states lack liability laws or mandatory server training laws for establishments serving alcohol; most states lack adequate numbers of alcoholic beverage control officers; and laws preventing sales to minors or intoxicated people are enforced only sporadically.[31] Further, some states have laws that prevent cities or counties from adopting

more stringent alcohol-control policies than those that exist at the state level. Strengthening these laws and ensuring their enforcement could help reduce alcohol overservice and create an environment that supports responsible beverage service by not placing law-abiding retailers at an economic disadvantage.

The authors wish to acknowledge the contribution of Dr. Ernest Sullivent, who did preliminary analyses related to this manuscript.

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 United States, 2000. JAMA 2004;291:1238–45.
2. CDC. Alcohol-related disease impact (ARDI). www.cdc.gov/alcohol/ardi.htm.
3. Midanik LT, Chaloupka FJ, Saitz R, et al. Alcohol-attributable deaths and years of potential life lost due to excessive alcohol use in the U.S. MMWR Morb Mortal Wkly Rep 2004;53:866–70.
4. National Highway Traffic Safety Administration. Alcohol-impaired driving. www-nrd.nhtsa.dot.gov/Pubs/810801.PDF.
5. Naimi TS, Brewer RD, Mokdad AH, et al. Binge drinking among U.S. adults. JAMA 2003;289:70–5.
6. Flowers N, Naimi TS, Brewer RD, et al. Patterns of alcohol consumption and alcohol-impaired driving in the United States. Alcohol Clin Exp Res 2008;32:639–44.
7. Quinlan KP, Brewer RD, Siegel P, et al. Alcohol-impaired driving among U.S. adults, 1993–2002. Am J Prev Med 2005;28:346–50.
8. 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.
9. Gruenewald PJ, Johnson FW, Treno AJ. Outlets, drinking and driving: a multilevel analysis of availability. J Stud Alcohol 2002;63:460–8.
10. Chang I, Lapham SC, Barton KJ. Drinking environment and sociodemographic factors among DWI offenders. J Stud Alcohol 1996;57:659–69.
11. Alcohol and Drug Prevention Division. Ventura County Place of Last Drink Survey. Ventura CA: Ventura County Behavioral Health, 2007.
12. Kairouz S, Greenfield TK. A comparative multi-level analysis of contextual drinking in American and Canadian adults. Addiction 2007;102:71–80.
13. Greenfield TK, Room R. Situational norms for drinking and drunkenness: trends in the U.S. adult population, 1979–1990. Addiction 1997;92:33–47.
14. Mayer RR, Forster JL, Murray DM, Wagenaar AC. Social settings and situations of underage drinking. J Stud Alcohol 1998;59:207–15.
15. Single E, Wortley S. Drinking in various settings as it relates to demographic variables and level of consumption: findings from a national survey in Canada. J Stud Alcohol 1993;54:590–9.
16. Treno AJ, Alaniz ML, Gruenewald PJ. The use of drinking places by gender, age and ethnic groups: an analysis of routine drinking activities. Addiction 2000;95:537–51.
17. Lang E, Stockwell T, Rydon P, et al. Drinking settings and problems of intoxication. Addiction Res 1995;3:141–9.
18. Casswell S, Zhang JF, Wyllie A. The importance of amount and location of drinking for the experience of alcohol-related problems. Addiction 1993;88:1527–34.
19. Stockwell T, Lang E, Rydon P. High risk drinking settings: the association of serving and promotional practices with harmful drinking. Addiction 1993;88:1519–26.
20. Gruenewald PJ, Stockwell T, Beel A, Dyskin EV. Beverage sales and drinking and driving: the role of on-premise drinking places. J Stud Alcohol 1999;60:47–53.
21. Midanik LT. Drunkeness, feeling the effects and 5+ measures. Addiction 1999;94:887–97.
22. National Institute of Alcohol Abuse and Alcoholism. NIAAA Council approves binge drinking definition. http://pubs.niaaa.nih.gov/publications/Newsletter/winter2004/Newsletter_Number3.htm.
23. 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.
24. Babor TF, Caetano R, Casswell S, et al. Alcohol: no ordinary commodity: research and public policy. New York: Oxford University Press, 2003.
25. National Institute of Alcohol Abuse and Alcoholism. Tenth Special Report to the U.S. Congress on Alcohol and Health. Bethesda MD: NIH, 2000.
26. Centers for Disease Control and Prevention. Guide to Community Preventive Services. Preventing excessive alcohol use. Available at http://thecommunityguide.org/alcohol/index.html. Accessed July 14, 2009.
27. Shults RA, Elder RW, Sleet DA, et al. Reviews of evidence regarding interventions to reduce alcohol-impaired driving. Am J Prev Med 2001;21(4S):66–88.
28. O'Donnell MA. Research on drinking locations of alcohol-impaired drivers: implications for prevention policies. J Pub Health Pol 1985;6:510–25.
29. McKnight JA, Streff FM. The effect of enforcement upon service of alcohol to intoxicated patrons of bars and restaurants. Accid Anal Prev 1994;26:79–88.
30. Stout EM, Sloan FA, Liang L, Davies HH. Reducing harmful alcohol-related behaviors: effective regulatory methods. J Stud Alcohol 2000;61:402–12.
31. Alcohol Epidemiology Program. Alcohol polices in the United States: highlights from the 50 States. Minneapolis: University of Minnesota, 2000.

State Alcohol-Use Estimates Among Youth and Adults, 1993–2005

David E. Nelson, MD, MPH, Timothy S. Naimi, MD, MPH, Robert D. Brewer, MD, MSPH, Hayley A. Nelson

Background: Underage drinking, particularly binge drinking, is an important public health problem that results in substantial premature mortality and morbidity. Little is known about the potential influence of the alcohol-use behaviors of adults on youth alcohol use at a population level. The purpose of this study was to examine the correlation of alcohol-use behaviors among youth with those of adults at a population level.

Methods: Data were analyzed in 2007 and 2008, using biennial 1993–2005 data from state school-based Youth Risk Behavior Surveys of students in grades 9–12, and from the Behavioral Risk Factor Surveillance System for adults aged ≥18 years. Pearson correlation coefficients (r) were used to compare state prevalence estimates for youth with those of adults for several alcohol-use measures.

Results: Overall and subgroup-specific state youth estimates of current drinking and binge drinking were generally moderately to strongly correlated with adult alcohol use (range of r-values for pooled estimates across all years: 0.35–0.68 for current drinking [$p<0.01$ for all correlations]; 0.24–0.60 for binge drinking [$p<0.01$ for all correlations]) and with youth and adult drinking-and-driving behaviors (range of r-values for pooled estimates: 0.12–0.52, $p<0.01$ for all but one correlation). Correlation coefficients were generally higher for girls with women and for youth with younger adults aged 18–34 years. The use of alcohol by youth before they were aged 13 years was not correlated with adult alcohol-use measures, and most youth alcohol-use measures were not correlated with adult heavy-alcohol use.

Conclusions: Most state youth alcohol-use estimates were correlated with state adult estimates. These findings have implications for underage-drinking control strategies and suggest that efforts to address this problem need to be targeted on a broader societal level.

(Am J Prev Med 2009;36(3):218–224) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Excessive alcohol use is the third leading preventable cause of death in the U.S.[1] It accounts for an average of approximately 79,000 deaths annually as well as substantial health morbidity and broader societal adverse consequences, such as violence, unintended pregnancy, and lost productivity.[2–5]

Binge drinking, which is generally defined as the consumption of ≥5 drinks in a row on a single occasion,[6] is a hazardous drinking pattern and is especially common among adolescents.[5,7–13] As with many health risk behaviors, alcohol use usually begins in adolescence.[4,5,9,14] In addition to the health risks posed by alcohol use during adolescence itself (e.g., motor-vehicle crashes), the early onset of alcohol use is associated with a substantially greater risk of alcohol misuse during adulthood.[9,14]

A substantial research literature exists on the factors associated with youth alcohol use, including genetics, demographics, psychological characteristics, and family and peer influences. Previous studies[4,5,15] have demonstrated a strong connection between youth and adult drinking at the household and the community levels. Somewhat surprisingly, little population-based research has been conducted on the relationship between youth and adult drinking at the state level.[8,16]

Assessing this relationship is important because young people may model their drinking patterns after adults beyond those living within their own households; in addition, youth alcohol use occurs within a broader societal context shaped by adults.[8,14,17–19] A need for a better understanding of the potential relationship between youth and adult drinking behavior on a population basis has practical implications, such as whether efforts to reduce underage drinking should be targeted to youth alone or to youth and adults more broadly.[14]

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 (Nelson DE, Naimi, Brewer), Atlanta; and the University of Georgia (Nelson HA), Athens, Georgia

Address correspondence and reprint requests to: David E. Nelson, MD, MPH, CDC, 4770 Buford Highway NE, Mailstop K67, Atlanta GA 30341-3717. E-mail: den2@cdc.gov.

0749-3797/09/$–see front matter
doi:10.1016/j.amepre.2008.10.018

To better assess potential population-based associations between youth and adult alcohol-use behaviors, biennial data from the state Youth Risk Behavior Surveys (YRBS) were correlated with data from the Behavioral Risk Factor Surveillance System (BRFSS). The purposes of this study were to determine if state alcohol-consumption patterns in youth and adults were correlated; if youth alcohol-impaired driving and adult alcohol-impaired driving were correlated; and to assess if there were changes in correlations over time. Because alcohol use among youth may be influenced more by younger-aged adults,[20] estimates for youth were compared to estimates for adults aged 18–34 years (younger adults) and adults aged ≥35 years (middle-aged and older adults). Estimates for boys were compared to those of men, and estimates for girls to those of women.

Methods

Youth Data

Youth alcohol-use prevalence estimates for the years 1993–2005 came from data tables published in the CDC's *Morbidity and Mortality Weekly Report* (MMWR) Surveillance Summaries from biennially conducted state YRBSs[7,21–26]; details are available elsewhere.[27] Briefly, states use the YRBS to anonymously collect data on health risk factors from students in grades 9–12. Students complete a self-administered paper-and-pencil questionnaire during a regular class period. Participation is anonymous and voluntary. Local parental-permission procedures are followed before survey administration.

Most states employ a two-stage cluster-sample design to produce representative samples of students in public schools. The total number of states with published data in the MMWR Surveillance Summaries ranged from 29 to 40. Median state sample sizes ranged from 1619 to 2760, and median state response rates ranged from 61% to 67.5%. Data from states with a representative sample of students, appropriate documentation, and an overall response rate of ≥60% were weighted to be representative of all students attending public schools in grades 9–12. Of all the states included in this study, only about two thirds had weighted data; unweighted data are representative of students who participated in the survey, not necessarily of all students statewide. (Because preliminary findings found virtually no differences in correlations when only weighted state YRBS data were used [data available from authors], this report includes YRBS data from every state published in each MMWR Surveillance Summary regardless of whether data were weighted or unweighted.)

Questions about alcohol use in general are contained on the YRBS. These include lifetime (ever) use of alcohol (respondent had at least 1 drink of alcohol on at least 1 day during his or her life); alcohol use for the first time before age 13 years (i.e., early age of initiation); current alcohol use (had at least 1 drink of alcohol on at least 1 day during the 30 days before the survey); and binge drinking (had ≥5 drinks of alcohol in a row within a couple of hours on at least 1 day during the 30 days before the survey). The YRBS also contain questions about alcohol use and motor vehicles, including driving after drinking alcohol (i.e., alcohol-impaired driving) and riding with a driver who had been drinking alcohol ≥1 time during the 30 days before the survey. The same wording for all the alcohol questions was used over the entire study period; YRBS questions are available at www.cdc.gov/healthyyouth/yrbs/data/index.htm.

Overall and gender-specific state data were available for all alcohol-use measures annually except for 1993; published data on alcohol-impaired driving and early age of initiation were available starting with the 1995 survey administration.[22] Because states have final decision-making authority about questions, not all alcohol questions were included each year. Each year the MMWR Surveillance Summaries contained no data on lifetime alcohol use from approximately two to seven states; in 1995 and 1997, no data on binge drinking from one state; and in 1995 and 2001, no data on early age of initiation from one state.

Adult Data

Biennial adult data (corresponding to the years for which YRBS data were available) were obtained from publicly available CDC BRFSS data sets for 1993–2005. (Although BRFSS data are collected annually, only biennial BRFSS data from odd-numbered years were used for this study in order to be consistent with the years that YRBS data on youth were available.) Details about the BRFSS are available elsewhere.[28] In brief, state-based random-digit-dial telephone surveys of non-institutionalized people aged ≥18 years are conducted monthly in all states; survey instruments contain questions on a variety of health risk measures. Overall, median state sample sizes for the years studied ranged from 2045 to 5812, and median state response rates ranged from 51.1% to 68.5%. Data for all states were weighted to be representative of each state's adult population.

Questions in the BRFSS covered current alcohol use; the frequency and quantity of alcohol use (consumption); and alcohol-impaired driving (the actual wording for BRFSS questions is available at www.cdc.gov/brfss). Current alcohol use was defined as having at least 1 drink of an alcohol-containing beverage (beer, wine, wine cooler, or liquor) in the past month from 1993 to 1999, and as having ≥1 drink in the past 30 days in 2005. In 2001 to 2003, respondents were asked about the number of days they drank an alcohol-containing beverage within the past 30 days; those reporting ≥1 day were considered current alcohol users.

Binge drinking was defined as drinking ≥5 alcohol-containing beverages on one or more occasions within the past month (1993–1999) or within the past 30 days (2001–2005). Heavy drinking was defined as having >2 drinks per day on average for men, and >1 drink per day on average for women, within the past month (1993–1999) or within the past 30 days (2001–2005). Alcohol-impaired driving was defined as having driven after perhaps having had too much to drink within the past month from 1993–1999; questions on alcohol-impaired driving on the BRFSS were not included on the core survey in 2001, 2003, or 2005, preventing correlations with this measure with YRBS data for these years.

Data Analyses

Data analyses occurred in 2007 and 2008 and were limited to the same states and years for which YRBS data were available. Pearson correlation coefficients (r) were used in all analyses,

Table 1. Correlations of state youth and adult alcohol-use measures: direct comparisons, overall and by subgroups, 1993–2005

	1993	1995	1997	1999	2001	2003	2005	1993–1999[a]	2001–2005	All years
Number of states	n=29	n=30	n=32	n=33	n=33	n=32	n=40			
Current drinking										
All youth/adults	0.34	0.51**	0.55**	0.52**	0.71**	0.45	0.57**	0.48**	0.55**	0.43**
Boys/men	0.25	0.41*	0.47**	0.52**	0.66**	0.40	0.52**	0.41**	0.51**	0.35**
Girls/women	0.44*	0.58**	0.59**	0.50**	0.70**	0.46**	0.59**	0.53**	0.56**	0.49**
All youth/adults aged 18–34	0.46*	0.62**	0.59**	0.59**	0.79**	0.58**	0.65**	0.56**	0.68**	0.55**
All youth/adults aged ≥35	0.29	0.45	0.51**	0.47**	0.64**	0.39	0.52**	0.44**	0.48**	0.35**
Binge drinking										
All youth/adults	0.16	0.36	0.42*	0.38*	0.64**	0.38*	0.46**	0.34**	0.49**	0.36**
Boys/men	0.00	0.27	0.32	0.34	0.63**	0.34	0.39*	0.25**	0.46**	0.27**
Girls/women	0.26	0.50	0.54**	0.35*	0.58**	0.33	0.41**	0.41**	0.44**	0.40**
All youth/adults aged 18–34	0.19	0.46*	0.50**	0.46**	0.71**	0.51**	0.52**	0.42**	0.60**	0.45**
All youth/adults aged ≥35	0.07	0.31	0.36*	0.29	0.54**	0.22	0.33*	0.29**	0.33**	0.24**
Alcohol-impaired driving										
All youth/adults	NA	0.62**	0.52**	0.52**	NA	NA	NA	0.53**	NA	NA
Boys/men	NA	0.55**	0.36*	0.53**	NA	NA	NA	0.45**	NA	NA
Girls/women	NA	0.51**	0.52**	0.37**	NA	NA	NA	0.45**	NA	NA
All youth/adults aged 18–34	NA	0.66**	0.58**	0.43*	NA	NA	NA	0.52**	NA	NA
All youth/adults aged ≥35	NA	0.49**	0.48**	0.54**	NA	NA	NA	0.49**	NA	NA

[a]1995–1999 only for alcohol-impaired driving
*p<0.05; **p<0.01
NA, not available

and significance, based on p-values <0.05, was used to assess if coefficients were different from 0. Correlations of 0.10–0.29 were considered weak, 0.30–0.49 moderate, and ≥0.50 strong.[29] Five broad types of prevalence correlations were estimated: overall youth with overall adults; overall youth with adults aged 18–34 years; overall youth with adults aged ≥35 years; boys with men; and girls with women.

Because of the differences in alcohol-use measures in the YRBS and BRFSS surveys, correlations were further categorized as direct or indirect. Direct comparisons consisted of correlations of the prevalence measures for current alcohol use, binge drinking, and alcohol-impaired driving for youth and adults (e.g., youth and adult binge-drinking estimates). Indirect comparisons consisted of correlating the remaining alcohol-use measures (e.g., youth binge drinking with adult heavy drinking).

Correlations were obtained separately by year, pooled for the time periods 1993–1999 and 2001–2005, and pooled across all years. Correlations using youth early age of initiation and youth alcohol-impaired driving were available only for 1995 to 2005, and comparisons with adults were performed for these years. Correlations of youth data with adult alcohol-impaired driving were conducted only for period 1993–1999. Direct comparisons (e.g., binge drinking among youth with binge drinking among adults) were performed both overall and for each subgroup for each year and for pooled years; indirect comparisons were limited to overall estimates and pooled years.

Results

The youth median overall state prevalence and the range for medians across the survey years for the alcohol-use measures were 78.4% (74.1%–80.8%) for lifetime alcohol use; 47.9% (42.8%–51.1%) for current alcohol use; 30.4% (26.3%–33.1%) for binge drinking; 30.3% (24.9%–34%) for early age of initiation; 13.9% (11.1%–15.2%) for alcohol-impaired driving; and 34.1% (27.2%–36.6%) for riding with a driver who had been drinking alcohol. The median overall adult state prevalence and the range for medians across the survey years for the alcohol-use measures were 55% (51.4%–58.4%) for current alcohol use; 15% (13.8%–16.9%) for binge drinking; 4.7% (3.2%–5.9%) for heavy drinking; and 2.1% (2%–2.4%) for alcohol-impaired driving.

Correlations of state youth and adult alcohol prevalence for direct-comparison measures are included in Table 1. Overall, the pooled estimates of current youth alcohol use were moderately correlated with current adult alcohol use (r =0.43, p<0.01). Demographic-specific estimates of current youth alcohol use were consistently correlated in a positive manner with current adult alcohol use (range of r-values for pooled estimates: 0.35–0.68; range of r-values for annual estimates: 0.25–0.79; Table 1), with nearly all coefficients significant at p<0.01. Coefficients for all comparisons were typically higher from 2001 to 2005 than in earlier years. Pooled and annual correlation coefficients for current alcohol use were generally higher among female youth and adults than among male youth and adults, and for youth estimates correlated with younger adults compared to older adults.

There were positive correlations for youth and adult binge-drinking prevalence, although correlation coefficients were slightly lower (the majority were moderately correlated) and annual findings somewhat less consistent than those for current drinking (Table 1 and

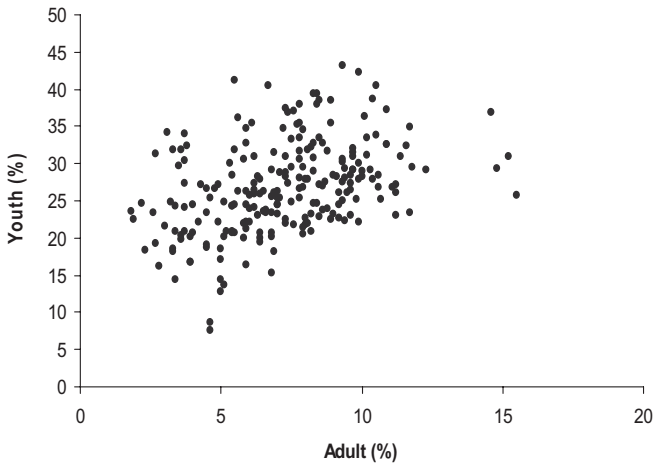


Figure 1. Correlation of binge drinking among youths and adults, 1993–2005

Figure 1; range of *r*-values for pooled estimates: 0.24–0.60; range of *r*-values for annual estimates: 0.00–0.71). As with current drinking, youth and adult binge-drinking correlations were generally stronger from 2001 to 2005, and were higher among girls and women and younger adults (Table 1). Youth alcohol-impaired driving estimates from 1995 to 1999 were moderately to strongly correlated with those of adult alcohol-impaired driving for all years and among all subgroups (Table 1; range of *r*-values for pooled estimates: 0.45–0.53; range of *r*-values for annual estimates: 0.36–0.66), with all coefficients significant at $p<0.01$.

Table 2 contains correlation coefficients of state overall youth and adult alcohol prevalence for indirect-comparison measures. State youth estimates of lifetime, current, and binge drinking were moderately to strongly correlated with state adult estimates of current alcohol use, binge drinking, and drinking-and-driving pooled across all years (range of *r*-values: 0.30–0.52, $p<0.01$ for all correlation coefficients) but not with adult heavy-drinking estimates. The correlation for youth riding-with-a-drinking-driver estimates with adult alcohol-impaired driving estimates from 1993 to 1999 was moderate (0.42, $p<0.01$).

Early age of initiation was not correlated with any adult alcohol-use measures. Youth estimates for alcohol-impaired driving and for riding with a drinking driver consistently correlated—albeit at a weak-to-moderate level—with adult binge-drinking estimates (range of *r*-values for pooled estimates: 0.18–0.44) but were inconsistently correlated with estimates of adult current alcohol use and heavy drinking.

Table 2. Correlations of state youth and adult alcohol-use measures: indirect comparisons, overall and pooled years, 1993–2005

	1993–1999[a,b]	2001–2005	All years[c]
Lifetime youth alcohol use with:			
Adult current alcohol use	0.40**	0.49**	0.40**
Adult binge drinking	0.30**	0.40**	0.31**
Adult heavy drinking	0.16	0.28**	0.12
Adult alcohol-impaired driving	0.33**	NA	NA
Current youth alcohol use with:			
Adult binge drinking	0.40**	0.52**	0.39**
Adult heavy drinking	0.09	0.30**	0.03
Adult alcohol-impaired driving	0.43**	NA	NA
Youth binge drinking with:			
Adult current alcohol use	0.32**	0.42**	0.31**
Adult heavy drinking	−0.01	0.14	−0.05
Adult alcohol-impaired driving	0.39**	NA	NA
Early age of initiation of alcohol use with:			
Adult current alcohol use	0.16	0.01	0.00
Adult binge drinking	0.10	0.05	0.02
Adult heavy drinking	0.12	0.11	−0.09
Adult alcohol-impaired driving	0.18	NA	NA
Youth alcohol-impaired driving with:			
Adult current alcohol use	0.15	0.25**	0.14*
Adult binge drinking	0.31**	0.44**	0.32**
Adult heavy drinking	−0.18	0.05	−0.18**
Youth riding with a drinking driver with:			
Adult current alcohol use	0.08	0.25**	0.05
Adult binge drinking	0.18*	0.40**	0.18**
Adult heavy drinking	−0.13	0.14	−0.20**
Adult alcohol-impaired driving	0.42**	NA	NA

[a]1995–1999 only for youth alcohol-impaired driving, and youth early age of initiation of alcohol use, with all adult alcohol-use measures
[b]1993–1999 only for all youth alcohol-use measures with adult alcohol-impaired driving
[c]1995–2005 only for youth alcohol-impaired driving, and youth early age of initiation of alcohol use, for correlations with adult alcohol-consumption measures
*$p<0.05$; **$p<0.01$
NA, not available

Discussion

This may be the first population-based study of the relationship between youth and adult alcohol-use measures at the state level. This study empirically demonstrated that state-level adult and youth alcohol use are generally correlated to a moderate or strong degree. These correlations were generally found for most state-level alcohol-use measures between youth and adults and were consistent over the 13-year time period, with some evidence that correlations were stronger in recent years. Correlations were higher for girls with women than for boys with men, and

for youth with younger adults than for youth with older adults. Youth estimates for early age of initiation did not correlate with adult alcohol-use measures, and youth alcohol-use measures did not correlate with adult heavy drinking. These findings provide further evidence of the need to address underage drinking through broader societal approaches that also influence excessive adult alcohol use.

Comparing these findings with prior work is difficult because there are few similar studies. However, this research is consistent with findings from a study by Nelson et al.,[8] which demonstrated that adult binge-drinking estimates for 40 states were strongly correlated with binge drinking among college students and with young adults aged 18–24 years and not attending college who resided in the same state ($r=0.43$ and 0.45, $p<0.01$, respectively). A study comparing YRBS findings for high school students with a telephone-administered YRBS survey of college students in Texas found that college students had a slightly higher prevalence of binge drinking and a slighter older age of first alcohol use than did high school students.[30]

The only study[31] involving a comparison of YRBS with BRFSS data was based on a survey of American Indians living on or near reservations in Montana. Although no comparisons were available on alcohol-use measures, that study found that the prevalence of youth tobacco use, physical inactivity, weight loss, and low levels of fruit and vegetable consumption was generally similar to, or higher than, that of adults. That study also found that risk-factor estimates about girls were generally more similar to those for women as opposed to estimates about boys compared to men.

A variety of factors at multiple levels influence youth alcohol use (e.g., religion, family members, peers).[4,5,14,15,17] One reason for the correlations between youth and adult alcohol-use measures is that youth drinking patterns may persist into adulthood.[32] Additionally, research has consistently shown that adolescents who begin drinking heavily at younger ages are at much greater risk of maintaining heavy use into adulthood and developing alcohol-use disorders.[5,15] However, these findings of generally moderate-to-strong correlations between population-level estimates of youth and adult alcohol-use behaviors suggest that environmental influences, such as social, cultural, and legal factors, may have an important effect on youth alcohol use.[14,17] This study also provides further evidence of the need for more scientifically rigorous research, such as intervention trials, on the effects of policies to reduce underage drinking.

Policies that result in higher alcohol excise taxes, in the enforcement of minimum legal drinking age laws, and in the restricted availability of alcohol reduce alcohol use among youth.[3,8,33–39] In the 2005 Nelson et al. study[8] of college students, alcohol-control policies—including those geared toward the general population rather than just underaged people—were independently associated with reduced binge drinking by college students, even after adjusting for the impact of same-state adult drinking patterns, emphasizing the impact that the policy environment may have on youth and adult drinking.

Possible reasons are unknown for the lack of state-level correlations between an early age of initiation of alcohol use among youth and adult alcohol use. Decisions by youth to begin drinking at a young age are complex and likely to be affected by other factors (e.g., genetics, parental alcohol use, peer pressure) besides adult alcohol-use patterns among nonhousehold adults.[4,5,15] Similarly, other factors may account for the general lack of positive correlations for youth alcohol-use measures with adult heavy-drinking estimates.

This study had limitations. This was an ecologic analysis, and correlation does not equal causation; the strong youth–adult correlations for most alcohol-use behaviors could be the result of other factors. Data were based on self-reports, which underestimate alcohol consumption, at least among adults.[40] There were wording differences in the questions used in the YRBS and BRFSS, which may have influenced correlations. There was the potential for a slight overlap between YRBS and BRFSS respondents, given that some high school students are aged ≥18 years, although the effect of this overlap on BRFSS estimates would be small. Response rates for the BRFSS, as with other telephone-based surveys, declined over time,[41] and the impact on correlations is not known. Both weighted and unweighted estimates were used from the YRBS; however, preliminary findings based on direct comparisons were similar only when weighted YRBS data were used. In addition, YRBS data are representative only of youth who attend school and not of all students in this age group. Finally, the generally stronger correlation between youth and adult alcohol-impaired driving should be interpreted cautiously, given that correlations could be calculated only through 1999.

This study's findings have implications for public health efforts to reduce excessive alcohol use. Regardless of the reasons for the youth–adult alcohol associations, the two are correlated. This is important because many citizens, policymakers, and even public health officials frame youth drinking as an age-related problem that should be addressed through youth-centered approaches. Certainly some youth-centered approaches are effective, such as the presence and enforcement of minimum legal drinking-age laws; laws that establish a lower legal blood alcohol concentration for young or inexperienced drivers (i.e., zero-tolerance laws); and school-based instructional programs to reduce riding with alcohol-impaired drivers.[36,38] A youth-centered approach, however, can result in a lack of focus on reducing excessive drinking among the entire population through the use of policy interventions.[42,43]

Higher alcohol taxes, for example, particularly those on beer, are inversely related to youth drinking, the frequency of youth drinking, heavy drinking by youth, youth motor-vehicle crashes, homicide, suicide, and youth violence.[33,44–48] More-comprehensive sets of state policies, including those geared toward the general population, are associated with less alcohol-impaired driving among college students,[18] including those who are underage[19]—e.g., one study found that states that lowered their legal limit for blood alcohol concentration to the 0.08 level experienced reductions in beer sales.[49] Similar findings have been shown for tobacco control.[50,51] Reducing underage drinking, as noted by the 2004 National Research Council/IOM Underage Drinking Report, will require a broader focus on reducing excessive drinking in the entire population.[14]

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the USDHHS or the CDC.

No financial disclosures were reported by the authors of this paper.

References

1. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the U.S., 2000. JAMA 2004;291:1238–45.
2. CDC. Alcohol & public health. Atlanta GA: CDC, 2008. www.cdc.gov/alcohol.
3. Babor TF, Caetano, R, Casswell S, et al. Alcohol and public policy: no ordinary commodity. New York: Oxford University Press, 2003.
4. National Institute on Alcohol Abuse and Alcoholism. Tenth special report to Congress on alcohol and health. Bethesda MD: National Institute on Alcohol Abuse and Alcoholism, NIH, USDHHS, 2000.
5. Galantier M, ed. Alcohol problems in adolescents and young adults: epidemiology, neurobiology, prevention, and treatment. New York: Springer, 2006.
6. Johnston LD, O'Malley PM, Bachman JG. National survey results on drug use, 1975–2000. Volume II: college students and young adults ages 19–40. NIH Pub. No. 01-4925. Bethesda MD: National Institute on Alcohol Abuse and Alcoholism, NIH, USDHHS, 2001.
7. Eaton DK, Kann L, Kinchen SA, et al. Youth risk behavioral surveillance—U.S., 2005. MMWR CDC Surveill Summ 2006;55(SS-05):1–108.
8. Nelson TF, Naimi TS, Brewer RD, Wechsler H. The state sets the rate: the relationship among state-specific college binge drinking, state binge drinking rates, and selected state alcohol control policies. Am J Public Health 2005;95:441–6.
9. USDHHS. A comprehensive plan for preventing and reducing underage drinking. Report to Congress. Washington DC: Substance Abuse and Mental Health Services Administration, 2006. www.stopalcoholabuse.gov/media/underagedrinking/pdf/underagerpttocongress.pdf.
10. Johnston LD, O'Malley PM, Bachman JG, Schulenberg JE. Monitoring the future: national survey results on drug use, 1975–2006. Volume I: Secondary school students. NIH Pub. No. 07-6205. Bethesda MD: National Institute on Drug Abuse, 2007.
11. Substance Abuse and Mental Health Services Administration (2007). 2006 National Survey on Drug Use and Health: detailed tables. www.oas.samhsa.gov/NSDUH/2k6NSDUH/tabs/TOC.htm.
12. Miller JW, Naimi TS, Brewer RD, Jones SE. Binge drinking and associated health risk behaviors among high school students. Pediatrics 2007;119:76–85.
13. Pacific Institute for Research and Evaluation. Drinking in America: myth, realities, and prevention policy. Calverton MD: Pacific Institute for Research and Evaluation, 2002.
14. National Research Council and IOM. Reducing underage drinking: a collective responsibility. Washington DC: The National Academies Press, 2004.
15. Romer D, ed. Reducing adolescent risk: toward an integrated approach. Thousand Oaks CA: Sage Publications, 2003.
16. Shults RA, Sleet DA, Elder RW, Ryan GW, Sehgal M. Association between state level drinking and driving countermeasures and self-reported alcohol-impaired driving. Inj Prev 2002;8:106–10.
17. Cook PJ, Moore MJ. Environment and persistence in youthful drinking patterns. In: Gruber J, ed. Risky behavior among youths: an economic analysis. Chicago IL: University of Chicago Press, 2001.
18. Wechsler H, Lee JE, Nelson TF, Kuo M. Underage college students' drinking behavior, access to alcohol, and the influence of deterrence policies: findings from the Harvard School of Public Health College Alcohol Study. J Am Coll Health 2002;50:223–36.
19. Wechsler H, Lee JE, Nelson TF, Lee H. Drinking and driving among college students: the influence of alcohol control policies. Am J Prev Med 2003;25:212–8.
20. Adams G. Reconsidering adult influences on adolescent drug use. Proceedings of the annual meeting of the American Statistical Association; 2006 Aug 10; Montreal, Canada.
21. Kann L, Warren CW, Harris W, et al. Youth risk behavior surveillance—U.S., 1993. MMWR CDC Surveill Summ 1995;44(SS-01):1–56.
22. Kann L, Warren CW, Harris W, et al. Youth risk behavior surveillance—U.S., 1995. MMWR CDC Surveill Summ 1996;45(SS-04):1–86.
23. Kann L, Kinchen SA, Williams BI, et al. Youth risk behavior surveillance—U.S., 1997. MMWR CDC Surveill Summ 1998;47(SS-03):1–92.
24. Kann L, Kinchen SA, Williams BI, et al. Youth risk behavior surveillance—U.S., 1999. MMWR CDC Surveill Summ 2000;49(SS-05):1–96.
25. Grunbaum JA, Kann L, Kinchen SA, et al. Youth risk behavior surveillance—U.S., 2001. MMWR CDC Surveill Summ 2002;51(SS-04):1–64.
26. Grunbaum JA, Kann L, Kinchen SA, et al. Youth risk behavior surveillance—U.S., 2003. MMWR CDC Surveill Summ 2004;53(SS-02):1–96.
27. Brener ND, Kann L, Kinchen SA, et al. Methodology of the youth risk behavior surveillance system. MMWR Recomm Rep 2004b;53(No. RR-12):1–14.
28. CDC. BRFSS annual survey data: survey data and documentation. www.cdc.gov/brfss/technical_infodata/surveydata.htm.
29. Cohen J. Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale NJ: Lawrence Erlbaum Associates, 1988.
30. Wiley DC, James G, Furney S, Jordan-Belver C. Using the Youth Risk Behavior Survey to compare risk behaviors of Texas high school and college students. J Sch Health 1997;67:45–9.
31. Nelson DE, Moon RW, Holtzman D, Smith P, Siegel PZ. Patterns of health risk behaviors for chronic disease: a comparison between adolescent and adult American Indians living on or near reservations in Montana. J Adolesc Health 1997;21:25–32.
32. O'Malley PM, Wagenaar AC. Effects of minimum drinking age laws on alcohol use, related behaviors and traffic crash involvement among American youth, 1976–1987. J Stud Alcohol 1991;52:478–91.
33. Chaloupka FJ, Grossman M, Saffer H. The effects of price on alcohol consumption and alcohol-related problems. Alcohol Res Health 2002;26:22–34.
34. Hahn R. Effectiveness of alcohol outlet density regulation in reducing excessive alcohol consumption and related harms. Proceedings of the meeting of the Task Force on Community Preventive Services; 2007 Feb 14–15; Atlanta GA.
35. Task Force on Community Preventive Services. Guide to community preventive services—excessive alcohol consumption. Atlanta GA: Task Force on Community Preventive Services, 2008. www.thecommunityguide.org/alcohol/default.htm.
36. Task Force on Community Preventive Services. Guide to community preventive services—motor vehicle occupant injury. Atlanta GA: Task Force on Community Preventive Services, 2008. www.thecommunityguide.org/mvoi/default.htm.
37. Shults RA, Elder RW, Sleet DA, et al; the Task Force on Community Preventive Services. Reviews of evidence regarding interventions to reduce alcohol-impaired driving. Am J Prev Med 2001:20(4S);66–84.
38. Elder RW, Nichols JL, Shults RA, et al. Effectiveness of school-based programs for reducing drinking and driving and riding with drinking drivers: a systematic review. Am J Prev Med 2005;28(5S):288–304.
39. Task Force on Community Preventive Services. Recommendations to reduce injuries to motor vehicle occupants: increasing child safety seat use, increasing safety belt use, and reducing alcohol-impaired driving. Am J Prev Med 2001:21(4S):16–22.

40. Harrison L, Hughes A, eds. The validity of self-reported drug use: improving the accuracy of survey estimates. NIDA Research Monograph 167. NIH Publication No. 97-4147. Bethesda MD: National Institute on Drug Abuse, NIH, USDHHS, 1997.
41. Kempf AM, Remington PL. New challenges for telephone survey research in the twenty-first century. Annu Rev Public Health 2007;28:113–26.
42. USDHHS. The Surgeon General's call to action to prevent and reduce underage drinking. Rockville MD: USDHHS, 2007.
43. Holder HD, Treno AJ. Moving toward a common evidence base for alcohol and other drug prevention policy. In: Stockwell T, Gruenewald PJ, Toumbourou JW, Loxley W, eds. Preventing harmful substance use: the evidence base for policy and practice. New York: John Wiley & Sons, 2005:351–66.
44. Dee TS. State alcohol policies, teen drinking and traffic accidents. J Public Economics 1999;72:289–315.
45. Coate D, Grossman M. Effects of alcoholic beverage prices and legal drinking ages on youth alcohol use. J Law Econ 1988;31:145–71.
46. Chaloupka FJ, Saffer H, Grossman M. Alcohol-control policies and motor-vehicle fatalities. J Legal Stud 1993;22:161–86.
47. Saffer H, Grossman M. Beer taxes, the legal drinking age, and youth motor vehicle fatalities. J Legal Stud 1987;16:351–74.
48. Lockhart SJ, Beck KH, Summons TG. Impact of higher alcohol prices on alcohol-related attitudes and perceptions of suburban, middle-class youth. J Youth Adolesc 1993;22:441–5.
49. Voas RB, Tippetts AS, Fell J. The relationship of alcohol safety laws to drinking drivers in fatal crashes. Accid Anal Prev 2000;32:483–92.
50. Jacobson PD, Lantz PM, Warner KE, et al. Combating teen smoking: research and policy strategies. Ann Arbor MI: University of Michigan Press, 2001.
51. Lantz PM. Youth smoking prevention policy: lessons learned and continuing challenges. In: Bonnie RJ, O'Connell ME, eds. Reducing underage drinking: a collective responsibility. Washington DC: The National Academies Press, 2004.

Health Literacy

Communication Strategies to Improve Patient Comprehension of Cardiovascular Health

Daniel J. Oates, MD, MSc; Michael K. Paasche-Orlow, MD, MA, MPH

Case presentation: A 67-year-old retired school bus driver presents to your office for an initial visit after having had an acute myocardial infarction, which is complicated by new-onset congestive heart failure. She comes to your office alone, with a bag of 5 pill bottles, and asks, "Do I really need all these pills?"

To care for themselves and participate in their health care, patients must be able to understand and act on information and instructions given to them by their healthcare providers. This concept is known as health literacy, which is defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and health services needed to make appropriate health decisions."[1] Basic literacy skills, such as proficiency in reading, writing, listening, interpreting images, and interacting with documents, as well as facility with numeric concepts and basic computation, are central to the concept of health literacy and greatly affect a patient's level of health literacy.

The Institute of Medicine, American Medical Association, American College of Physicians, and the Joint Commission have targeted health literacy as a cross-cutting priority area for quality improvement to transform US health care.[2–5] Patients with the largest disease burdens are often those with the least ability to understand and use health information. This is due in part to a lack of focus on patient education and poor communication skills by clinicians. In this article, we discuss the prevalence of limited health literacy, its impact on health outcomes and healthcare utilization, and strategies that providers may use to enhance their communication skills.

The Problem

According to the 2003 National Assessment of Adult Literacy, a 30 000-household US Department of Education survey, 36% of US adults possess basic or below-basic health literacy skills.[6] For people with basic health literacy, most documents such as patient education brochures, informed consent forms, notices of privacy protection, patient bills of rights, and even pill bottles are far too complex. The prevalence of limited health literacy is higher for those with low educational attainment, the elderly, racial and ethnic minorities, and people with chronic disease.[7] Indeed, more than 50% of those 80 to 84 years old and more than 70% of patients 85 years old and older have marginal or limited health literacy.[8]

The Impact

Patients with limited health literacy have worse diabetic control[9]; often present with more advanced diseases, such as prostate cancer[10]; use fewer preventative services[11]; and are up to twice as likely to be hospitalized.[12] Additionally, older adults with limited health literacy have a hazard ratio for mortality over a 5-year period of 1.52 compared with those with normal health literacy.[13] Many factors account for this worse health status, including an increasingly complex healthcare system, difficulties accessing healthcare, limitations in patient-provider communication, and the failure of providers to promote self-management and recognize patient barriers to communication and comprehension.[14]

Numerous barriers to healthcare access exist for those with limited health literacy. Insurance companies and gov-

From the Section of Geriatrics (D.J.O.) and Section of General Internal Medicine (M.K.P.-O.), Department of Medicine, Boston University School of Medicine, Boston, Mass.

Correspondence to Daniel J. Oates, MD, MSc, Geriatrics, Robinson 2, Boston Medical Center, 72 E Concord St, Boston, MA 02118. E-mail daniel.oates@bmc.org

(*Circulation*. 2009;119:1049-1051.)

***Circulation* is available at http://circ.ahajournals.org** **DOI: 10.1161/CIRCULATIONAHA.108.818468**

Table. Clear Communication Strategies

Guiding Principles	Specific Steps
Clinical skills	1. Avoid jargon.
	2. Use simple sentence structure and plain language.
	3. Speak slowly.
	4. Use analogies, if appropriate (eg, "Getting a pacemaker is like replacing the electrical wiring in your house").
	5. Limit the amount of information discussed:
	Focus on 2 or 3 key points per visit and repeat them.
	Use others (office staff, VNA, home physical therapist, etc) to help reinforce key points.
Be specific	1. Use clear, action-oriented directives.
	2. Stress action steps the patient should take.
	3. Stress concrete, specific steps that the patient can take.
	4. Minimize information about anatomy and physiology.
	5. Focus instead on answering the patient's question, "What do I need to do?"
Use multiple forms of communication	1. Use more than 1 communication modality to give the most important information.
	2. Pictures can help convey complex information or explain procedures.
	3. Videos or interactive computer programs may also be useful.
	4. Get feedback from patients to make certain such patient education materials work with your patients.
Help patients ask questions	1. Create an environment conducive to patients asking questions.
	Instead of asking, "Do you have any questions?" you can ask, "What questions do you have for me?"
	2. Empower your patients to always leave medical encounters knowing the answer to the question, "What do I need to do?"
Confirm comprehension	1. Conduct "teach back."
	Part A: "Tell me what you'll tell your family about what we talked about."
	Part B: Focus feedback on aspects not understood.
	Part C: Reevaluate comprehension ("close the loop") and provide additional feedback until mastery has been exhibited.

VNA indicates Visiting Nurse Association.

ernment programs often introduce hurdles for those seeking care in the form of application procedures and paperwork, which deter those with literacy problems from seeking care, often owing to embarrassment or perceived shame from their limited literacy.[15]

Barriers can be present within the patient-provider relationship itself that make adequate communication and comprehension difficult. Providers often assume that their patients are functionally literate and communicate with them assuming they are able to read and comprehend information, although this often is not the case.[16] Clinicians can often be rushed and therefore make patients feel rushed and embarrassed to ask questions. The office visit can be a daunting interaction, especially for those with limited health literacy. Patients often prefer to be quiet than to admit that they do not understand their doctor's instructions. They fear that their limited literacy skills will be revealed.[15]

Strategies for Clear Communication

Numerous strategies are available that clinicians can implement that will help their patients overcome limited health literacy (Table).[17] Some of these communication techniques appear easy to implement; however, these strategies often require practice and the participation and training of an interdisciplinary team, as well as feedback from patients.

The goal is to help patients become informed and activated.[18] This cannot be achieved without a welcoming environment in which patients are comfortable asking questions. Shame is a prominent emotion that patients with limited literacy associate with medical encounters. Everything from registration to referrals should be made clear and simple. If you are not hearing questions, patients do not feel welcome to ask. Who are the people in your healthcare setting with the responsibility to elicit and answer patients' questions? Do they help patients feel comfortable asking questions? There are many ways to distribute this responsibility of eliciting and answering questions, but if the tasks are not clearly defined, achievement of the objective is unlikely.

Avoiding the use of medical jargon during the encounter is another important way to improve patient comprehension. Medical providers often use terms that are straightforward to them, yet may not be so to patients. Commonly heard jargon such as the words "echo," "stress

test," and "EKG" may confuse patients and make them fearful unless these words are explained. Use of jargon can be a subconscious technique providers use to assert their role as a health professional and exhibit the mastery they have of their topic area. Unfortunately, it does not promote patient understanding. To make matters worse, even simple words can function as jargon. For example, medical providers tend to use the term "diet" to refer to all the food a person consumes. Patients, however, tend to use the word "diet" to refer to an effort to lose weight. It can be hard to identify and drop the jargon; feedback from non-health professionals can be useful. Taking time to explain in plain terms the action steps you want patients to take will help improve patient understanding, and it can be an effective way for providers to show that it is important to them that their patients understand.

Universal Precautions

The ultimate way to ensure that communication with your patient has been successful is to check. In doing this, physicians often ask, "So, do you understand?" (while getting up and walking for the door, training the patient to respond "yes"). This is not a helpful check for comprehension. A more effective technique is to conduct a "teach back," in which you ask the patient to explain to you or teach back the critical action items from the encounter. You may ask, "We talked about several things today. I want to be sure that it is clear what you are going to do, so please tell me, what is the plan?" or "When you go home, what will you tell your partner about what you need to do every day?" Such questions are helpful in determining the extent of understanding and also what parts of the action plan the patient may not have understood fully. Clinicians can then provide immediate feedback and educational efforts to correct items the patient did not comprehend. This may need to take a different form than simply repeating the idea. The success of this teaching then needs to be evaluated with another round of teach back to determine whether the information has been imparted successfully.[19]

Conclusions

Integration of the clear communication techniques outlined here may take practice and training for a wide range of clinical staff; however, the high prevalence and significant clinical impact of limited health literacy warrant the expenditure of time and resources. Implementation of the communication techniques presented will help create a prepared and proactive clinical team that will be able to empower patients with limited health literacy to become informed.[20]

Sources of Funding

Dr Oates is supported in part by a Geriatric Academic Career Award from the Health Resources and Services Administration (HRSA), No. 1 K01 HP00020-01.

Disclosures

None.

References

1. Ratazan SC, Parker RM. Introduction. In: Selden CR, Zorn M, Ratazan SC, Parker RM, compilers. *National Library of Medicine Current Bibliographies in Medicine: Health Literacy.* NLM publication No. CBM 2000-1. Bethesda, Md: National Institutes of Health, US Department of Health and Human Services; 2000.
2. Institute of Medicine, Committee on Identifying Priority Areas for Quality Improvement. *Priority Areas for National Action: Transforming Health Care Quality.* Washington, DC: National Academies Press; 2003.
3. "What Did the Doctor say?" Improving Health Literacy to Protect Patient Safety. Available at: http://www.jointcommission.org/NR/rdonlyres/D5248B2E-E7E6-4121-8874-99C7B4888301/0/improving_health_literacy.pdf. Accessed August 28, 2008.
4. Health literacy: report of the Council on Scientific Affairs, Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. *JAMA*. 1999;281:552–557.
5. ACP Foundation. Promoting Health Literacy. Available at: http://foundation.acponline.org. Accessed August 28, 2008.
6. Kunter M, Greenberg E, Jin Y, Paulsen C. *The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy (NECS 2006-483).* Washington, DC: US Department of Education, National Center for Education Statistics; 2006.
7. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielson-Bohlman LT, Rudd, RR. The prevalence of limited health literacy. *J Gen Intern Med*. 2005;20:175–184.
8. Gazmararian JA, Baker DW, Williams MV. Health literacy among Medicare enrollees in a managed care organization. *JAMA*. 1999; 281:545–551.
9. Schillnger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, Palacios J, Sullivan GD, Bindman AB. Association of health literacy with diabetes outcomes. *JAMA*. 2002;288:475–482.
10. Bennett CL, Ferreira NR, Davis TC, Kaplan J, Weinberger M, Kuzel T, Seday MA, Sartor O. Relationship between literacy, race and stage of presentation among low income patients with prostate cancer. *J Clin Oncol*. 1998;16:3101–3104.
11. Scott TL, Gazmararian JA, Williams MV, Baker DW. Health literacy and preventative healthcare use among Medicare managed care enrollees in a managed care organization. *Med Care*. 2002;40:475–482.
12. Baker DW, Parker RM, Williams MV. Health literacy and risk of hospital admission. *J Gen Intern Med*. 1998;13:794–798.
13. Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA. Health literacy and mortality among elderly persons. *Arch Intern Med*. 2007;167:1503–1509.
14. Paasche-Orlow MK, Parker RM. Improving the effectiveness of patient education: a focus on limited health literacy. In: King TE, Wheeler M, Fernandez A, eds. *Medical Management of Vulnerable and Underserved Patients: Principles, Practice and Populations.* New York, NY: McGraw-Hill; 2007.
15. Parikh NS, Parker RM, Nurss JR, Baker DW, Williams MV. Shame and health literacy: the unspoken connection. *Patient Educ Couns*. 1996;27:33–39.
16. Horowitz CR, Monteith S, McLaughlin M, Sisk JE, Chatterjee S. Low health literacy is common and unhealthy: do we recognize it in our own patients? *J Gen Intern Med*. 2004;19(suppl 1):176.
17. Hironaka LK, Paasche-Orlow MK. The implications of health literacy on patient-provider communication. *Arch Dis Child*. 2008;93:428–432.
18. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? *Eff Clin Pract*. 1998;1:2–4.
19. Schillinger D, Piette J, Grumbach K, Wang F, Wilson C, Daher C, Leong-Grotz K, Castro C, Bindman AB. Closing the loop: physician communication with diabetic patients who have low health literacy. *Arch Intern Med*. 2003;163:83–90.
20. Paasche-Orlow MK, Schillinger D, Green SM, Wagner EH. How health care delivery systems can begin to address the challenge of limited health literacy. *J Gen Intern Med*. 2006;21:884–887.

National Survey of Patients' Bill of Rights Statutes

Michael K. Paasche-Orlow[1], Dan M. Jacob[2], Mark Hochhauser[3], and Ruth M. Parker[4]

[1]Section of General Internal Medicine Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [2]Healthcare Analytics, LLC, New York, NY, USA; [3]Readability Consultant, Golden Valley, MN, USA; [4]Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA.

BACKGROUND: Despite vigorous national debate between 1999–2001 the federal patients' bill of rights (PBOR) was not enacted. However, states have enacted legislation and the Joint Commission defined an accreditation standard to present patients with their rights. Because such initiatives can be undermined by overly complex language, we surveyed the readability of hospital PBOR documents as well as texts mandated by state law.

METHODS: State Web sites and codes were searched to identify PBOR statutes for general patient populations. The rights addressed were compared with the 12 themes presented in the American Hospital Association's (AHA) PBOR text of 2002. In addition, we obtained PBOR texts from a sample of hospitals in each state. Readability was evaluated using Prose, a software program which reports an average of eight readability formulas.

RESULTS: Of 23 states with a PBOR statute for the general public, all establish a grievance policy, four protect a private right of action, and one stipulates fines for violations. These laws address an average of 7.4 of the 12 AHA themes. Nine states' statutes specify PBOR text for distribution to patients. These documents have an average readability of 15th grade (range, 11.6, New York, to 17.0, Minnesota). PBOR documents from 240 US hospitals have an average readability of 14th grade (range, 8.2 to 17.0).

CONCLUSIONS: While the average U.S. adult reads at an 8th grade reading level, an advanced college reading level is routinely required to read PBOR documents. Patients are not likely to learn about their rights from documents they cannot read.

KEY WORDS: patient rights; readability; policy; literacy.
J Gen Intern Med 24(4):489–94
DOI: 10.1007/s11606-009-0914-z

Received May 27, 2008
Revised December 3, 2008
Accepted January 12, 2009
Published online February 3, 2009

BACKGROUND

In 2001, both the U.S. House of Representatives and U.S. Senate passed bills to create a Federal Patients' Bill of Rights (PBOR). While the Senate version of the bill reversed certain elements of the Employee Retirement Income Security Act (ERISA), by allowing patients to sue in state and federal courts for denials of care by managed care organizations, the House version of the bill did not provide such a right and President Bush was reported to have threatened to veto the bill if it included such a provision.[1] The bill was moved to a House–Senate conference to work out differences between House-passed and Senate-passed bills, but these negotiations failed. Despite this, many states enacted Patients' Bill of Rights laws.[2,3]

The concept of patients' rights represents a cultural shift that began to emerge 40 years ago when notions of informed consent and autonomy were first endorsed by court opinion and institutional policy.[4,5] In 1973, the American Hospital Association

(AHA) presented the first patients' bill of rights.[6] The 12 themes addressed in this initial document (e.g., right to respectful care, right to refuse treatment, right to confidentiality, right to refuse participation in research) have remained in subsequent versions (Table 1), and in the 1990s the Joint Commission phased in a requirement to inform every patient about their rights as a national standard for hospital accreditation (RI.2.20).

Unfortunately, efforts to advance patients' rights can be thwarted by inadequate attention to the complexity and language of the materials presented to patients. For example, while the average U.S. adult reads at an 8th grade reading level, informed consent documents and notices of privacy practices typically require the reading capacity of a high school graduate.[7,8] We hypothesized that PBOR texts are also written at a level of complexity that far exceeds patients' average capacity. We therefore undertook a survey to determine the readability of PBOR texts in the United States. We included PBOR texts from a sample of U.S. hospitals and all PBOR texts designated by state law to be given to all patients. We performed the following three additional analyses of state PBOR statutes: 1) comparison of the rights delineated in state law to the themes advanced in the 2002 version of the American Hospital Association PBOR; 2) abstraction of any enforcement powers that are delineated within the statute; and 3) evaluation of the presence of PBOR texts in languages other than English for those states with mandatory language defined within the statute.

Table 1. Frequency of American Hospital Association Patients' Bill of Rights Themes in State Statutes and Hospital Documents

Theme The patient has the right to:	State Statutes (N = 23), %	Hospital Documents (N=240), %
1. Considerate and respectful care	78	97
2. Obtain current and understandable information	87	93
3. Refuse recommended treatment	87	97
4. Have an advanced directive	35	95
5. Privacy	87	93
6. Confidential communications and records	78	92
7. Review records	43	88
8. Indicated medical care including transfer to another facility	39	90
9. Be informed of business relationships that influence care	17	40
10. Refuse participation in research	74	58
11. Reasonable continuity of care	43	87
12. Be informed of charges as well as policies for patient responsibilities and resolution of conflicts	74	57

METHODS

Data Sources

We obtained state PBOR statutes by searching all 50 state government Web sites and legal codes in the Lexis-Nexus Data base. If this information was unclear, we contacted the legal counsel for the state Department of Public Health and Welfare and/or the legal counsel for the State Legislature. The focus of this analysis was PBOR material for general patient populations. As such, PBOR legislation intended for specific patient populations (e.g., psychiatric patients) or special circumstances (e.g., long-term care) were not included.

To obtain a sample of hospital PBOR documents, we used the U.S. News and World Report 2006 alphabetical state listing of the nation's "best hospitals"; in each state we searched the publicly available Web sites for every fourth general hospital on the list with the goal of obtaining 5 different PBOR documents from each state. We designated a document as different from other documents in the state sample if the language, excluding institutional names, was not exactly the same. In addition, documents had to be at least 300 words long to be included. This served to exclude documents that are merely advertisements or outlines of actual PBOR texts and ensured an adequate word count for readability analysis. In circumstances where multiple hospitals on the list had identical PBOR documents, we retained one copy of the PBOR and continued to search for additional documents. We continued to search the list until we found five unique documents of sufficient length per state or the list was exhausted by cycling through the list four times. All Web sites were accessed between July and August 2006.

Readability and Language Availability

Readability analyses were conducted on each hospital PBOR using three software programs; Prose: The Readability Analyst, Grammatik 6.0, and Wstyle: Writing Style Analyzer (1992).[9] For any state that designated the specific PBOR text to be presented to patients, the readability of such text was evaluated in the same fashion. In addition, for each state that designated the specific PBOR text to be presented to patients, we searched relevant Web sites for approved text in other languages.

Prose provides grade level estimates for eight readability formulas. The upper limit for most readability formulas is grade 17, which represents a 1st year graduate school reading level. Grammatik 6.0 software (1994) analyzes a text's sentence and vocabulary complexity. Wstyle categorizes writing style as Very Poor, Poor, Weak, Satisfactory, Good, Very Good, and Excellent.

Analysis of Themes

The specific rights that are protected in each state statute were abstracted and compared with the 12 themes in the 2002 version of the American Hospital Association PBOR. This process was conducted independently by two coders (MPO and DJ), who designated each AHA theme as present, present but altered, or not present. In addition, state PBOR themes not included in the AHA PBOR document were documented. Each instance of disagreement among reviewers was reevaluated in a joint conference for final classification until agreement was reached.

Protected Remedies

Any recourse delineated within the statute was abstracted. We also noted instances where the statute specifically limits a person's options to pursue legal remedies for breach of the rights delineated in the statue.

Statistical Analysis

We used the Wilcoxon signed-rank test to compare the average reading grade level of documents required by state statutes to the average reading grade level of the hospital sample in those states. The reading grade levels of PBOR documents of hospitals in states with a PBOR text defined by statute were compared to the reading grade levels for PBOR documents of hospitals in other states with use of the Wilcoxon rank sum test. All significance tests were two-tailed. Analyses were conducted with Stata version 8 (College Station, TX).

RESULTS

In two states, no relevant legislation was identified. In 25 states, PBOR laws existed exclusively for the protection of specific patient populations. Of the 23 states with PBOR legislation for general patient populations, nine states' laws presented a specific PBOR document for distribution to patients. We analyzed a total of 240 hospital PBOR documents from all 50 states; we did not find five unique hospital PBOR documents in Delaware (4), Hawaii (3), North Dakota (2), South Dakota (2) and Utah (4).

Table 2. Most Common Non-American Hospital Association Patients' Bill of Rights Themes in State Statutes and Hospital Documents

Non-AHA Themes The patient has the right to:	State Statutes (N=23), %	Hospital Documents (N=240), %
1. File a grievance	100	71
2. Examine and receive an explanation of the itemized bill regardless of source of payment	57	75
3. Respect for dignity and worth despite diagnosis	50	46
4. Visitation (and right to exclude visitors)	43	33
5. Prompt pain assessment, management, and relief	43	67
6. Have communication needs met (interpreter services, large print documents, etc.)	36	63
7. Exercise their rights without regard to sex, race, economic status, educational background, color, religion, ancestry, nation origin, sexual orientation or marital status, or the source of payment for care	29	67
8. Freedom from seclusion and restraint, unless clinically required or necessary to protect hospital staff	29	42
9. Receive care in a safe setting and help accessing protective services	21	58
10. Consideration of the ethical, cultural, spiritual, or psychosocial issues that arise in provision of care	14	33

Readability

The average reading grade level for the 240 hospital PBOR texts was 14.1 (95% confidence interval 13.9 to 14.3, range 8.2 to 17.0). The average reading grade level for each state's hospital sample of PBOR texts was 14.1 (95% confidence interval 13.8 to 14.4; range, 12.0, Maine, to 16.6, Minnesota). Nine states stipulated within their statute the actual PBOR text to be distributed to patients. The average reading grade level for these nine documents was 15.2 (95% confidence interval 13.8 to 16.7; range 11.6, New York, to 17, Minnesota) as seen in Table 3. Hospitals in these nine states rarely presented the text exactly as prescribed by state law (1 of 45). The reading grade level of hospital PBOR texts in these nine states was lower than the language specified by state law (14.7 vs. 15.2, p=0.14) and higher than the average reading grade level of hospital PBOR documents in other states (14.7 versus 14.0, p= 0.05). Table 4 presents examples of excerpts from hospital PBOR texts for four common themes.

Text Presented in Other Languages in State Statutes

In six of the nine states that present statutory PBOR texts, the state presented the mandatory text exclusively in English; three of these states presented a PBOR document in Spanish and two of these states also presented documents in additional languages (New York: Italian, Russian, Greek, Chinese, Yiddish, and Creole; Minnesota: Hmong, Somali, Russian, and Laotian).

Specific Themes

Of the 12 AHA themes, state statutes included an average of 7.4 themes and hospital documents included an average of 9.8 themes. As seen in Table 1, the AHA theme that is least commonly presented is the right to be informed of business relationships that influence care. In the 23 state statutes and the 240 hospital documents there were 95 themes not addressed in the AHA document (e.g., pain management including opiates, receiving an itemized bill, and freedom from restraints). The most common non-AHA themes are presented in Table 2.

Table 3. Readability Statistics for Patients' Bill of Rights as Codified in State Law

State	Reading Grade Level[1]	Flesch Reading Ease[2]	Sentence Complexity[3]	Vocabulary Complexity[4]	Writing Style[5]
New York	11.6	52: Fairly difficult	25	55	Satisfactory
Pennsylvania	12.9	48: Difficult	43	56	Weak
California	15.0	35: Difficult	45	67	Weak
Florida	15.2	36: Difficult	75	50	Poor
Texas	16.1	27: Very difficult	50	66	Poor
New Jersey	16.3	29: Very difficult	55	66	Very poor
Massachusetts	16.5	18: Very difficult	70	55	Poor
New Hampshire	16.6	23: Very difficult	78	58	Poor
Minnesota	17.0	15: Very difficult	84	66	Poor
Average	15.3	31: Difficult	58	60	Poor

[1] ***Reading Grade Level*** *is the average of eight readability formulas as calculated by* ***Prose: The Readability Analyst Software*** *(1988-1991)*
[2] ***Flesch Reading Ease*** *as calculated by* ***Prose: The Readability Analyst Software*** *(1988-1991).*
[3] ***Sentence Complexity*** *(100 = most complex) as calculated by* ***Grammatik 6.0 Software*** *(1994). Score is based on the number of words and clauses in a document.*
[4] ***Vocabulary Complexity*** *(100 = most complex) as calculated by* ***Grammatik 6.0 Software*** *(1994). Score is based on the number of syllables in a document and a comparison to a word list of unusual or difficult words.*
[5] ***Writing Style*** *as calculated by* ***WStyle. Writing-Style Analyzer Software*** *(1992). Score is based on: 1) Active Voice—portion of sentences using only active verbs; 2) Word economy—ratio of words that convey meaning (verbs, nouns, adjectives, and adverbs) to supporting words (propositions, articles, etc.); 3) Readability—difference between the document's readability grade and the target-reader's grade; 4) Word choice—ratio of direct, active verbs and concrete nouns to abstract nouns and verbs transformed to nouns.*

Table 4. Examples of Patients' Bill of Rights Text in Four Common Domains*

Readability Level	DOMAIN	
	Right to Refuse Care	**Right to Privacy of Records**
5th grade	*Tell us what medical care you want and what medical care you do not want.*	*We do not share your records unless you give us permission.*
8th grade	"Let you choose whether to accept or refuse treatments."	"Keep your hospital and medical records private."
12th grade	"You have the right to consent to or refuse treatment, as permitted by law, throughout your hospital stay. If you refuse a recommended treatment, you will receive other needed and available care."	"You have the right to expect that treatment records are confidential unless you have given permission to release information or reporting is required or permitted by law. When the hospital releases records to others, such as insurers, it emphasizes that the records are confidential."
16th grade	"The patient has the right to make decisions about the plan of care prior to and during the course of treatment and refuse a recommended treatment or plan of care to the extent permitted by law and hospital policy and to be informed of the medical consequences of this action. In case of such refusal, the patient is entitled to other appropriate care and services that the hospital provides or be transferred to another hospital. The hospital should notify patients of any policy that might affect patient choice within the institution."	"The patient has the right to expect that all communications and records pertaining to his/her care will be treated as confidential by the hospital, excepting cases such as suspected abuse and public health hazards when reporting is permitted or required by law. The patient has the right to expect that the hospital will emphasize the confidentiality of this information when it releases it to any other parties entitled to review information in these records."
	Right to Know Names of Providers	**Right to See Bill**
5th grade	*The doctors and nurses must tell you their names.*	*You have the right to see your bill.*
8th grade	"Tell you the names and roles of the people caring for you."	"Show you your bill and explain it to you, no matter how it is paid."
12th grade	"Be informed of the name and position of the doctor who will be in charge of your care in the hospital."	"You have the right to an examination and explanation of your bill, regardless of how it is paid."
16th grade	"Upon request, to obtain from the facility in charge of his care the name and specialty, if any, of the physician or other person responsible for his care or the coordination of his care."	"Every such patient or resident of said facility in which billing for service is applicable to such patient or resident, upon reasonable request, shall receive from a person designated by the facility an itemized bill reflecting laboratory charges, pharmaceutical charges, and third party credits and shall be allowed to examine an explanation of said bill regardless of the source of payment. This information shall also be made available to the patient's attending physician."

* *Quotations denote verbatim excerpts from hospital documents. The readability level represents the overall reading level of the document from which the excerpt was taken. Text that is not in quotations and presented in italics was written by the authors*

Recourse

Each state's statute established an internal and external grievance policy. In most of these states, complaints may be directed to the State Department of Health and in several states complaints are directed to the board of registration. For example, in Vermont complaints are directed to the board of medicine and failure to comply with any provision of the Patients' Bill of Rights law may constitute a basis for disciplinary action against a physician. In one state, Illinois, the law stipulated fines for violations and in four states (Arizona, Massachusetts, Maine, and Texas), the statute protects a private civil right of action. For example, under Texas law "A plaintiff who prevails in a suit under this section may recover actual damages, including damages for mental anguish even if an injury other than mental anguish is not shown."[10] In contrast, the Florida statute included language to explicitly restrict patients' legal options: "This section shall not be used for any purpose in any civil or administrative action and neither expands nor limits any rights or remedies provided under any other law."[11]

DISCUSSION

Our findings suggest that PBOR documents presented in U.S. hospitals far exceed the reading capacity of the majority of adults. In addition, these documents commonly fail to include themes designated by state law and by the American Hospital Association. While close to half of the states in the U.S. have Patients' Bill of Rights legislation for the general public, the specific rights named in these laws vary and few of these laws incorporate remedies other than a mechanism to file complaints. Furthermore, in nine states statutory language to be presented to patients is very complex and is usually exclusively presented in English.

These observations may not be surprising for people who know that other documents such as informed consent forms and notices of privacy protection have also been shown to be overly complex. Efforts to empower patients are undermined by legal jargon in many instances. Similarly, efforts to cultivate communication skills and inculcate the importance of patient education in trainees are hampered by the mixed message presented by patients' rights documents that patients cannot read. Students may be taught that they should care about health literacy and low English proficiency while simultaneously observing what may appear as institutional indifference in the domain of patients' rights documents.

There are several reasons why clinicians and other patient advocates should particularly care about the readability and language accessibility of PBOR documents. Patients' Bill of Rights documents are publicly presented. They are among the initial points of patient engagement. Complex public documents may serve to train patients to be more passive in their care and

may instill fear in patients with limited literacy or English proficiency. Many clinicians probably view the PBOR as a health system issue that does not directly impact clinical practice or their relationships with patients. However, a well-presented PBOR document has the capacity to encourage patient activation and trust in those providing services. The current research, which demonstrates that PBOR documents are frequently not understandable to patients, reveals a missed opportunity to present the patient care mission in a clear manner.

In the 1970s, the patients' rights movement was advanced because physicians were perceived as too powerful.[12] At that time, patients had to advocate for the right to be given information about their diagnosis and prognosis.[13] By the 1990s, when the concept of a patients' bill of rights was introduced in Congress, the topic was advanced by a consumer rights movement due to a sense that managed care companies and insurers were too powerful.[14] Instead of protecting a right to refuse treatment from paternalist physicians, consumers wanted to secure a right to choose their providers and have access to treatments being denied by payors.

The American Hospital Association, which has long been an advocate for a patients' bill of rights, changed their format in 2006 to a brochure called "The Patient Care Partnership," which contains the same themes and "informs patients about what they should expect during their hospital stay with regard to their rights."[15] While the brochure is a clear departure from the legal jargon of prior PBOR documents advanced by the American Hospital Association (and is presented on their Web site in Arabic, Chinese, English, Russian, Spanish, Tagalog, and Vietnamese), the English text is still written at an 11th grade reading level.

As seen in Table 4, where we present examples written at a 5th grade level, the themes of the PBOR can be written in plain English. In most states, hospitals are free to revise their PBOR documents; however, in nine states (CA, FL, MA, MN, NH, NJ, NY, PA and TX) statutes should be amended either to allow hospitals to write their own language or to present the official state PBOR in plain English. A note of caution is warranted. According to Robert Gunning, developer of the Fog readability formula: "Like all good inventions, readability yardsticks can cause harm in misuse. They are handy statistical tools to measure complexity in prose…*But they are not formulas for writing.*"[16] Authors who replace long words with short words that are similarly arcane have not improved the actual readability, even if they do reduce their readability score.[17,18]

Different formulas report grade levels that vary by two to four grades, partly because they are based on different levels of reader comprehension. Because the SMOG formula is based on 100% reader comprehension, it tends to score higher than other formulas which are based on 35%—70% reader comprehension. Rather than using a single formula that might bias the results by scoring "high" or "low," we used Prose software because it provides the average grade level estimates of eight readability formulas. In addition, we provide further analyses to exhibit the level of complexity of the PBOR documents.

There are limitations to readability software programs. First, the same formula in different programs may give different grade levels due to variations in algorithms used to count sentences and syllables.[17] Second, formulas do not take into account a PBOR's organization, font size, font family, etc. Third, these formulas cannot account for the background knowledge of the readers, their motivation, cultural experiences, etc. Despite these limitations, the formulas do provide a reasonable and cost-effective way of assessing how clearly PBORs are written.

Interested hospitals and legislatures may benefit from consulting specialists in adult basic education, readability, and improving patient care systems in this process. Patients and their advocates can also play an important role. In addition, plain language versions in other languages should be commissioned. Similarly, hospitals can improve patients' comprehension of their rights by supplementing their print material with other educational methods such as video or interactive multimedia that can be developed. A promising proposal for a National Health Literacy Act, to establish a national center for health literacy at the Agency for Healthcare Research and Quality as well as provide funding for State Health Literacy offices, is currently being vetted.[19] Resources of this kind could help avoid future instances of legislatures compelling hospitals to present unreadable legal jargon to patients.

The strengths of this study that lend weight to our conclusions are the amount of text analyzed, the blind sampling within every state, and the complete evaluation of state statutes. Nonetheless, several limitations should be kept in mind. First, we surveyed only hospital PBOR texts that were available through institutional Web sites. Although it is likely that the materials presented on institutional Web sites accurately reflect local practices, additional materials were not examined. Second, we did not attempt to evaluate the conceptual complexity of the content. It is possible that variations in conceptual complexity influence readability as well. Third, we evaluated readability using the average of eight readability formulas and three measures of syntax and semantics: sentence complexity, vocabulary complexity, and writing style. While this represents a significant advance over the vast majority of published analyses which are based simply on the Flesch–Kincaid scale, or other single metrics of readability, additional factors that affect legibility and understandability, such as the type font, layout, and length, were not evaluated in this project. Similarly, we were not able to evaluate the readability of PBOR documents in languages other than English to determine, for example, if the Minnesota State PBOR, which is at a graduate school level in English, is also at a 17th grade level in Hmong, Somali, Russian and Laotian. Fourth, we report the remedies offered within statutes; however, this does not reflect the volume or types of complaints that these statutes have actually generated. We made multiple attempts to determine details of these programs, but were not able to obtain records on complaints or otherwise assess the consequences of PBOR statutes. It would be valuable to know how patients and states use these programs.[20]

When a hospital PBOR document is missing a theme that is recommended by the AHA or required by state statute, it is unclear if this represents an accidental lapse or a purposeful departure. The absence of themes from PBOR documents, however, does not change clinical standards. For example, the least common AHA PBOR theme presented in hospital documents and state statutes relates to the disclosure of business relationships that may influence care. Nonetheless, professional standards dictate disclosure of such relationships.[21]

Promoting patients' rights has had many years of regulatory support from the AHA and the Joint Commission. Similarly, almost half the states in our country have shown legislative support for a bill of rights to protect all patients. These laws do

not establish a right to health care. Yet, patients' rights statutes are designed to promote the ethical and humane treatment of patients. These goals will not be realized by presenting patients with documents they are not able to read and understand.

Acknowledgement: *This study was funded by the Pfizer Clear Health Communication Initiative.*

Conflict of Interest Statement: *Dan M. Jacob is an employee of Healthcare Analytics, LLC, New York, NY and Mark Hochhauser is a Readability Consultant, Golden Valley, MN. The other authors have no potential conflicts of interest to report.*

Corresponding Author: *Michael K. Paasche-Orlow; Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA (e-mail: mpo@bu.edu).*

REFERENCES

1. **Mariner WK.** What recourse? Liability for managed-care decisions and the Employee Retirement Income Security Act. N Engl J Med. 2000;343 (8):592–96.
2. **Binette MJ.** Patients' bill of rights: legislative cure-all or prescription for disaster? North Carol Law Rev. 2003;81:653–96.
3. **Blaes P.** Promotion and enforcement of patients' rights. Med Law. 2004;23(2):289–97.
4. **Annas GJ.** A national bill of patients' rights. N Engl J Med. 1998;338 (10):695–9.
5. **Monaghan JC.** Whatever happened to the patient's bill of rights? Med Econ. 1975;52(16):109–10.
6. **Annas GJ.** A.H.A. Bill of Rights. Trial. 1973;9(6):59–61.
7. **Paasche-Orlow MK, Jacob DM. Powell JN.** Notices of Privacy Practices: a survey of the Health Insurance Portability and Accountability Act of 1996 documents presented to patients at US hospitals. Med Care. 2005;43(6):558–64.
8. **Paasche-Orlow MK, Taylor HA, Brancati FL.** Readability standards for informed-consent forms as compared with actual readability. N Engl J Med. 2003;348(8):721–6.
9. MicroBrothers Software. Prose: The readability analyst. Reference Manual. 1988. Boulder Colorado.
10. Patient Rights. Tx Stat. title 25, §133.42.(2007).
11. Florida Patient's Bill of Rights and Responsibilities. Fla. Stat. title 29, § 381.026 (2007).
12. **Annas GJ.** The Rights of Hospital Patients: The Basic ACLU Guide to a Hospital Patient's Rights. New York: Discus Books; 1975.
13. **Veatch RM.** Bioethics discovers the Bill of Rights. Natl Forum. 1989;69 (4):11–3.
14. **Rosenbaum S.** Managed care and patients' rights. JAMA. 2003;289 (7):906–7.
15. "The Patient Care Partnership," (Accessed December, 23, 2008, at http://www.aha.org/aha/issues/Communicating-With-Patients/pt-carepartnership.html).
16. **Gunning R.** The Technique of Clear Writing. New York: McGraw Hill; 1952:29.
17. **Hochhauser M.** Some overlooked aspects of consent form readability. IRB. 1997;19(5):5–9.
18. **Hochhauser M.** Informed consent and patient's rights documents: a right, a rite, or a rewrite? Ethics Behav. 1999;9(1):1–20.
19. **Coleman N.** "A bill to ensure that all Americans have basic health literacy skills to function effectively as patients and health care consumers." S.2424 (Accessed December, 23, 2008, at http://www.govtrack.us/congress/bill.xpd?bill=s110-2424).
20. **Zaremski MJ.** Patients rights and accountability: can there exist rights without remedies in an American legal and legislative framework? Med Law. 2003;22(3):429–50.
21. **Coyle SL.** Physician-industry relations. Part 1: individual physicians. Ann Intern Med. 2002;136(5):396–402.

Patient Education and Counseling 75 (2009) 293–294

Contents lists available at ScienceDirect

Patient Education and Counseling

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

Editorial

Bridging the International Divide for Health Literacy Research

Over the past few decades, *Patient Education and Counseling* has been one of the leading journals that has supported the emerging field of health literacy. This is quite fitting, as the research has included a multidisciplinary assortment of investigators concerned with the journal's core aspirations: improving the communication between patients and their providers and developing methods for patients to be more educated and activated about their health. This agenda grows out of an intuition that views health care as unnecessarily complicated and insufficiently dedicated to the mission of patient education and patient empowerment. Thus, it is truly an honor to present this special issue addressing international perspectives on health literacy research as part of the 75th volume anniversary of the journal.

The collection of papers we present in this special issue reveal both the strides being taken to remediate the problem of limited health literacy as well as the great growth potential that remains for this field of inquiry. We specifically have highlighted through the article assortment how health literacy has become an international phenomenon. Though the research is still dominated by work being done in the Unites States, the issue includes articles from Australia, Great Britain, Japan, and Korea. The expansion of health literacy research to Asia is particularly notable as the groundwork is now clearly being laid to extend the field into populations that communicate with character-based languages.

We have also included material from a broad range of content and methods. There are three papers that focus on measurement: Shapira et al. evaluate and refine prior tests of health numeracy; Lee et al. present data for a new Korean Health Literacy Scale; and Yost et al., describe a new English and Spanish computer-based instrument. There are also four papers that evaluate innovative intervention strategies to address literacy barriers in health communication. Bickmore et al. present findings on the efficacy of an animated computer avatar system, while Kandula et al. review the development and testing of a multimedia diabetes education program for use in community health centers. Wallace et al, evaluate another diabetes self-management intervention that combined a clinical practice protocol with literacy-appropriate patient education materials and a brief counseling strategy. Lastly, Rudd et al. describe results of using literacy-appropriate patient education materials with brief counseling for patients with arthritis. The work by Clement et al. appropriately complements these papers through a systematic review of complex health literacy interventions.

The extent, association, and implications of limited literacy on health outcomes are also further explored by many of the authors in this issue as well. There are two papers that investigate the role of literacy in colorectal cancer screening programs (Von Wagner et al. and Smith et al.), while the impact parental literacy may have on children's health is examined by two research teams (Shone et al. and Hironka et al.) in the context of asthma and medication adherence. Pandit et al. determine the nature of the relationship between literacy, education, and hypertension outcomes; Apter et al. present data on the role of health literacy on quality of life in the context of patients with asthma. The complexity of oral communication in the context of genetic counseling is explored in an innovative fashion by Roter et al.

Two articles recognize the additional barrier of language in the context of health literacy. Sudore et al. compare the impact of literacy and limited English proficiency in healthcare, and Fang and colleagues reveal alarming health literacy and language barriers for basic concepts relating to stroke among patients in an anticoagulation clinic. Finally, Tokuda et al. present the first evaluation of the prevalence of health literacy barriers in Japan and show that health literacy is linked to health-related quality of life.

Close scrutiny of the articles reveals that much work remains to further ground and unite the field across international borders. To begin, there appears to be significant variation in definitions. Some authors avoid the term 'health literacy' when discussing basic literacy skills in a health care context while others are clearly comfortable with this label. The most commonly used tools to measure health literacy among the papers in this special issue are versions of the Rapid Estimate of Adult Literacy in Medicine (REALM) and the Test of Functional Health Literacy Assessment (TOFHLA). In fact, neither of these instruments are tests of health literacy as it is typically defined. The REALM is a word recognition test and depending on which version of the TOFHLA was used, there are aspects of both reading fluency and numeracy present. Both of these tests were designed from general literacy measures and more accurately should be regarded as assessing basic literacy skills framed within a healthcare context. Basic literacy skills such as these are core components of any definition of health literacy, but authors disagree about how to discuss results of studies using these instruments. Two of the papers in this issue address the topic of definitions in the context of developing new measurement tools. Both Lee et al. who presents data for the Korean Health Literacy Scale, and Yost et al. who presents data on an English and Spanish computer-based instrument, developed a broad array of test stimuli to move beyond measuring basic literacy skills. These authors clearly aspire to capture a more expansive health literacy concept. What remains to be seen is whether such measurement is worthwhile. The literature in health literacy, after all, has been substantiated on data derived from studies that have used crude instruments like the REALM and TOFHLA, so the relationship between other domains of health literacy (that might be defined on the basis of new instruments) and health outcomes remains to be

0738-3991/$ – see front matter
doi:10.1016/j.pec.2009.05.001

seen. Regardless, existing tools have been highly predictive for clearly identifying an at-risk population.

Even further, the term health literacy is increasingly referred to as a broader public health concept in both the United States and internationally. Clearly, the large amount of interest in helping patients and families access, understand, and use health information allows for such diffusion. Those in the health literacy field must be open to a growing duality that research in this line of inquiry can refer to the study of (1) the knowledge and skills a person needs (e.g., to gather, understand, and comply with medical instructions), and (2) the preparation and outreach that must be undertaken by health systems and other relevant institutions to convey roles, responsibilities, and information within healthcare and support their ability to perform relevant tasks. The former addresses an individual cognitive and psychosocial skill set, while the latter targets attributes of the health system. Both are appropriate foci for health literacy research and despite the broad array of measures and intervention targets that fall within this construct, this should not be viewed as problematic. Health literacy embodies the goals of clear health communication and patient engagement in healthcare, requiring a diverse worldview and multidisciplinary perspective.

We hope this issue generates further global interest in health literacy and recognition of its importance in the context not only for advancing clear health communication, but also, for promoting healthcare equity, quality, and safety. An identifiable marker of the journal's success will be to see future publications from even more countries seeking to address the problem. As *Patient Education and Counseling* has an international audience, it is quite befitting for it to continue to serve as a preeminent venue for advancing the field among diverse populations and health systems. We thank the editors, reviewers, and publication staff for supporting this issue, and the authors for sharing their laudable work.

Michael K. Paasche-Orlow[a,*], Kirsten McCaffery[b]
Michael S. Wolf[c]

[a]*Department of Medicine, Boston University School of Medicine, 801 Massachusetts Avenue, 2nd floor, Boston, MA 02118, USA*
[b]*School of Public Health, University of Sydney, Edward Ford Building, Sydney, NSW 2006, Australia*
[c]*Institute for Healthcare Studies, Northwestern University, Chicago, IL 60611, USA*

*Corresponding author. Tel.: +1 617 414 5877; fax: +1 617 414 4676
E-mail address: mpo@bu.edu (M.K. Paasche-Orlow)

Long-Term Trends in Myocardial Infarction Incidence and Case Fatality in the National Heart, Lung, and Blood Institute's Framingham Heart Study

Nisha I. Parikh, MD, MPH; Philimon Gona, PhD; Martin G. Larson, ScD; Caroline S. Fox, MD, MPH; Emelia J. Benjamin, MD, ScM; Joanne M. Murabito, MD, ScM; Christopher J. O'Donnell, MD, MPH; Ramachandran S. Vasan, MD; Daniel Levy, MD

Background—Whereas the prevalence of coronary heart disease risk factors has declined over the past decades in the United States, acute myocardial infarction (AMI) rates have been steady. We hypothesized that this paradox is due partly to the advent of increasingly sensitive biomarkers for AMI diagnosis.

Methods and Results—In Framingham Heart Study participants over 4 decades, we compared the incidence and survival rates of initial AMI diagnosis by ECG (AMI-ECG) regardless of biomarkers with those based exclusively on infarction biomarkers (AMI-marker). We used Poisson regression to calculate annual incidence rates of first AMI over 4 decades (1960 to 1969, 1970 to 1979, 1980 to 1989, and 1990 to 1999) and compared rates of AMI-ECG with rates of AMI-marker. Cox proportional-hazards analysis was used to compare AMI case fatality over 4 decades. In 9824 persons (54% women; follow-up, 212 539 person-years; age, 40 to 89 years), 941 AMIs occurred, including 639 AMI-ECG and 302 AMI-marker events. From 1960 to 1999, rates of AMI-ECG declined by ≈50% and rates of AMI-marker increased ≈2-fold. Crude 30-day, 1-year, and 5-year case fatality rates in 1960 to 1969 and 1990 to 1999 were 0.20 and 0.14, 0.24 and 0.21, and 0.45 and 0.41, respectively. Age- and sex-adjusted 30-day, 1-year, and 5-year AMI case fatality declined by 60% in 1960 to 1999 (P for trend <0.001), with parallel declines noted after AMI-ECG and AMI-marker.

Conclusions—Over the past 40 years, rates of AMI-ECG have declined by 50%, whereas rates of AMI-marker have doubled. Our findings offer an explanation for the apparently steady national AMI rates in the face of improvements in primary prevention. **(*Circulation*. 2009;119:1203-1210.)**

Key Words: biomarkers ■ electrocardiography ■ epidemiology ■ myocardial infarction

During the past 4 decades, death rates from coronary heart disease (CHD) have declined by >60%.[1–4] Between nearly one half to upwards of three quarters of the decline in CHD mortality has been attributed to improvements in primary prevention and risk factor modification.[5–10] Awareness, treatment, and control of 3 key risk factors—hypertension, hypercholesterolemia, and smoking—have improved in recent decades.[1,11] Despite these improvements, hospitalization rates for acute myocardial infarction (AMI) have remained relatively stable over the past 5 decades.[1,4,12] The reasons for the paradoxical stability of AMI rates in the face of declining CHD risk factor prevalence are not clear.

Editorial p 1189
Clinical Perspective p 1210

Whereas ECG criteria for AMI have not changed appreciably over the past 50 years, several different biomarkers of varying sensitivity and specificity have been introduced for the detection of AMI. Early on, serum glutamic oxalacetic transaminase and lactic dehydrogenase were used, in conjunction with clinical information, to diagnose AMI. In more recent times, serum markers of myocardial cell damage, including creatine phosphokinase (CPK), lactic dehydrogenase isoenzymes, CPK-MB, and troponin, have been introduced sequentially to diagnose AMI and have been firmly

Received September 30, 2008; accepted December 18, 2008.

From the National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Mass (N.I.P., P.G., M.G.L., C.S.F., E.J.B., J.M., C.J.O., R.S.V., D.L.); Beth Israel Deaconess Medical Center, Cardiovascular Division, Boston, Mass (N.I.P.); Boston University, Boston, Mass (N.I.P., P.G., M.G.L.); Brigham and Women's Hospital, Harvard Medical School, Boston, Mass (C.S.F.); Cardiology, Section of General Internal Medicine and Preventive Medicine Sections, Boston University School of Medicine, Boston, Mass (E.J.B., J.M., R.S.V., D.L.); Division of Cardiology, Massachusetts General Hospital, Boston (C.J.O.); and Center for Population Studies of the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Md (C.J.O., C.S.F., D.L.).

Guest Editor for this article was Philippe Gabriel Steg, MD.

Correspondence to Daniel Levy, MD, Framingham Heart Study, 73 Mount Wayte Ave, Suite 2, Framingham, MA 01702–5803. E-mail levyd@nih.gov

Circulation **is available at http://circ.ahajournals.org** **DOI: 10.1161/CIRCULATIONAHA.108.825364**

incorporated into international guidelines for AMI case definition.[13] Compared with diagnosis based solely on history and ECG, AMI diagnosis based on serial biomarker measurements has substantially increased the detection of AMI cases.[14–16]

Previous investigations of US trends in AMI incidence comparing different diagnostic criteria have been hospital based and have encompassed limited time periods for their analysis.[14,17] The Framingham Heart Study, which has >50 years of physician-validated AMI data on a community-based cohort, offers a unique setting to study trends in AMI incidence and case fatality rates based on the following AMI diagnostic criteria: AMI by ECG diagnosis (AMI-ECG) regardless of biomarker elevation, which offers an unbiased assessment of long-term trends, and AMI by biomarker diagnosis (AMI-marker) in the absence of diagnostic ECG changes, which reflects changing methods in clinical practice. The sum of these 2 mutually exclusive approaches represents total AMI.

We hypothesized a priori that rates of AMI-ECG have declined in the long term (consistent with improvements in CHD risk factors) while the rates of AMI-marker have increased (owing to greater biomarker sensitivity), resulting in a relatively steady rate of total AMI incidence over a 40-year time interval. Accordingly, we analyzed 40-year trends in the incidence of first AMI and for the 2 mutually exclusive AMI subgroups of AMI-ECG and AMI-marker. Such an approach may shed light on the paradoxical stability of national AMI rates in the setting of improvements in CHD risk factors and declining rates of CHD mortality.

Secondarily, we assessed time-period trends in mortality after AMI and its subcomponents, AMI-ECG and AMI-marker. This analysis will help promote understanding of the relative effectiveness of secondary prevention efforts over time when considered in conjunction with analyses of time-period changes in the incidence of initial AMI, which reflect advances in primary prevention.

Methods

The Framingham Heart Study is a community-based prospective observational study that began in 1948, enrolling 5209 men and women in the original study cohort.[18] Original cohort members attended clinic examinations approximately every 2 years. In 1971, 5124 men and women enrolled in the Framingham Heart Study offspring cohort, which included the children and spouses of the children of the original cohort. Participant examinations for the offspring cohort occurred approximately every 4 to 8 years; the design and methodology have been described elsewhere.[19] This investigation included original and offspring cohort members.

We considered all original and offspring cohort members 40 to 89 years of age who were free of AMI (recognized and unrecognized) at their first Framingham clinic examination in each decade of study (1960s, 1970s, 1980s, 1990s). Our final sample size consisted of 9824 individuals. Each individual could enter the sample multiple times on the basis of eligibility for time period and age group. For example, a participant 35 years of age in 1960 would not contribute follow-up time to the first time period until he or she turned 40 in 1965, thereafter contributing 5 years. That participant would contribute 5 years to second time period and so on until the patient died or developed AMI. Similarly, a patient 75 years old in 1960 contributed at most 5 years to the last period. Participants provided written informed consent, and the study protocol was approved by the Boston University Medical Center Institutional Review Board.

Table 1. Characteristics of Framingham Heart Study Participants at the Start of Each Time Period of Study

	1960s	1970s	1980s	1990s
Men, n	1768	2205	2145	2147
Women, n	2366	2687	2662	2632
Mean age, y	53±8	56±10	60±12	60±13
Women, %	57	55	55	55
Total cholesterol, mg/dL	248±45	226±43	221±41	210±39
Systolic BP, mm Hg	136±23	136±22	133±20	133±21
Diastolic BP, mm Hg	84±12	82±11	79±10	78±10
Body mass index, kg/m^2	25.8±4.1	26.5±4.3	26.6±4.4	27.0±4.8
Glucose, mg/dL	82±23	106±18	95±39	99±30
Hypertension, %	43	47	47	48
Diabetes mellitus, %	4	8	10	8
Smoking, %	53	28	30	35

BP indicates blood pressure. Values are mean±SD when appropriate.

Risk Factor Assessment

At each routine clinic visit, participants underwent physical examination, 12-lead ECG, anthropometry, and laboratory assessment of vascular risk factors. Details on the ascertainment of risk factors have been previously described.[19] Participants with systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg (mean reading of 2 readings taken by an examining physician) or receiving medication for the treatment of hypertension were defined as having hypertension. Plasma glucose and total cholesterol were measured. Diabetes mellitus was defined (throughout the study period) as fasting plasma glucose ≥126 mg/dL, a nonfasting glucose of ≥200 mg/dL, or treatment with either insulin or hypoglycemic agents. Participants were considered to be current smokers if they smoked on average at least 1 cigarette per day during the year before examination.

Serum Biomarkers of MI

Several serum biomarkers were used for AMI diagnosis during the study time period. Specific diagnostic biomarkers and the decades during which they were used for AMI diagnosis in the Framingham Heart Study included the following: serum glutamic oxalacetic transaminase beginning in the mid-1950s, lactic dehydrogenase in the 1960s, CPK in the 1970s, CPK-MB and lactic dehydrogenase isoenzyme in the 1980s, and troponin in the late 1990s. We did not use prespecified cut points to determine biomarker elevation because variability was present in assays used in the various hospitals from which medical records were collected. Thus, we considered a biomarker elevated if it exceeded the reference limit provided by the hospital laboratory report at the time of AMI hospitalization, according to the available medical record/chart.

Ascertainment of AMI and AMI Case Fatality

Framingham Heart Study participants are under continuous surveillance for cardiovascular disease events and death. The surveillance process included physician-administered questions about cardiovascular events during each routine follow-up Framingham Heart Study clinic visit and a mailed health history update questionnaire (which, before the late 1990s, consisted of a brief questionnaire for those who had not attended examinations and, after the late 1990s, included detailed sections about interim cardiac events and hospitalizations). If a participant reported a possible interim event, all pertinent medical records were collected and reviewed by an events adjudication committee consisting of 3 physicians who reviewed all available hospitalization records, physician office visit notes, and pathology reports.[20] AMIs were diagnosed on the basis of ischemic

Table 2. Decade-Specific Incidence Rates of Overall AMI, AMI-ECG, and AMI-Marker per 10 000 Person-Years Among Men

	1960–1969	1970–1979	1980–1989	1990–1999
Overall AMI				
Age, y				
40–49	43.80	22.05	29.43	24.71
50–59	62.62	53.86	59.75	39.78
69–69	74.05	84.33	97.90	55.11
70–79	152.17	83.84	135.56	117.36
80–89	*	98.09	129.90	166.03
Events, n	130	143	190	144
AMI-ECG				
Age, y				
40–49	39.39	18.17	23.75	15.96
50–59	53.50	40.63	43.80	21.67
69–69	63.25	63.59	71.74	30.01
70–79	121.15	56.31	87.65	52.05
80–89	*	66.69	85.10	75.20
Events, n	112	105	131	72
AMI-marker				
Age				
40–49	4.41	3.88	5.68	8.75
50–59	9.12	13.23	15.94	18.11
69–69	10.80	20.73	26.15	25.11
70–79	31.01	27.53	47.92	65.31
80–89	*	31.40	44.79	90.84
Events, n	18	38	59	72

Person-years of observation: 122 560.
*No events for age group/time period.

Table 3. Decade-Specific Incidence Rates of Overall AMI, AMI-ECG, and AMI-Marker per 10 000 Person-Years Among Women

	1960–1969	1970–1979	1980–1989	1990–1999
Overall AMI				
Age, y				
40–49	6.17	3.96	4.18	3.99
50–59	10.34	11.35	9.95	7.53
69–69	23.51	34.16	31.35	20.05
70–79	51.17	35.96	45.96	45.21
80–89	*	85.53	89.53	130.00
Events, n	47	81	100	106
AMI-ECG				
Age, y				
40–49	5.55	3.26	3.37	2.57
50–59	8.84	8.56	7.29	4.10
69–69	20.09	25.76	22.97	10.91
70–79	40.74	24.15	29.72	20.05
80–89	*	58.15	58.66	58.88
Events, n	38	59	72	50
AMI-marker				
Age				
40–49	0.62	0.70	0.81	1.4
50–59	1.51	2.79	2.66	3.43
69–69	3.43	8.40	8.37	9.13
70–79	10.43	11.81	16.25	25.16
80–89	*	27.38	30.87	71.13
Events, n	9	22	28	56

Person-years of observation: 89 979.
*No events for age group/time period.

chest discomfort with diagnostic ECG changes (based on chart review) with or without diagnostic biomarker changes (AMI-ECG) or ischemic chest discomfort with diagnostic serum biomarkers of infarction but without diagnostic ECG changes (AMI-marker). ECG criteria for AMI included development of pathological Q waves of ≥0.04 seconds, often accompanied by ST elevation and followed by serial changes indicating a reversion of these ECG changes toward normal. We chose to exclude persons with unrecognized/silent AMI because it is impossible to determine the exact date of occurrence, which is assigned a midpoint between the last ECG without an abnormality and the first one manifesting Q-wave changes.

Case fatality was assessed within 30 days and at 1 and 5 years. For 1- and 5-year mortality, deaths occurring within the first 30 days were excluded from analysis. We did this to obtain a truer sense of how many "later" case fatalities occurred after AMI (because a large proportion of post-AMI deaths occur within 30 days of the index event as opposed to later). Furthermore, pathophysiologically "early" death resulting from AMI is likely different from "later" deaths.

Statistical Methods

Prevalence rates and means (±SD) of cardiovascular disease risk factors were calculated for the study sample at the first examination cycle in each decade of study. We used Poisson regression to calculate annual incidence rates of first AMI over 4 time periods (1960 to 1969, 1970 to 1979, 1980 to 1989, and 1990 to 1999) and compared rates of AMI-ECG with rates of AMI-marker. We tested for sex×age group, sex×time period, age group×time period, age group×AMI type, and time period×AMI type interactions for incidence rate trends; given multiple statistically significant probability values for these interactions, we present age- and sex-specific AMI incidence rates for each time period. Additionally, with small numbers of events for the oldest and youngest age groups, we provide trends in (log-transformed) event rates across the 4 time periods for the age groups of 50 to 59, 60 to 69, and 70 to 79 years for men and women separately. We calculated tests of trend for overall AMI, AMI-ECG, and AMI-marker across time periods, with the 1960s serving as the referent decade (using a model accounting for the interactions listed above). We used Cox proportional-hazards models to calculate age- and sex-adjusted case fatality curves and 30-day, 1-year and 5-year case fatality rates after all AMI, AMI-ECG, and AMI-marker for each of the 4 time periods (with 1960 to 1969 serving as the referent period). The follow-up period for case fatality was until the end of 2006. The assumption of proportionality of hazards was satisfied over the 5-year follow-up period after AMI (P for time to death×period interaction >0.32 for overall AMI, AMI-ECG, and AMI-marker). A 2-sided value of $P<0.05$ was considered to indicate statistical significance. All statistical analyses were performed with the use of the SAS statistical software (version 9.0; SAS Institute, Inc, Cary, NC).

Dr Levy had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors have read and agree to the manuscript as written.

Results

Characteristics of Study Sample

Study participant characteristics by decade are shown in Table 1. Of the 9824 participants, 54% were women;

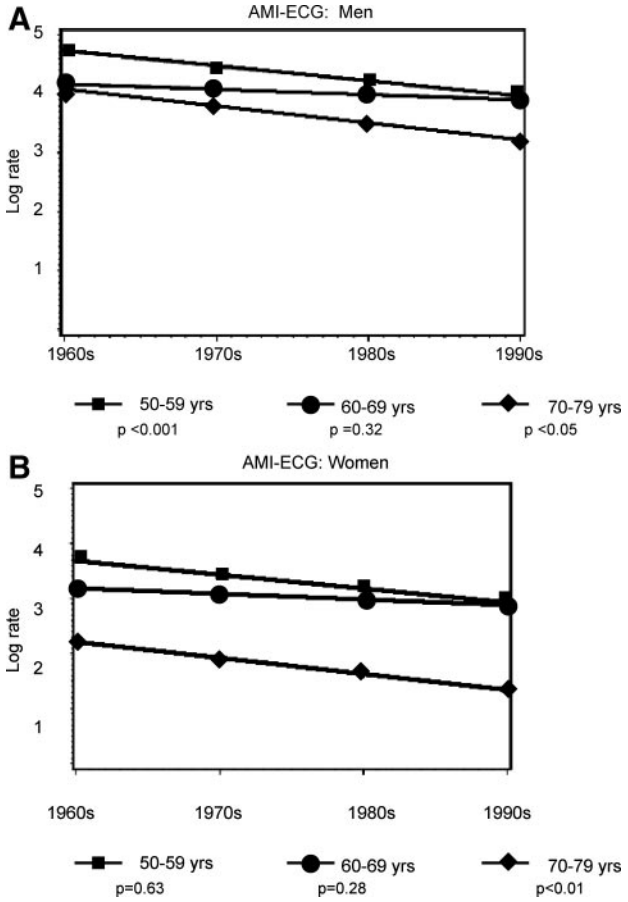


Figure 1. Temporal trends for age-range–specific incidence rates in AMI-ECG from 1960 to 1999 among men (A) and women (B).

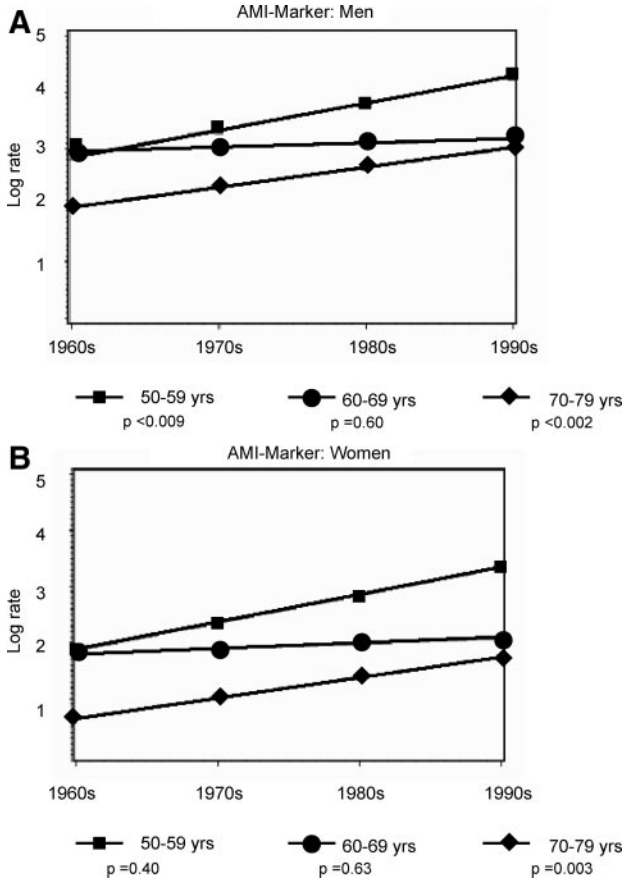


Figure 2. Temporal trends for age-range–specific incidence rates in AMI-marker from 1960 to 1999 among men (A) and women (B).

follow-up time was 212 539 person-years. The mean age of participants at the start of each time period ranged from 53 years in the 1960s to 60 years in 1990s. Smoking rates, total cholesterol concentrations, and systolic and diastolic blood pressures decreased from 1960 to 1999.

Trends in Overall AMI, AMI-ECG, and AMI-Marker Rates

Overall, 941 first AMIs occurred, including 639 AMI-ECG events (68%) and 302 AMI-marker events (32%). Age- and sex-specific incidence rate trends from the 1960s to the 1990s are presented in Tables 2 and 3 for men and women, respectively.

Rates of AMI-ECG declined by ≈50% and rates of AMI-marker doubled over the study period (Tables 2 and 3, Figures 1 and 2). Among men, statistically significant declines in AMI-ECG were noted in the age groups of 50 to 59 years (P for trend <0.001) and 70 to 79 years (P for trend <0.05) (Figure 1A). In women, statistically significant declines in AMI-ECG were noted among those 70 to 79 years of age (P for trend <0.01) (Figure 1B). Among men, statistically significant increases in AMI-marker were noted in those 50 to 59 and 70 to 79 years of age (P for trend <0.01 for both) (Figure 2A); in women, statistically significant increases in AMI-marker were noted among those 70 to 79 years of age (P for trend <0.01) (Figure 1B). Trends for AMI-ECG and AMI-marker were largely flat for the 60-to 69-year-old group.

Trends in 30-Day, 1-Year, and 5-Year AMI Case Fatality

Five-year case fatality rates after overall AMI decreased steadily from 1960 to 1999 (P for trend <0.001) (Table 4, Figure 3A). Trends in 5-year case fatality rates after AMI-ECG and after AMI-marker mirrored overall 5-year AMI case fatality trends (Figure 3B and 3C). Similarly, decreases were seen in 30-day and 1-year case fatality rates after overall AMI, AMI-ECG, and AMI-marker (Table 4). A particularly large shift was found toward decreased case fatality between the 1970s and 1980s (Figure 3A through 3C).

Discussion

Principal Findings

In a community-based cohort of 9824 men and women followed up for a 4-decade interval, we found that AMI-ECG rates declined ≈50% with a concomitant 2-fold increase in rates of AMI-marker. The 30-day, 1-year, and 5-year case fatality rates after overall AMI declined by 50% to 75% from 1960 to 1999, with parallel declines in case fatality after both AMI-ECG and AMI-marker over this period. We conclude that national MI trend data may be biased by a diagnostic drift resulting from the advent of diagnostic biomarker tests for

Table 4. Age-and Sex-Adjusted 30-Day, 1-Year, and 5-Year Mortality Rates for Overall AMI, AMI-ECG, and AMI-Marker Among Framingham Heart Study Participants

Outcome	1960–1969	1970–1979	1980–1989	1990–1999	*P*, Trend Test
At 30 d					
All AMI					
Deaths, n	35	45	47	34	
HR (95% CI)	Referent	0.66 (0.42–1.05)	0.45 (0.28–0.71)	0.27 (0.16–0.45)	<0.0001
AMI-ECG					
Deaths, n	32	38	35	23	
HR (95% CI)	Referent	0.71 (0.44–1.16)	0.50 (0.30–0.82)	0.38 (0.21–0.69)	0.0004
AMI-marker					
Deaths, n	3	7	12	11	
HR (95% CI)	Referent	0.72 (0.18–2.81)	0.46 (0.12–1.82)	0.22 (0.06–0.89)	0.006
At 1 y					
All AMI					
Deaths, n	(42)	(66)	(73)	(54)	
HR (95% CI)	Referent	0.83 (0.56–1.23)	0.58 (0.39–0.87)	0.35 (0.22–0.54)	<0.0001
AMI-ECG					
Deaths, n	37	49	55	29	
HR (95% CI)	Referent	0.80 (0.52–1.25)	0.68 (0.44–1.05)	0.42 (0.25–0.71)	<0.001
AMI-marker					
Deaths, n	5	17	18	25	
HR (95% CI)	Referent	1.11 (0.41–3.04)	0.43 (0.15–1.22)	0.31 (0.11–0.87)	<0.001
At 5 y					
All AMI					
Deaths, n	80	104	111	104	
HR (95% CI)	Referent	0.73 (0.54–0.98)	0.47 (0.35–0.64)	0.36 (0.26–0.50)	<0.006
AMI-ECG					
Deaths, n	70	78	72	46	
HR (95% CI)	Referent	0.70 (0.50–0.98)	0.47 (0.33–0.66)	0.36 (0.24–0.53)	<0.001
AMI-marker					
Deaths, n	10	26	39	58	
HR (95% CI)	Referent	1.06 (0.51–2.21)	0.59 (0.28–1.22)	0.45 (0.22–0.93)	0.001

HR indicates hazard ratio.

AMI; we were able to identify and quantify the possible magnitude of this effect within our study setting. These findings may explain the paradoxical stability of AMI rates in the United States despite concomitant improvements in CHD risk factors.

Temporal Trends in AMI

Several epidemiological studies conducted in United States have demonstrated steady rates of AMI from the 1970s to the 1990s,[4,12,17,21] whereas data from the World Health Organization Monitoring Trends and Determinants in Cardiovascular Disease (WHO-MONICA) project demonstrated modest declines in rates of AMI from 1985 to 1991.[22] Data from the Worcester Heart Attack Study similarly demonstrated modest declines in the incidence of first AMI.[23] Differences in study design and event ascertainment may have accounted for the differing results between these prior studies. In our study, particularly in men, overall AMI trends appear to be decreasing in a parallel fashion compared with AMI-ECG. In women, overall AMI rates were steady to decreased.

Defining AMI in population studies and clinical research is essential for accurate disease surveillance, clinical trial design and conduct, and healthcare resource allocation.[13,24,25] Several prior studies demonstrating trends in AMI rates have used international diagnostic codes (*International Classification of Diseases* [ICD]) for hospital discharges to identify AMI cases[21,26–28] and may be subject to "diagnostic drift".[29] Specifically, diagnostic coding of AMI during hospitalizations may have increased as a result of changes in reimbursement practices and by the use of more sensitive biomarkers of infarction.[24,29] Temporal-trend estimates of AMI based on ICD codes have shown steady rates[4,21] over the past several decades. In contrast, AMI-ECG rates in our study sample declined by ≈50% from 1960 to 1999. AMI-ECG represents a relatively "unbiased" estimate of AMI that has not been influenced by the advent of increasingly sensitive biomarkers of infarction in recent decades.[17] Not surprisingly, AMI-marker rates in our study increased over this same time period in a manner consistent with prior data in the WHO-MONICA

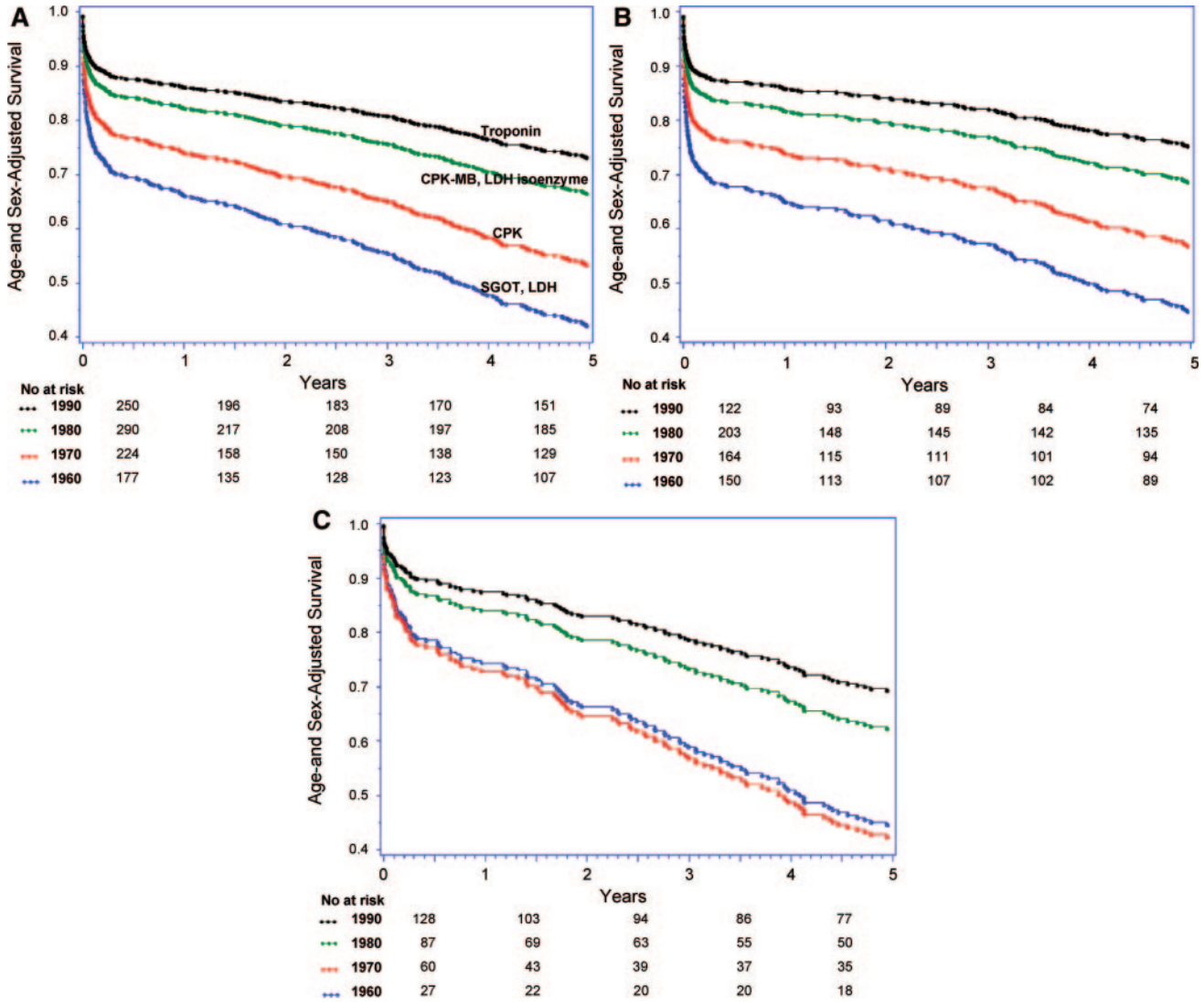


Figure 3. Up to 5-year case fatality after overall AMI (A), after AMI-ECG by decade (B), and after AMI-marker by decade (C), with the major biomarker used during each decade.

study, which showed higher AMI rates using biomarker-based definitions (troponin) compared with ECG-based definitions.[15] Similarly, an investigation in the Minnesota Heart Study demonstrated a 50% increase in AMI detection in 1980 when CPK and CPK-MB information was added to the Minnesota Heart Study AMI diagnostic algorithm.[14] Additional studies have mirrored these findings, showing that troponin-influenced AMI diagnosis has increased the AMI detection rate compared with AMI diagnosis based on CPK-MB and total CPK.[30,31] Our data extend these findings by demonstrating that biomarker-influenced AMI diagnosis has yielded a doubling in rates of AMI-marker over the 40-year period spanning 1960 to 1999.

The proportion of overall AMI diagnosed by ECG (68%) was similar to figures reported in a prior report from the Minnesota Heart Survey.[17] That investigation concluded that incident AMI-ECG rates were steady from 1975 to 1985 and declined from 1985 to 1995.[17] We extend these findings by providing data from 2 additional decades of observation. Our results demonstrate a 50% to 60% decline in AMI-ECG rates from 1960 to 1999. AMI-ECG likely represents a more advanced form of MI; declines demonstrated in out-of-hospital sudden cardiac death[2,12,32–34] (attributable to improved primary prevention efforts)[32] have likely contributed to some degree to the declines in the incidence of AMI-ECG.

Another possible explanation for the decline in AMI-ECG and the relative rise in AMI-marker may have to do with decreases in time from the onset of symptoms to hospital presentation and treatment (data from the National Registry of Myocardial Infarction),[35] which are thought to be due to public health education efforts and guideline implementation, which have collectively stressed the need to decrease door-to-intervention time for AMI.[35] On the other hand, other studies of community-based individuals and clinical trial participants have shown no temporal declines in prehospital delay during AMI.[36–38]

Case Fatality Rates

Our finding that AMI case fatality declined from 1960 to 1999 is consistent with studies conducted in the United

States[1,3,4,23] and Europe[22] that demonstrated declines in overall AMI case fatality over the past 20 to 40 years. Several studies have demonstrated that out-of-hospital sudden cardiac death has declined substantially over the past several decades.[2,12,32–34] We extend these findings by demonstrating that case fatality rates after AMI-ECG and after AMI-marker have declined to a similar degree.

Prior studies have suggested that improvements in primary prevention account for 40% to 50% of the reduction in CHD mortality in the Unites States from 1968 to 2000.[6,10] Our finding of a 50% decline in incidence of first AMI using an AMI definition for which bias is inherently low (ie, AMI-ECG) implies that primary prevention efforts also have influenced the incidence of AMI.

Strengths and Limitations

The availability of 4 decades of physician-validated AMI and case fatality data and the ability to separate AMI-ECG and AMI-marker are unique strengths of our investigation. Indeed, AMI diagnosis relying on ICD coding may have a sensitivity of only 60% compared with physician-validated AMI diagnosis.[39] Our adjudication committee had access to simultaneous ECG and biomarker information; therefore, the ECG adjudication could have been biased by knowledge of biomarker information. However, we believe that if such a bias were introduced, it would have biased results toward a greater proportion of AMI-ECG over time. We could not separate the contribution of specific biomarkers to AMI diagnosis among AMI-marker cases. Our study sample is largely white of European descent; therefore, our findings may not be applicable to other ethnic groups or other geographic regions. We had a relatively small number of events when sex, specific age groups, and 4 time periods are considered, possibly limiting our statistical power to detect differences. We had a limited number of subjects in each sex, age group, and time period. We did not provide confidence intervals for the trend analyses for the incidence rates of AMI-ECG and AMI-marker; in addition, because of limited statistical power, we did not test the interaction term of MI type with time period.

Implications of our Findings

The diagnosis of AMI is evolving; therefore, it is a challenge to accurately characterize the "true" epidemiology of AMI. However, our data demonstrate that although AMI-ECG rates have declined, this decline was offset by rising AMI-marker rates. Because the most sensitive biomarkers (ie, troponin) were not available in the earlier study decades (1960s to 1980s), AMI-marker earlier on may have been underdiagnosed. Regardless, the advent of increasingly sensitive biomarkers for AMI diagnosis has substantially influenced AMI detection rates in the United States over the past several decades.

Conclusions

Over the past 40 years, AMI-ECG rates have declined by 50% and AMI-marker rates have doubled, offering a possible explanation for apparently steady national rates of overall AMI in the face of improvements in primary prevention.

Source of Funding

This work was supported by National Institutes of Health/National Heart, Lung, and Blood Institute contract N01-HC-25195.

Disclosures

None.

References

1. *NHLBI Morbidity and Mortality Chart Book*. Bethesda, Md: National Heart, Lung, and Blood Institute; 2004.
2. Fox CS, Evans JC, Larson MG, Kannel WB, Levy D. Temporal trends in coronary heart disease mortality and sudden cardiac death from 1950 to 1999: the Framingham Heart Study. *Circulation*. 2004;110:522–527.
3. Guidry UC, Evans JC, Larson MG, Wilson PW, Murabito JM, Levy D. Temporal trends in event rates after Q-wave myocardial infarction: the Framingham Heart Study. *Circulation*. 1999;100:2054–2059.
4. Rosamond WD, Chambless LE, Folsom AR, Cooper LS, Conwill DE, Clegg L, Wang CH, Heiss G. Trends in the incidence of myocardial infarction and in mortality due to coronary heart disease, 1987 to 1994. *N Engl J Med*. 1998;339:861–867.
5. Unal B, Critchley JA, Capewell S. Explaining the decline in coronary heart disease mortality in England and Wales between 1981 and 2000. *Circulation*. 2004;109:1101–1107.
6. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, Giles WH, Capewell S. Explaining the decrease in U.S. deaths from coronary disease, 1980–2000. *N Engl J Med*. 2007;356:2388–2398.
7. Laatikainen T, Critchley J, Vartiainen E, Salomaa V, Ketonen M, Capewell S. Explaining the decline in coronary heart disease mortality in Finland between 1982 and 1997. *Am J Epidemiol*. 2005;162:764–773.
8. Beaglehole R. Medical management and the decline in mortality from coronary heart disease. *BMJ (Clin Res Ed)*. 1986;292:33–35.
9. Capewell S, Morrison CE, McMurray JJ. Contribution of modern cardiovascular treatment and risk factor changes to the decline in coronary heart disease mortality in Scotland between 1975 and 1994. *Heart*. 1999;81:380–386.
10. Goldman L, Cook EF. The decline in ischemic heart disease mortality rates: an analysis of the comparative effects of medical interventions and changes in lifestyle. *Ann Intern Med*. 1984;101:825–836.
11. Gregg EW, Cheng YJ, Cadwell BL, Imperatore G, Williams DE, Flegal KM, Narayan KM, Williamson DF. Secular trends in cardiovascular disease risk factors according to body mass index in US adults. *JAMA*. 2005;293:1868–1874.
12. Rosamond WD, Folsom AR, Chambless LE, Wang CH. Coronary heart disease trends in four United States communities: the Atherosclerosis Risk in Communities (ARIC) study 1987–1996. *Int J Epidemiol*. 2001;30(suppl 1):S17–S22.
13. Thygesen K, Alpert JS, White HD, Jaffe AS, Apple FS, Galvani M, Katus HA, Newby LK, Ravkilde J, Chaitman B, Clemmensen PM, Dellborg M, Hod H, Porela P, Underwood R, Bax JJ, Beller GA, Bonow R, Van der Wall EE, Bassand JP, Wijns W, Ferguson TB, Steg PG, Uretsky BF, Williams DO, Armstrong PW, Antman EM, Fox KA, Hamm CW, Ohman EM, Simoons ML, Poole-Wilson PA, Gurfinkel EP, Lopez-Sendon JL, Pais P, Mendis S, Zhu JR, Wallentin LC, Fernandez-Aviles F, Fox KM, Parkhomenko AN, Priori SG, Tendera M, Voipio-Pulkki LM, Vahanian A, Camm AJ, De Caterina R, Dean V, Dickstein K, Filippatos G, Funck-Brentano C, Hellemans I, Kristensen SD, McGregor K, Sechtem U, Silber S, Tendera M, Widimsky P, Zamorano JL, Morais J, Brener S, Harrington R, Morrow D, Lim M, Martinez-Rios MA, Steinhubl S, Levine GN, Gibler WB, Goff D, Tubaro M, Dudek D, Al-Attar N. Universal definition of myocardial infarction. *Circulation*. 2007;116:2634–2653.
14. Burke GL, Edlavitch SA, Crow RS. The effects of diagnostic criteria on trends in coronary heart disease morbidity: the Minnesota Heart Survey. *J Clin Epidemiol*. 1989;42:17–24.
15. Salomaa V, Koukkunen H, Ketonen M, Immonen-Raiha P, Karja-Koskenkari P, Mustonen J, Lehto S, Torppa J, Lehtonen A, Tuomilehto J, Kesaniemi YA, Pyorala K. A new definition for myocardial infarction: what difference does it make? *Eur Heart J*. 2005;26:1719–1725.
16. Lee TH, Goldman L. Serum enzyme assays in the diagnosis of acute myocardial infarction: recommendations based on a quantitative analysis. *Ann Intern Med*. 1986;105:221–233.
17. Crow RS, Hannan PJ, Jacobs DR Jr, Lee SM, Blackburn H, Luepker RV. Eliminating diagnostic drift in the validation of acute in-hospital myo-

cardial infarction: implication for documenting trends across 25 years: the Minnesota Heart Survey. *Am J Epidemiol*. 2005;161:377–388.
18. Dawber TR, Kannel WB. An approach to longitudinal studies in a community: the Framingham study. *Ann N Y Acad Sci*. 1963;107:539–556.
19. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol*. 1979;110:281–290.
20. Cupples LA, D'Agostino RB Sr. Some risk factors related to the annual incidence of cardiovascular disease and death using pooled repeated biennial measurements: Framingham Study, 30-year follow-up. In: Kannel WB, Wolf PA, Garrison RJ, eds. *The Framingham Heart Study: An Epidemiologic Investigation of Cardiovascular Disease*. Washington, DC: NIH Publication; 1987:87–203.
21. Fang J, Alderman MH. Dissociation of hospitalization and mortality trends for myocardial infarction in the United States from 1988 to 1997. *Am J Med*. 2002;113:208–214.
22. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M, Tolonen H, Ruokokoski E, Amouyel P. Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations: Monitoring Trends and Determinants in Cardiovascular Disease. *Lancet*. 1999;353:1547–1557.
23. Goldberg RJ, Yarzebski J, Lessard D, Gore JM. A two-decades (1975 to 1995) long experience in the incidence, in-hospital and long-term case-fatality rates of acute myocardial infarction: a community-wide perspective. *J Am Coll Cardiol*. 1999;33:1533–1539.
24. Luepker RV, Apple FS, Christenson RH, Crow RS, Fortmann SP, Goff D, Goldberg RJ, Hand MM, Jaffe AS, Julian DG, Levy D, Manolio T, Mendis S, Mensah G, Pajak A, Prineas RJ, Reddy KS, Roger VL, Rosamond WD, Shahar E, Sharrett AR, Sorlie P, Tunstall-Pedoe H. Case definitions for acute coronary heart disease in epidemiology and clinical research studies: a statement from the AHA Council on Epidemiology and Prevention; AHA Statistics Committee; World Heart Federation Council on Epidemiology and Prevention; the European Society of Cardiology Working Group on Epidemiology and Prevention; Centers for Disease Control and Prevention; and the National Heart, Lung, and Blood Institute. *Circulation*. 2003;108:2543–2549.
25. Alpert JS. Will the real myocardial infarction please stand up? *Clin Chem*. 2006;52:795–796.
26. Arciero TJ, Jacobsen SJ, Reeder GS, Frye RL, Weston SA, Killian JM, Roger VV. Temporal trends in the incidence of coronary disease. *Am J Med*. 2004;117:228–233.
27. McGovern PG, Pankow JS, Shahar E, Doliszny KM, Folsom AR, Blackburn H, Luepker RV. Recent trends in acute coronary heart disease: mortality, morbidity, medical care, and risk factors: the Minnesota Heart Survey Investigators. *N Engl J Med*. 1996;334:884–890.
28. McGovern PG, Jacobs DR Jr, Shahar E, Arnett DK, Folsom AR, Blackburn H, Luepker RV. Trends in acute coronary heart disease mortality, morbidity, and medical care from 1985 through 1997: the Minnesota Heart Survey. *Circulation*. 2001;104:19–24.
29. Assaf AR, Lapane KL, McKenney JL, Carleton RA. Possible influence of the prospective payment system on the assignment of discharge diagnoses for coronary heart disease. *N Engl J Med*. 1993;329:931–935.
30. Falahati A, Sharkey SW, Christensen D, McCoy M, Miller EA, Murakami MA, Apple FS. Implementation of serum cardiac troponin I as marker for detection of acute myocardial infarction. *Am Heart J*. 1999;137:332–337.
31. Kontos MC, Fritz LM, Anderson FP, Tatum JL, Ornato JP, Jesse RL. Impact of the troponin standard on the prevalence of acute myocardial infarction. *Am Heart J*. 2003;146:446–452.
32. Goraya TY, Jacobsen SJ, Kottke TE, Frye RL, Weston SA, Roger VL. Coronary heart disease death and sudden cardiac death: a 20-year population-based study. *Am J Epidemiol*. 2003;157:763–770.
33. Rea TD, Crouthamel M, Eisenberg MS, Becker LJ, Lima AR. Temporal patterns in long-term survival after resuscitation from out-of-hospital cardiac arrest. *Circulation*. 2003;108:1196–1201.
34. Bunch TJ, White RD, Friedman PA, Kottke TE, Wu LA, Packer DL. Trends in treated ventricular fibrillation out-of-hospital cardiac arrest: a 17-year population-based study. *Heart Rhythm*. 2004;1:255–259.
35. Rogers WJ, Canto JG, Lambrew CT, Tiefenbrunn AJ, Kinkaid B, Shoultz DA, Frederick PD, Every N. Temporal trends in the treatment of over 1.5 million patients with myocardial infarction in the US from 1990 through 1999: the National Registry of Myocardial Infarction 1, 2 and 3. *J Am Coll Cardiol*. 2000;36:2056–2063.
36. Goldberg RJ, Yarzebski J, Lessard D, Gore JM. Decade-long trends and factors associated with time to hospital presentation in patients with acute myocardial infarction: the Worcester Heart Attack study. *Arch Intern Med*. 2000;160:3217–3223.
37. Gibler WB, Armstrong PW, Ohman EM, Weaver WD, Stebbins AL, Gore JM, Newby LK, Califf RM, Topol EJ. Persistence of delays in presentation and treatment for patients with acute myocardial infarction: the GUSTO-I and GUSTO-III experience. *Ann Emerg Med*. 2002;39:123–130.
38. McGinn AP, Rosamond WD, Goff DC Jr, Taylor HA, Miles JS, Chambless L. Trends in prehospital delay time and use of emergency medical services for acute myocardial infarction: experience in 4 US communities from 1987–2000. *Am Heart J*. 2005;150:392–400.
39. Rosamond WD, Chambless LE, Sorlie PD, Bell EM, Weitzman S, Smith JC, Folsom AR. Trends in the sensitivity, positive predictive value, false-positive rate, and comparability ratio of hospital discharge diagnosis codes for acute myocardial infarction in four US communities, 1987–2000. *Am J Epidemiol*. 2004;160:1137–1146.

CLINICAL PERSPECTIVE

Whereas the prevalence of coronary heart disease risk factors has declined over past decades in the United States, acute myocardial infarction (AMI) rates have been steady. Because the diagnosis of AMI is evolving, it is a challenge to characterize the "true" epidemiology of AMI accurately. Among Framingham Heart Study participants, we found that over the past 40 years, rates of AMI diagnosed by ECG have declined by 50%, whereas rates of AMI diagnosed by biomarkers have doubled. The advent of increasingly sensitive biomarkers for AMI diagnosis has substantially influenced AMI detection rates in the United States over the past several decades. Our findings offer an explanation for the apparently steady national AMI rates in the face of improvements in primary prevention.

CLINICAL RESEARCH STUDY

THE AMERICAN JOURNAL of MEDICINE®

Breastfeeding in Infancy and Adult Cardiovascular Disease Risk Factors

Nisha I. Parikh, MD, MPH,[a,b] Shih-Jen Hwang, PhD,[a,c] Erik Ingelsson, MD, PhD,[a] Emelia J. Benjamin, MD, ScM,[a,d,e] Caroline S. Fox, MD, MPH,[a,c,f] Ramachandran S. Vasan, MD,[a,d] Joanne M. Murabito, MD, ScM[a,j]

[a]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Mass; [b]Cardiovascular Division, Beth Israel Deaconess Medical Center, Boston, Mass; [c]National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Md; [d]Preventive Medicine and Cardiology Sections, Boston University School of Medicine, Boston, Mass; [e]Epidemiology Section, Boston University School of Public Health, Boston, Mass; [f]Brigham and Women's Hospital, Harvard Medical School, Boston, Mass; [j]Section of General Internal Medicine, Boston University School of Medicine, Boston, Mass.

ABSTRACT

BACKGROUND: Public health recommendations advocate breastfeeding in infancy as a means to reduce obesity in later life. Several prior studies relating breastfeeding to cardiovascular risk factors have been limited by lack of adjustment for maternal and participant confounding factors.

METHODS: We ascertained breastfeeding history via questionnaire from mothers enrolled in the Framingham Offspring Study. In their young to middle-aged adult children enrolled in the Framingham Third Generation, we examined the relations between maternal breastfeeding history (yes, no) and cardiovascular risk factors, including body mass index (BMI), high-density lipoprotein (HDL) cholesterol, total cholesterol, triglycerides, fasting blood glucose, and systolic and diastolic blood pressure levels. We applied generalized estimating equations to account for sibling correlations and adjusted for maternal and participant lifestyle, education, and cardiovascular risk factors.

RESULTS: In Third Generation participants (n = 962, mean age = 41 years, 54% were women), 26% of their mothers reported breastfeeding. Compared with non-breastfed individuals, breastfed adult participants had lower multivariable-adjusted BMI (26.1 kg/m^2 vs 26.9 kg/m^2, $P = .04$) and higher HDL cholesterol levels (HDL 56.6 mg/dL vs 53.7 mg/dL, $P = .01$). On additional adjustment for BMI, the association between breastfeeding and HDL cholesterol was attenuated ($P = .09$). Breastfeeding was not associated with total cholesterol, triglycerides, fasting blood glucose, systolic blood pressure, or diastolic blood pressure.

CONCLUSION: Breastfeeding in infancy is inversely associated with adult BMI and positively associated with HDL cholesterol. Associations between breastfeeding and BMI may mediate the association between breastfeeding and HDL cholesterol.

 • *The American Journal of Medicine (2009) 122, 656-663*

KEYWORDS: Body mass index; Breastfeeding; Early nutrition; High-density lipoprotein cholesterol; Infancy; Lactation; Risk factors

Recent US and World Health Organization recommendations strongly advocate breastfeeding in infancy as a means toward not only reducing infant infections but also protecting against adverse adult health outcomes such as obesity.[1-3] Prior epidemiologic evidence suggests that breastfeeding in infancy also may have protective effects on cardiovascular disease risk factor profiles[4-9] and cardiovascular disease risk[4,10] in adulthood. Previous studies have demonstrated that breastfeeding in infancy can lead to small reductions in adolescent and adult blood pressure levels,[4,6,11] decreased total cholesterol and low-density lipoprotein cholesterol (LDL-C) levels in adulthood,[4,5,8] and modest decreases in adult body mass index (BMI).[4]

Funding: Supported by a National Institute of Health/National Heart, Lung, and Blood Institute, contract N01-HC-25195, 2K24 HL 04334 (RSV).

Conflict of Interest: None.

Authorship: All authors had access to the data and played a role in writing this manuscript.

Requests for reprints should be addressed to Joanne M. Murabito, MD, ScM, 73 Mount Wayte Ave, #2, Framingham, MA 01702.

E-mail address: murabito@bu.edu

0002-9343/$ -see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.amjmed.2008.11.034

Some prior studies have been limited by self-reported as opposed to directly measured maternal[12] and offspring cardiovascular disease risk factors.[12] Furthermore, some prior studies have been limited by the failure to account for potential maternal and offspring confounders, including socioeconomic status.[8] Highlighting the importance of accounting for socioeconomic status, a recent investigation in the Nurses Health Study did not demonstrate a significant association between breastfeeding and BMI on adjustment for socioeconomic status.[12]

CLINICAL SIGNIFICANCE

- Being breastfed in infancy for 1 month or more is associated with higher adult HDL levels and lower mean adult body mass index.
- Our study suggests that the benefits of breastfeeding extend beyond childhood to adult health outcomes.

Detailed risk factor ascertainment, sociodemographic data collection, and maternal breastfeeding report among Framingham Heart Study Offspring mothers and their adult children in the Third Generation cohort allowed the opportunity to extend previous data on the association of breastfeeding in infancy with several cardiovascular disease risk factors in adulthood. We hypothesized that breastfeeding in infancy would be protective for cardiovascular disease risk factors, but that these associations would be attenuated after accounting for maternal and participant socioeconomic and lifestyle characteristics.

MATERIALS AND METHODS

Study Sample

Participants for this study were part of the Third Generation cohort of the Framingham Heart Study; their mothers were members of the Offspring cohort. The Original Framingham Heart Study Cohort[13] and Framingham Offspring cohorts have been described.[14] Between July 1996 and May 1997, a breast health survey was mailed to Offspring cohort women that included questions regarding breastfeeding history of each of their children. The design and selection criteria for women chosen to receive the breast health survey have been described.[15] Briefly, women free of breast cancer with a first-degree female relative enrolled in the Framingham Heart Study were sampled on the basis of 1 of 3 criteria: women having a mother or sister with documented breast cancer; women having a mother or sister with a non-gynecologic cancer; and women with mothers or sisters free of documented cancer. A total of 683 participants (77%) returned the questionnaire (**Figure**).

Among these 683 offspring participants, participants were excluded who completed the questionnaire who did not have a child (n = 60), whose children did not attend the Third Generation study (n = 142), or who returned the questionnaire with incomplete or inconsistent information (n = 88). Because of these exclusion criteria, only 393 (44%) of the offspring women survey sample were included in the analysis. These 393 mothers provided the source for the Third Generation cohort participants (n = 962) (**Figure**). The 393 offspring women included in this study when compared with offspring women with children enrolled in the Third Generation cohort who are not included in this study (n = 1114) had similar cardiovascular disease risk factor profiles at enrollment in the Framingham Heart Study with the exception of a slightly higher diastolic blood pressure (77 vs 75 mm Hg) and lower rates of cigarette smoking (42% vs 53%) ($P < .05$) (Appendix, available online). The Boston University Medical Center Institutional Review Board approved the main study protocols for the Framingham Offspring and Third Generation cohort, and all participants signed written informed consent.

Risk Factor Collection

In the current study, we used information on cardiovascular risk factors from the first and second offspring examinations (Offspring cohort mothers) and the first Third Generation examination (adult progeny). Details regarding the ascertainment of risk factors have been described.[16] Diabetes was defined as fasting plasma glucose ≥ 126 mg/dL or treatment with either insulin or oral hypoglycemic agents. Lipids were measured on 12-hour fasting venous blood samples collected in tubes containing 0.1% EDTA. Plasma was separated by ultracentrifugation, and plasma lipid concentrations (total cholesterol and HDL-C) were measured as previously described.[17] HDL-C was measured after precipitation of apo B-containing lipoproteins, and low-density lipoprotein cholesterol concentrations were estimated using the Friedewald formula.[18] Intra-assay coefficients of variation for the Third Generation cohort cholesterol, triglycerides, and high-density lipoprotein were 0.5%, 1.1%, and 1.4%, respectively; interassay coefficients of variation were 1.1%, 1.8%, and 3.0%, respectively. Seated blood pressure was measured by

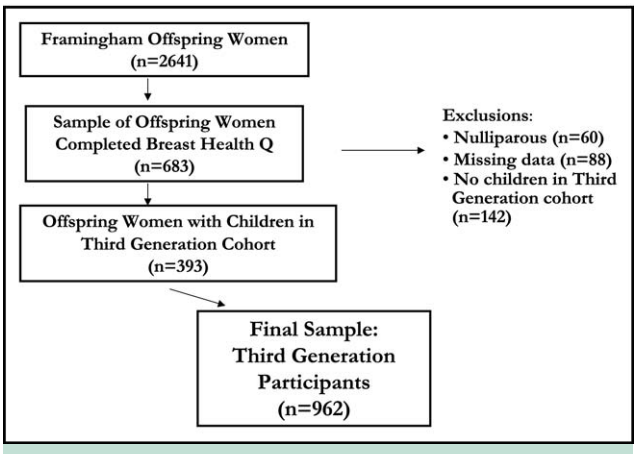


Figure Creation of study sample based on Framingham Offspring (maternal) breastfeeding information.

a trained physician after the participant had rested for 5 minutes, and the average of 2 physician-obtained readings was used. Hypertension was defined as systolic blood pressure ≥ 140 mm Hg, diastolic blood pressure ≥ 90 mm Hg, or use of blood pressure-lowering medications. Medication use was ascertained by physicians by detailed review of participant medication lists, and Third Generation participants were asked to bring medication bottles to the clinic examination, including lipid-lowering medications, oral contraceptive pills, and hormone therapy use.

Participants were considered to be current smokers if they smoked at least 1 cigarette per day for the year before examination. Categories of BMI were defined according to National Heart Lung and Blood Institute and the World Health Organization guidelines[19,20] as follows: normal weight (BMI 18.5-25 kg/m^2), overweight (25 ≤ BMI < 30 kg/m^2), and obese (≥30 kg/m^2). Data regarding highest educational degree obtained were gathered via questionnaire and categorized as follows: high school diploma or equivalent or less, Associate's degree/junior college, Bachelor's degree, and Master's degree or doctorate. The physical activity index was reflective of physical activity performed in a typical 24-hour period using a structured questionnaire that asked participants to report the number of hours asleep; at rest; and in slight, moderate, and heavy activity in a typical day.[21,22] Moderate-to heavy alcohol intake was defined as consumption of more than 14 drinks per week in men or 7 drinks per week in women. Prevalent cardiovascular disease was defined as recognized myocardial infarction, coronary insufficiency (prolonged chest pain accompanied by reversible ischemic electrocardiographic changes), angina pectoris, stroke, transient ischemic attack, or intermittent claudication using previously described criteria.[23]

Statistical Methods

Descriptive statistics of Third Generation participant characteristics were grouped by maternal breastfeeding status (age- and sex-adjusted generalized estimated equation models were used to compare characteristics of participants by breastfeeding status). Descriptive statistics of maternal characteristics were presented according to whether mothers breastfed none, some, or all of their children. We examined cardiovascular risk factors described above as end points in the Third Generation Cohort. No cardiovascular events were examined in our study sample. Generalized estimated equation models were used to assess relations between dichotomous (ever vs never) breastfeeding status and the following Third Generation participant cardiovascular disease risk factors: BMI, total cholesterol, HDL cholesterol, triglycerides, fasting blood glucose, systolic blood pressure, and diastolic blood pressure. The intraclass correlation of BMI was 0.02 (P = .47) among Third Generation Cohort siblings. We applied generalized estimated equation models to account for related observations given the presence of siblings in the Third Generation cohort. Statistical models were constructed with adjustment for the following: model 1—age, sex, hypertension treatment, lipid treatment, smoking status, birth order, oral contraceptive use, hormone replacement use, physical activity index, and education level; model 2—model 1 covariates plus maternal smoking status, maternal education level, and maternal BMI at study entry.

In secondary analyses, the dependent variable BMI was additionally adjusted for HDL cholesterol level, and the outcome HDL cholesterol was additionally adjusted for the following covariates: participant BMI, participant alcohol intake, and maternal HDL cholesterol. We also analyzed the dependent variables of cardiovascular disease risk factors dichotomously using clinically meaningful cut points. We tested potential effect modification by educational level. Finally, to assess the association of breastfeeding in infancy with an overall healthy lifestyle, we related breastfeeding status with participant higher education, smoking status, and physical activity.

All statistical analyses were performed using SAS statistical software (version 8.1). A P value of less than .05 was considered to be statistically significant.

RESULTS

Third Generation Participant Characteristics

Study sample characteristics grouped by breastfeeding status are shown in **Table 1**. Twenty-six percent of participants were reported by mothers to have been breastfed in infancy. Of those individuals who were breastfed, the median breastfeeding duration was 4 months (range 1-22 months), and 29.6% were breastfed for more than 6 months. A higher prevalence of breastfed individuals had higher education levels and a lower prevalence of diabetes (**Table 1**).

Maternal Characteristics

Characteristics of mothers by whether they breastfed none, all, or some of their children are shown in **Table 2**. Mothers who breastfed all of their children also had the highest education levels, were the leanest, and were least likely to smoke.

Breastfeeding Status and Cardiovascular Disease Risk Factors in Adulthood

In model 1, which adjusted for participant cardiovascular disease risk factors, physical activity, and education, breastfeeding (ever vs never) was associated with a lower BMI (P = .03) (**Table 3**). Additionally adjusting for maternal factors in model 2 (maternal smoking, education, and BMI) did not materially change the association between breastfeeding status and BMI (P = .04; adjusted mean BMI among those breastfed vs not breastfed was 26.1 vs 26.9 kg/m^2, respectively).

In model 1, adjusting for participant cardiovascular disease risk factors, physical activity, and education, breastfeeding was associated with a higher HDL cholesterol level (P = .01). Additionally adjusting for maternal factors (smoking, education, and maternal BMI) did not materially change the association between breastfeeding status and HDL cholesterol level (P = .01; adjusted mean HDL cho-

Table 1 Third Generation Participant Characteristics by Breastfeeding Status

Characteristics	Third Generation Participant Breastfeeding Status	
Means (SD) or (%)	No n = 712	Yes n = 250
Age (y)	41 ± 7	41 ± 9
Women (%)	54.8	50.0
Birth order (among siblings)	2.4 + 1.4	2.0 ± 1.2
Systolic blood pressure (mm Hg)	117 ± 15	118 ± 15
Diastolic blood pressure (mm Hg)	76 ± 10	75 ± 10
Fasting blood glucose (mg/dL)	95 ± 19.5	94 ± 10
Total cholesterol (mg/dL)	190 ± 34	190 ± 32.5
HDL cholesterol (mg/dL)[a]	54 ± 16	56 ± 15
Triglycerides (mg/dL)	116 ± 90.5	108 ± 79
BMI (kg/m^2)	26.9 ± 5.4	26.3 ± 5.0
Waist circumference (cm)	93 ± 15	93 ± 14
Physical activity index	37.6 ± 8	37.6 ± 8
Hypertension, %	7.8	8.0
Lipid treatment, %	5.9	7.2
Diabetes mellitus, %	3.0	1.2
Obesity (BMI ≥ 30), %	23.5	20.4
Overweight (25 ≤ BMI < 30), %	35.8	32.0
Total cholesterol > 200, %	36.5	36.3
Triglyceride > 150, %	21.4	16.0
Fasting glucose > 126, %	2.7	0.8
Oral contraceptive use, %	20.4	14.8
Hormone replacement therapy use, %	3.8	3.2
Smoking, %	17.0	15.3
Moderate alcohol intake,[b] %	15.4	18.2
Education level[a]		
High school or less, %	17.5	9.6
Some college, %	32.3	30.9
Bachelor's degree, %	35.7	41.0
Master's degree or higher, %	14.5	18.5

HDL = high-density lipoprotein; BMI = body mass index.
[a]Age- and sex-adjusted $P < .05$.
[b]Defined as > 7 drinks per week in women and > 14 drinks per week in men.

lesterol concentrations among those breastfed vs not breastfed were 56.6 mg/dL relative to 53.7 mg/dL, respectively).

Breastfeeding (ever vs never) was not associated with participant total cholesterol, triglycerides, fasting blood glucose, systolic blood pressure, or diastolic blood pressure in either model 1 or model 2 (**Table 3**).

Secondary Analyses

Association between Breastfeeding Status and Body Mass Index and High-density Lipoprotein Cholesterol Categories. Breastfeeding was inversely associated with low HDL cholesterol levels (<40 mg/dL in men and < 50 mg/dL in women) even after accounting for participant and maternal cardiovascular disease risk factors, lifestyle, and socioeconomic characteristics (multivariable-adjusted odds ratio = 0.63 [0.42-0.96], $P = .03$) (**Table 4**). Breastfeeding was not significantly associated with any other dichotomized risk factors in fully adjusted models. (**Table 4**).

Additional Model Adjustments. Additionally adjusting multivariable model 2 for participant alcohol intake did not materially change the positive association between breastfeeding status and HDL cholesterol. Similarly, additionally adjusting model 2 for maternal HDL cholesterol did not materially change the positive association between breastfeeding status and HDL cholesterol.

Effect Modification. There was no evidence of effect modification by educational level on the associations between breastfeeding with HDL or with BMI.

Association between Breastfeeding and Healthy Lifestyle. In fully adjusted models there were no significant associations between breastfeeding and having achieved a Bachelor's or higher degree (odds ratio = 1.20 [0.84-1.72]), having a physical activity score of 37 or more (1.20 [0.84-1.73]), or current cigarette smoking (1.17 [0.75-1.52]).

DISCUSSION

Summary of Findings

In a community-based sample of 962 men and women in early to middle age, maternal report of breastfeeding was associated with modestly lower participant BMI and higher participant HDL cholesterol concentrations. Maternal report of breastfeeding was not significantly associated with offspring total cholesterol, triglycerides, fasting blood glucose, or systolic or diastolic blood pressure levels. Breastfeeding was associated with higher mean HDL cholesterol concentrations even after accounting for participant and maternal education, lifestyle factors, and cardiovascular disease risk factors. However, the association between maternal breastfeeding and participant HDL cholesterol appeared to be attenuated by adjustment for participant BMI.

Breastfeeding and Body Mass Index

In keeping with our data, prior studies have found an inverse association between breastfeeding in infancy and adolescent and adult adiposity.[4,7,12,24-27] Higher growth rates in early infancy among formula-fed compared with breastfed infants have been demonstrated in randomized trials of low birth weight and preterm infants,[28] as well as in observational studies among normal birth weight babies.[29-31] In contrast with some prior reports, a significant attenuation in the association between breastfeeding and lower BMI was not found on adjustment for maternal and participant socioeconomic status defined using educational attainment. Furthermore, BMI is a moderately heritable trait,[32] yet adjustment for maternal BMI did not significantly diminish the associations. One prior study also demonstrated a significant inverse association between breastfeeding and childhood

Table 2 Selected Maternal Characteristics at Study Entry by Breastfeeding Status[a]

Characteristic	Maternal Breastfeeding Report Breastfed		
Means (SD) or N (%)	No Children n = 250	Some Children (Not All) n = 63	All Children n = 80
Age at study entry	35.8 ± 7.6	35.3 ± 9.0	40.1 ± 7.6
Parity (No. live births)	3.3 ± 1.6	2.8 ± 1.4	3.7 ± 1.5
Systolic blood pressure (mm Hg)	117.6 ± 14.2	116.7 ± 14.6	121.1 ± 18.6
Diastolic blood pressure (mm Hg)	77.2 ± 9.1	76.4 ± 10.8	78.9 ± 10.3
Total cholesterol (mg/dL)	192.9 ± 36.9	187.3 ± 33.8	196.6 ± 40.8
LDL cholesterol (mg/dL)	119.4 ± 33.4	114.0 ± 35.1	123.0 ± 36.9
HDL cholesterol (mg/dL)	56.9 ± 14.3	58.2 ± 13.0	56.9 ± 17.3
Triglycerides (mg/dL)	82.0 ± 57.9	69.9 ± 40.0	84.0 ± 52.4
BMI (kg/m^2)	24.2 ± 4.5	23.8 ± 3.6	24.5 ± 4.6
Hypertension, n (%)	27 (11)	12 (19)	12 (19)
Diabetes mellitus, n (%)	1 (0.4)	0	0
Smoking, n (%)	100 (42)	24 (38)	24 (30)
Overweight, n (%)	41 (16)	11 (17)	16 (20)
Obesity, n (%)	35 (15)	8 (13)	7 (9)
Education level[b]			
Less than high school n (%)	13 (6)	11 (18)	6 (9)
Some college n (%)	155 (68)	34 (57)	28 (42)
Bachelor's degree n (%)	55 (24)	12 (20)	21 (32)
Master's degree or higher n (%)	5 (2)	3 (5)	11 (17)

BMI = body mass index; SD = standard deviation; LDL = low-density lipoprotein; HDL = high-density lipoprotein.

[a]Risk factors taken at first examination the woman achieved age ≥ 20 y (5 mothers were aged < 20 y at examination cycle 1; therefore, examination cycle 2 data were used).

[b]Data available in 60 women who breastfed some children, 228 women who did not breastfeed, and 66 women who breastfed all of their children.

overweight even after adjusting for maternal obesity.[33] Our BMI data were ascertained directly in both mothers and study sample participants rather than self-reported.[12] Self-reported data used in prior studies may have led to some outcome misclassification with resultant biasing of measures toward the null value. Finally, our study was conducted in a sample unselected for sex and occupation, in contrast with prior investigations conducted among female registered nurses.[12]

Exact mechanisms by which breast milk confers protection against offspring weight gain are not known, but aggregate data from several recent studies suggest that adipokines may potentially mediate the association. A recent laboratory study in rats has suggested that delayed weaning

Table 3 Least Square Means for Adulthood Cardiovascular Disease Risk Factors by Breastfeeding Status in Infancy

	Model 1[a]			Model 2[b]		
Risk Factor	Not Breastfed N = 712	Breastfed N = 250	*P* Value	Not Breastfed N = 712	Breastfed N = 250	*P* Value
BMI, kg/m^2	27.0 (26.6–27.4)	26.1 (25.5–26.8)	.03	26.9 (26.4–27.3)	26.1 (25.4–26.7)	.04
HDL cholesterol, mg/dL	53.8 (52.5–54.8)	56.1 (54.5–58.2)	.01	53.7 (52.5–54.9)	56.6 (54.7–58.5)	.01
Total cholesterol, mg/dL	190.7 (187.6–193.2)	190.7 (186.9–195.4)	.8	190.4 (187.3–193.2)	189.2 (185.0–194.3)	.8
Triglycerides, mg/dL	117.2 (110.1–124.1)	109.1 (99.1–119.6)	.2	115.7 (108.3–123.0)	109.4 (97.8–121.1)	.4
Fasting blood glucose, mg/dL	94.8 (93.5–96.2)	93.9 (92.3–95.1)	.3	94.8 (93.4–96.4)	93.5 (91.8–95.0)	.2
Systolic blood pressure, mm Hg	117.6 (116.3–118.8)	117.1 (115.8–119.1)	.9	117.6 (116.3–118.8)	117.5 (115.9–119.6)	.9
Diastolic blood pressure, mm Hg	75.9 (75.1–76.6)	74.8 (73.7–76.1)	.2	75.8 (74.9–76.5)	74.9 (73.7–76.3)	.3

BMI = body mass index; HDL = high-density lipoprotein.

If we do not account for multiple testing, then $\alpha < 0.05$ is significant. With the most conservative approach, with a Bonferroni correction (accounting for 7 different dependent variables/tests), $\alpha < 0.05/7$ or 0.007 is significant.

[a]Model 1 covariates: age, sex, hypertension treatment, lipid treatment, smoking status, birth order, oral contraceptive use, hormone replacement use, physical activity, and education level.

[b]Model 2 covariates: model 1 variables plus maternal smoking status, maternal education level, and maternal BMI at study entry.

Table 4 Odds Ratios for Dichotomized Adulthood Cardiovascular Disease Risk Factors by Breastfeeding Status in Infancy

	Model 1[a] OR (95% CI)	Model 2[b] OR (95% CI)
BMI > 30 kg/m^2	0.78 (0.52–1.16)	0.75 (0.47–1.21)
Total cholesterol > 200 mg/dL	1.01 (0.73–1.40)	0.91 (0.63–1.32)
HDL cholesterol < 40 mg/dL in men or < 50 mg/dL in women	0.64 (0.43–0.94)	0.63 (0.42–0.96)
Triglycerides > 150 mg/dL	0.64 (0.42–0.97)	0.68 (0.44–1.05)
Fasting glucose > 126 mg/dL	0.30 (0.08–1.17)	0.40 (0.09–1.70)

OR = odds ratio; CI = confidence interval; BMI = body mass index; HDL = high-density lipoprotein.

[a]Model 1 covariates: age, sex, hypertension treatment, lipid treatment, smoking status, birth order, oral contraceptive use, hormone replacement use, physical activity, and education level.

[b]Model 2 covariates: model 1 variables plus maternal smoking status, maternal education level, and maternal BMI at study entry.

(meaning continuation of breast milk and delayed introduction of solid food) reduces plasma levels of the appetite-related peptide, ghrelin, and gastric ghrelin cell development.[34] Ghrelin concentration increases during specific stages in rat infancy were formerly thought to be age related as opposed to diet related.[34] Furthermore, leptin concentrations in human breast milk have been demonstrated to inversely correlate with human infant weight gain up until 2 years of age.[35,36] In a separate randomized prospective study of feeding among preterm infants, serum leptin to fat mass ratio measured in adolescence was demonstrated to be lower in those randomized to donated banked breast milk compared with formula feeding while in infancy.[37] Levels of other novel adipokines in human breast milk, including epidermal and adipocyte fatty acid binding protein, have been demonstrated to positively correlate with infant birth weight.[38]

Breastfeeding and High-density Lipoprotein Cholesterol

Whereas prior data suggest that breastfeeding is related to increases in total and LDL-cholesterol levels in infancy and adulthood,[5] fewer studies have specifically examined the association between breastfeeding and later-life HDL cholesterol levels. A recent study in a British birth cohort born in 1958 did not demonstrate an association between breastfeeding for more than 1 month and adult levels of HDL cholesterol.[27] A Dutch prospective study of adults aged 48 to 53 years demonstrated a lower total to HDL cholesterol ratio among breastfed compared with formula-fed individuals.[8] In a randomized prospective study of feeding among preterm infants, serum total to HDL cholesterol ratio measured in adolescence was lower among those previously randomized to banked donated breast milk compared with formula feeding.[39]

Choice of Covariates for Adjustment

We thought it was particularly important to adjust for education, BMI, and smoking status. Breastfeeding is more prevalent among women with a higher education, which in turn is associated with a number of positive health indicators, including increased HDL cholesterol,[40] lower BMI,[41] and abstinence from smoking. Lower BMI and abstinence from smoking in turn are associated with higher HDL cholesterol levels.[42]

Breastfeeding and Other Cardiovascular Disease Risk Factors

The absence of a significant association between breastfeeding in infancy and later-life blood pressure in our sample is consistent with several prior investigations that did not find a significant association between breastfeeding and adult blood pressure.[8,43,44] Furthermore, findings from a recent meta-analysis of several previously published studies raise the concern that the inverse association between breastfeeding and blood pressure from other studies may have been subject to selection or publication bias.[6]

We found no association between breastfeeding status and fasting blood glucose levels. This is in keeping with prior studies in adolescents[45] and middle-aged men[4] showing no association between breastfeeding in infancy and later-life insulin resistance (as measured by homeostasis model assessment). We did not specifically study differences in rates of diabetes by breastfeeding status because the prevalence of diabetes was too low in our sample for meaningful analysis.

Given the lack of association between breastfeeding status and several cardiovascular disease risk factors studied, we assessed our statistical power to detect modest effects for associations between breastfeeding and the cardiovascular disease risk factors for which we did not detect significant associations. Taking sibling correlation into account, we had 80% power to detect a systolic blood pressure difference of 2.9 mm Hg, diastolic blood pressure difference of 2.0 mm Hg, total cholesterol difference of 7.0 mg/dL, triglyceride difference of 17.2 mg/dL, and fasting glucose difference of 2.9 mg/dL. We had more modest power to detect smaller mean differences.

STRENGTHS AND LIMITATIONS

Direct and routine assessment of cardiovascular risk factors for 2 generations of participants to account for both maternal covariates and offspring cardiovascular disease risk factors is a unique strength of our study. Risk factors were measured in offspring in adulthood, whereas most prior reports examined the relation of breastfeeding to childhood risk factors. Several limitations should be acknowledged as well. Breastfeeding assessment was done decades after the birth of participants, which could have led to recall bias. However, the recall of whether or not a women breastfed her child has been shown to be accurate for ≥ 20 years later.[46] Furthermore, our study relied on maternal compared

with self-reported breastfeeding history, which has been demonstrated to be more accurate.[47] We did not adjust for other components of infant diet or account for birth weight in our multivariable analysis. It has been demonstrated that low birth weight infants tend to breastfeed for shorter durations and tend to have rapid catch-up growth, which is associated with later-life obesity.[48] We also did not account for paternal factors because not all Third Generation participants have fathers in the Framingham Offspring cohort. Our study participants are of white European ancestry; therefore, these findings may not be generalizable to other ethnic populations. We accounted for socioeconomic status by means of highest education degree obtained, which might not have fully accounted for socioeconomic differences.[49] HDL subfractions, which have demonstrated accuracy in predicting cardiovascular disease,[50] were not measured in this study. Although the incubation period between exposure and outcome in our study is relatively long, it is being increasingly recognized that exposures in early life affect adult health. In turn, evidence suggests that cholesterol and BMI measured in middle age confer later higher lifetime cardiovascular disease risk.[51]

Potential selection bias from the breast health survey sampling scheme cannot be excluded; however, Offspring mothers with adult children enrolled in the Third Generation cohort not included in our study did not differ with respect to BMI and HDL cholesterol levels from the Offspring mothers included in the study. Because the alternatives to breastfeeding in the 1960s and 1970s differed from what is available today, these comparisons might not be relevant to current long-term breastfeeding effects. We did not assess the exclusivity of breastfeeding within our study framework and were unable to carry out an analysis of risk factor levels among siblings discordant for breastfeeding because we had few of these in our study sample to permit a meaningful analysis. A discordant sibling pair analysis may have permitted better control of unmeasured potentially confounding maternal and family level factors. We did not account for the dietary intake of participants (ie, fat, carbohydrate, and protein intake). We may have limited power to detect very modest differences in blood pressure, total cholesterol, triglycerides, and fasting glucose. We did not account for multiple testing in our interpretation of results. By using the most conservative approach, given that model 1 and model 2 adjustments were highly correlated, and there were 7 separate dependent variables, the Bonferroni correction would have yielded an α level for significance of $0.05/7 = 0.007$. Finally, this is an observational study, and therefore we cannot infer causality.

IMPLICATIONS AND DIRECTIONS FOR FUTURE STUDY

Our findings confirm previous reports of a protective association between breastfeeding and later-life adiposity (as measured by BMI). Although the net reductions in BMI demonstrated in our study are modest, the beneficial effect at the population level may have important public health relevance. The risk of death from cardiovascular disease and congestive heart failure has been demonstrated to increase even with small incremental increases in BMI,[52,53] suggesting that even modest differences in excess adiposity may increase cardiovascular disease mortality risks. Furthermore, the mechanisms underlying the association between lower adulthood BMI among individuals breastfed in infancy are arguably of considerable importance. Our findings taken in conjunction with recent experimental evidence linking adipokines to breast milk and infant weight suggest that further elucidating mechanisms relating nutrition in early life and cardiometabolic risk factor profile in later life is an important area of research. Furthermore, informed decisions about whether or not to breastfeed affect more than 4 million women annually[54] who give birth in the United States. Thus, understanding the association of breastfeeding with cardiovascular disease risk factors in later life remains an important public health issue.

CONCLUSIONS

Breastfeeding in infancy was associated with a modestly reduced BMI and elevated HDL cholesterol levels in adulthood after accounting for several participant and maternal characteristics. The association between breastfeeding and HDL cholesterol was attenuated on accounting for participant BMI. Studies elucidating the mechanisms underlying nutrition in early life and adiposity in later life are warranted.

References

1. Gartner LM, Morton J, Lawrence RA, et al. Breastfeeding and the use of human milk. *Pediatrics.* 2005;115:496-506.
2. Office of Women's Health. US Department of Health and Human Services 2006. Available at: http://www.womenshealth.gov/breastfeeding/index.cfm?page=Campaign. Last accessed October 17, 2006.
3. Primary care interventions to promote breastfeeding: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med.* 2008;149:560-564.
4. Martin RM, Ben Shlomo Y, Gunnell D, et al. Breast feeding and cardiovascular disease risk factors, incidence, and mortality: the Caerphilly study. *J Epidemiol Community Health.* 2005;59:121-129.
5. Owen CG, Whincup PH, Odoki K, et al. Infant feeding and blood cholesterol: a study in adolescents and a systematic review. *Pediatrics.* 2002;110:597-608.
6. Owen CG, Whincup PH, Gilg JA, Cook DG. Effect of breast feeding in infancy on blood pressure in later life: systematic review and meta-analysis. *BMJ.* 2003;327:1189-1195.
7. Owen CG, Martin RM, Whincup PH, et al. The effect of breastfeeding on mean body mass index throughout life: a quantitative review of published and unpublished observational evidence. *Am J Clin Nutr.* 2005;82:1298-1307.
8. Ravelli AC, van der Meulen JH, Osmond C, et al. Infant feeding and adult glucose tolerance, lipid profile, blood pressure, and obesity. *Arch Dis Child.* 2000;82:248-252.
9. Rudnicka AR, Owen CG, Strachan DP. The effect of breastfeeding on cardiorespiratory risk factors in adult life. *Pediatrics.* 2007;119:e1107-e1115.
10. Rich-Edwards JW, Stampfer MJ, Manson JE, et al. Breastfeeding during infancy and the risk of cardiovascular disease in adulthood. *Epidemiology.* 2004;15:550-556.

11. Martin RM, Ness AR, Gunnell D, et al. Does breast-feeding in infancy lower blood pressure in childhood? The Avon Longitudinal Study of Parents and Children (ALSPAC). *Circulation*. 2004;109:1259-1266.
12. Michels KB, Willett WC, Graubard BI, et al. A longitudinal study of infant feeding and obesity throughout life course. *Int J Obes (Lond).* 2007;31:1078-1085. Epub 2007 Apr 24.
13. Dawber TR, Kannel WB. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci.* 1963;107:539-556.
14. Feinleib M, Kannel WB, Garrison RJ, et al. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* 1975;4:518-525.
15. Murabito JM, Evans JC, Larson MG, et al. Family breast cancer history and mammography: Framingham Offspring Study. *Am J Epidemiol.* 2001;154:916-923.
16. Kannel WB, Feinleib M, McNamara PM, et al. An investigation of coronary heart disease in families. The Framingham Offspring Study. *Am J Epidemiol*. 1979;110:281-290.
17. McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. *Clin Chim Acta.* 1987; 166:1-8.
18. 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.
19. 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):51S-209S.
20. WHO Consultation on Obesity. *Obesity: Preventing and Managing the Global Epidemic.* Geneva, Switzerland: World Health Organization: WHO Technical Report Series 894; 2000.
21. Kannel WB, Sorlie P. Some health benefits of physical activity. The Framingham Study. *Arch Intern Med.* 1979;139:857-861.
22. Sherman SE, D'Agostino RB, Cobb JL, Kannel WB. Physical activity and mortality in women in the Framingham Heart Study. *Am Heart J.* 1994;128:879-884.
23. Cupples LA, D'Agostino RB, Sr. Some risk factors related to the annual incidence of cardiovascular disease and death using pooled repeated biennial measurements: Framingham Study, 30-year follow-up. In: Kannel, WB, Wolf PA, Garrison RJ, eds. *The Framingham Heart Study: An Epidemiologic Investigation of Cardiovascular Disease.* Washington, DC: NIH Publication; 1987:87-203.
24. Arenz S, Ruckerl R, Koletzko B, von Kries R. Breast-feeding and childhood obesity—a systematic review. *Int J Obes Relat Metab Disord.* 2004;28:1247-1256.
25. Gillman MW, Rifas-Shiman SL, Berkey CS, et al. Breast-feeding and overweight in adolescence. *Epidemiology.* 2006;17:112-114.
26. Victora CG, Barros F, Lima RC, et al. Anthropometry and body composition of 18 year old men according to duration of breast feeding: birth cohort study from Brazil. *BMJ.* 2003;327:901.
27. Rudnicka AR, Owen CG, Strachan DP. The effect of breastfeeding on cardiorespiratory risk factors in adult life. *Pediatrics.* 2007;119:e1107-e1115.
28. McGuire W, Anthony MY. Formula milk versus term human milk for feeding preterm or low birth weight infants. *Cochrane Database Syst Rev.* 2001:CD002971.
29. Butte NF, Wong WW, Hopkinson JM, et al. Infant feeding mode affects early growth and body composition. *Pediatrics.* 2000;106: 1355-1366.
30. Dewey KG, Heinig MJ, Nommsen LA, et al. Growth of breast-fed and formula-fed infants from 0 to 18 months: the DARLING Study. *Pediatrics.* 1992;89(6 Pt 1):1035-1041.
31. Michaelsen KF, Larsen PS, Thomsen BL, Samuelson G. The Copenhagen Cohort Study on Infant Nutrition and Growth: breast-milk intake, human milk macronutrient content, and influencing factors. *Am J Clin Nutr.* 1994;59:600-611.
32. Coady SA, Jaquish CE, Fabsitz RR, et al. Genetic variability of adult body mass index: a longitudinal assessment in Framingham families. *Obes Res.* 2002;10:675-681.
33. Mayer-Davis EJ, Rifas-Shiman SL, Zhou L, et al. Breast-feeding and risk for childhood obesity: does maternal diabetes or obesity status matter? *Diabetes Care.* 2006;29:2231-2237.
34. Fak F, Friis-Hansen L, Westrom B, Wierup N. Gastric ghrelin cell development is hampered and plasma ghrelin is reduced by delayed weaning in rats. *J Endocrinol.* 2007;192:345-352.
35. Miralles O, Sanchez J, Palou A, Pico C. A physiological role of breast milk leptin in body weight control in developing infants. *Obesity (Silver Spring).* 2006;14:1371-1377.
36. Dundar NO, Anal O, Dundar B, et al. Longitudinal investigation of the relationship between breast milk leptin levels and growth in breast-fed infants. *J Pediatr Endocrinol Metab.* 2005;18:181-187.
37. Singhal A, Farooqi IS, O'Rahilly S, et al. Early nutrition and leptin concentrations in later life. *Am J Clin Nutr.* 2002;75:993-999.
38. Bronsky J, Karpisek M, Bronska E, et al. Adiponectin, adipocyte fatty acid binding protein, and epidermal fatty acid binding protein: proteins newly identified in human breast milk. *Clin Chem.* 2006;52:1763-1770.
39. Singhal A, Cole TJ, Fewtrell M, Lucas A. Breastmilk feeding and lipoprotein profile in adolescents born preterm: follow-up of a prospective randomised study. *Lancet.* 2004;363:1571-1578.
40. Demakakos P, Nazroo J, Breeze E, Marmot M. Socioeconomic status and health: the role of subjective social status. *Soc Sci Med.* 2008;67: 330-340.
41. Maty SC, Lynch JW, Raghunathan TE, Kaplan GA. Childhood socioeconomic position, gender, adult body mass index, and incidence of type 2 diabetes mellitus over 34 years in the Alameda County Study. *Am J Public Health.* 2008;98:1486-1494.
42. Linn S, Fulwood R, Carroll M, et al. Serum total cholesterol: HDL cholesterol ratios in US white and black adults by selected demographic and socioeconomic variables (HANES II). *Am J Public Health.* 1991;81:1038-1043.
43. Kolacek S, Kapetanovic T, Luzar V. Early determinants of cardiovascular risk factors in adults. B. Blood pressure. *Acta Paediatr*. 1993; 82:377-382.
44. Leeson CP, Kattenhorn M, Deanfield JE, Lucas A. Duration of breast feeding and arterial distensibility in early adult life: population based study. *BMJ.* 2001;322:643-647.
45. Lawlor DA, Riddoch CJ, Page AS, et al. Infant feeding and components of the metabolic syndrome: findings from the European Youth Heart Study. *Arch Dis Child.* 2005;90:582-588.
46. Kark JD, Troya G, Friedlander Y, et al. Validity of maternal reporting of breast feeding history and the association with blood lipids in 17 year olds in Jerusalem. *J Epidemiol Community Health.* 1984;38:218-225.
47. Troy LM, Michels KB, Hunter DJ, et al. Self-reported birthweight and history of having been breastfed among younger women: an assessment of validity. *Int J Epidemiol.* 1996;25:122-127.
48. Lande B, Andersen LF, Henriksen T, et al. Relations between high ponderal index at birth, feeding practices and body mass index in infancy. *Eur J Clin Nutr.* 2005;59:1241-1249.
49. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research: one size does not fit all. *JAMA.* 2005;294:2879-2888.
50. Buring JE, O'Connor GT, Goldhaber SZ, et al. Decreased HDL2 and HDL3 cholesterol, Apo A-I and Apo A-II, and increased risk of myocardial infarction. *Circulation.* 1992;85:22-29.
51. Lloyd-Jones DM, Leip EP, Larson MG, et al. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation.* 2006;113:791-798.
52. Calle EE, Thun MJ, Petrelli JM, et al. Body-mass index and mortality in a prospective cohort of U.S. adults. *N Engl J Med*. 1999;341:1097-1105.
53. Kenchaiah S, Evans JC, Levy D, et al. Obesity and the risk of heart failure. *N Engl J Med.* 2002;347:305-313.
54. Martin JA, Hamilton BE, Sutton PD, et al. Births: final data for 2004. *Natl Vital Stat Rep.* 2006;55:1-101.

APPENDIX

	Offspring Women Not Sampled n = 1114	Offspring Women Sampled n = 393	*P* Value
Age, y	37	36	.21
BMI, kg/m^2	24.9	24.4	.45
SBP, mm Hg	117	118	.23
DBP, mm Hg	75	77	<.001
Total cholesterol, mg/dL	193	192	.80
HDL cholesterol, mg/dL	57	57	.67
Current smoking	53%	42%	.001

BMI = body mass index; SBP = systolic blood pressure; DBP = diastolic blood pressure; HDL = high-density lipoprotein.

Cardiovascular disease risk factors among Offspring women with breastfeeding information (n = 393) versus Offspring women not in our sample but with children enrolled in Third Generation-n = 1114), means for continuous variables and percentages where denoted.

nature genetics

Meta-analysis of genome-wide association data identifies two loci influencing age at menarche

John R B Perry[1,23], Lisette Stolk[2–4,23], Nora Franceschini[5,23], Kathryn L Lunetta[6,7,23], Guangju Zhai[8,23], Patrick F McArdle[9,23], Albert V Smith[10,23], Thor Aspelund[10,11], Stefania Bandinelli[12], Eric Boerwinkle[13], Lynn Cherkas[8], Gudny Eiriksdottir[10], Karol Estrada[2], Luigi Ferrucci[14], Aaron R Folsom[15], Melissa Garcia[16], Vilmundur Gudnason[10,11], Albert Hofman[3,4], David Karasik[6,17], Douglas P Kiel[6,17], Lenore J Launer[16], Joyce van Meurs[2,4], Michael A Nalls[18], Fernando Rivadeneira[2–4], Alan R Shuldiner[9], Andrew Singleton[18], Nicole Soranzo[8,19], Toshiko Tanaka[20], Jenny A Visser[2], Michael N Weedon[1], Scott G Wilson[8,21], Vivian Zhuang[7], Elizabeth A Streeten[9,23], Tamara B Harris[16,23], Anna Murray[1,23], Tim D Spector[8,23], Ellen W Demerath[15,23], André G Uitterlinden[2–4,23] & Joanne M Murabito[6,22,23]

We conducted a meta-analysis of genome-wide association data to detect genes influencing age at menarche in 17,510 women. The strongest signal was at 9q31.2 ($P = 1.7 \times 10^{-9}$), where the nearest genes include *TMEM38B, FKTN, FSD1L, TAL2* and *ZNF462*. The next best signal was near the *LIN28B* gene (rs7759938; $P = 7.0 \times 10^{-9}$), which also influences adult height. We provide the first evidence for common genetic variants influencing female sexual maturation.

Menarche is the start of menstruation and occurs at a mean age of approximately 13 years, normally about 2 years after the onset of puberty[1]. Twin and family studies suggest a significant genetic component to menarcheal age, with at least 50% heritability[2–4]. Linkage and candidate gene studies have not confirmed any loci that influence normal variation in age at menarche[4,5]. Genome-wide association (GWA) studies have been successful in identifying many variants associated with complex disease and quantitative traits and we

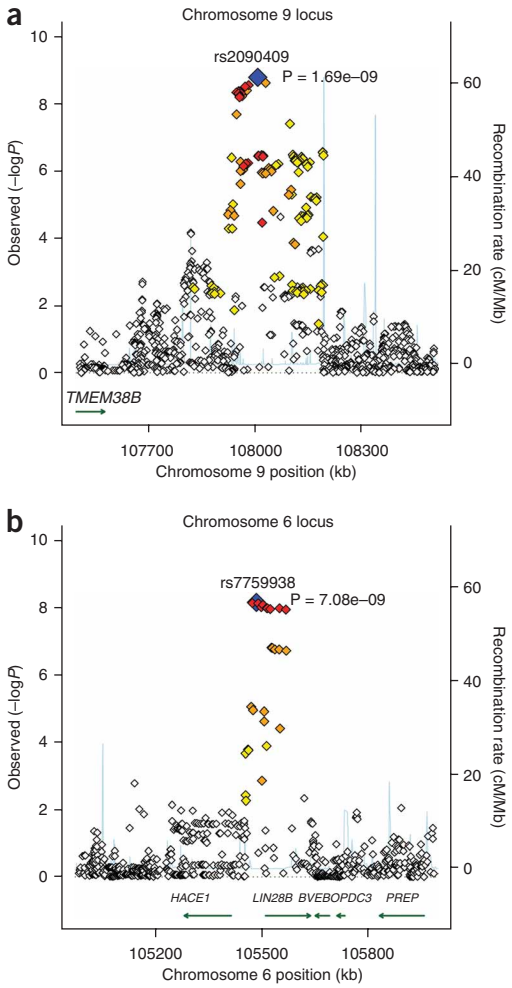


Figure 1 Genomic context (based on NCBI B36) of the top two independent signals at 9q31.2 and 6q21 plotted against association $-\log_{10} P$ values. Only UCSC Refseq genes are shown. r^2 between each SNP and the top signal is color coded: >0.8, red; 0.5–0.8, orange; 0.2–0.5, yellow; <0.2, unfilled. Blue represents SNP with lowest *P* value. Chromosome positions are based on build hg18.

[1]Institute of Biomedical and Clinical Science, Peninsula Medical School, Exeter, UK. [2]Department of Internal Medicine. [3]Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands. [4]Netherlands Genomics Initiative–sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA). [5]Department of Epidemiology, University of North Carolina Gillings School of Global Public Health, Chapel Hill, North Carolina, USA. [6]The National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [7]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [8]Department of Twin Research & Genetic Epidemiology, King's College London, London, UK. [9]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA. [10]Icelandic Heart Association, Kopavogur, Iceland. [11]University of Iceland, Reykjavik, Iceland. [12]Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy. [13]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA. [14]Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, USA. [15]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA. [16]Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA. [17]Hebrew SeniorLife Institute for Aging Research and Harvard Medical School, Boston, Massachusetts, USA. [18]Laboratory of Neurogenetics, National Institute of Aging, Bethesda, Maryland, USA. [19]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [20]Medstar Research Institute, National Institute on Aging, Baltimore, Maryland, USA. [21]School of Medicine & Pharmacology, University of Western Australia and Department of Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. [22]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA. [23]These authors contributed equally to this work. Correspondence should be addressed to A.G.U. (a.g.uitterlinden@erasmusmc.nl).

Received 12 February; accepted 21 April; published online 17 May 2009; doi:10.1038/ng.386

Table 1 Genome-wide significant associations with age at menarche

SNP	Study	*N*	Allele	Frequency	Imputation quality	Effect (years)	s.e.	*P* value
rs2090409	ARIC	4,247	A	0.31	1.00	–0.10	0.04	0.004
	FHS	3,801	A	0.31	0.99	–0.07	0.04	0.07
	RSI	3,175	A	0.34	1.01	–0.08	0.04	0.06
	TwinsUK	2,276	A	0.32	0.99	–0.11	0.03	0.0003
	AGES-Reykjavik	1,849	A	0.27	0.99	–0.08	0.05	0.09
	RSII	1,000	A	0.35	1.06	–0.15	0.07	0.04
	InCHIANTI	597	A	0.33	0.98	–0.26	0.09	0.005
	HAPI Heart Study	565	A	0.28	0.95	0.03	0.10	0.7809
	Meta-analysis	17,510	A	0.31	1.00	–0.10	0.02	1.7×10^{-9}
rs7759938	ARIC	4,247	C	0.33	0.94	0.12	0.04	0.001
	FHS	3,801	C	0.33	0.90	0.10	0.04	0.009
	RSI	3,175	C	0.31	1.00	0.06	0.04	0.16
	TwinsUK	2,276	C	0.32	0.98	0.06	0.03	0.04
	AGES-Reykjavik	1,849	C	0.34	1.00	0.08	0.04	0.07
	RSII	1,000	C	0.30	0.94	0.09	0.08	0.27
	InCHIANTI	597	C	0.29	0.98	0.24	0.09	0.008
	HAPI Heart Study	565	C	0.23	0.91	0.34	0.11	0.002
	Meta-analysis	17,510	C	0.33	0.96	0.09	0.02	7.0×10^{-9}

Meta-analysis *P* values are corrected by individual-study genomic control inflation factors. Alleles are based on forward strand and positions on NCBI build 36. Meta-analysis frequency is calculated as weighted average across all studies. Imputation quality refers to the imputation quality score generated by MACH (oevar) / SNPTEST (proper_info).

therefore used this approach to identify genes involved in determining age at menarche. As earlier age at menarche is associated with shorter stature and obesity, the identified variants may not only clarify the genetic control of female sexual maturation but may also point to regulatory mechanisms involved in normal human growth and obesity.

We carried out a meta-analysis of 17,510 females from eight different population-based cohorts: Age/Gene Environment Susceptibility-Reykjavik Study (AGES-Reykjavik), Atherosclerosis Risk in Communities (ARIC) Study, Framingham Heart Study (FHS), Amish HAPI Heart Study, InCHIANTI Study, Rotterdam Study I and II and TWINS UK Study (**Supplementary Note** online). Women of European descent, with self-reported age at menarche between 9 and 17 years (representing the 1st to 99th percentile, with mean age at menarche of 13.12 (s.d. 1.5) years), were included. Agreement between adult-recalled and prospectively collected age at menarche is reported to be good (κ statistic = 0.81)[6]. Each study conducted a GWA analysis using linear regression or linear mixed-effects models with an additive genetic model adjusting for birth year or birth cohort (FHS), with additional adjustments for population structure when appropriate. Approximately 2.55 million autosomal SNPs, imputed with reference to the HapMap CEU panel, passed quality control criteria. We then conducted a meta-analysis using a fixed-effects model based on inverse variance weighting. Full details of cohorts and methods are given in **Supplementary Table 1** and **Supplementary Methods** online.

Twenty-eight SNPs passed the conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$ and were at either 9q31.2 or 6q21 (**Supplementary Table 2** and **Supplementary Fig. 1** online). The 18 SNPs on chromosome 9 were in linkage disequilibrium (LD), with $r^2 > 0.31$, as were the 10 SNPs on chromosome 6, with $r^2 > 0.96$ (**Fig. 1** and **Supplementary Table 2**). To identify more than one signal that could account for the association findings, we carried out conditional analysis adjusting for the SNP with the lowest *P* value in the region (rs7759938 for chromosome 6 and rs2090409 for chromosome 9). Within 1 Mb flanking each SNP, the lowest adjusted *P* values for association with age at menarche were $P = 0.0017$ and $P = 0.0077$ for chromosomes 9 (1,030 SNPs) and 6 (775 SNPs), respectively. These findings suggest a single signal accounting for the associations at each locus. The quantile-quantile plot (**Supplementary Fig. 2** online) showed modest deviation away from the null when these top two signals were removed, suggesting the presence of additional loci for this trait.

The strongest signal at 9q31.2 was observed with rs2090409, where each A allele was associated with approximately a 5-week reduction in menarcheal age ($P = 1.7 \times 10^{-9}$). All studies showed consistent evidence of association with the same direction of effect in all but one study, similar effect sizes and *P* values between 0.8 and 0.0003 (**Table 1**). The recombination region containing rs2090409 includes only a hypothetical gene (*BC039487*). Outside of this, the only RefSeq gene within a 1-Mb window is a transmembrane protein gene, *TMEM38B*, which is approximately 400 kb proximal to the GWAS signal. In mice, *TMEM38B* is expressed strongly in brain and the null mutation is neonatal lethal. Within 2 Mb of the signal, genes include *SLC44A1*, *FKTN*, *FSD1L*, *TAL2* and *ZNF462*, none of which is an obvious candidate gene for involvement in menarche. However, a SNP in *ZNF462*, 650 kb from our signal but not in LD ($r^2 = 0.086$), has been previously associated with variation in height[6].

The 6q21 signal was within a recombination interval that included only one gene, *LIN28B* (**Fig. 1**) and was also associated with approximately a 5-week reduction in menarcheal age per T allele (rs7759938; $P = 7.0 \times 10^{-9}$). The effect was consistent across all studies, with *P* values between 0.27 and 0.001 (**Table 1**). A common variant in the *LIN28B* gene has previously been associated with normal variation in adult height[7]. The most significant menarche-associated variant (rs7759938) and the previously reported height variant (rs314277) lie within 28.7 kb of each other and are likely to represent the same signal, as $r^2 = 0.26$ and $D' = 1$ in HapMap. The allele associated with earlier age at menarche is associated with decreased height, which is consistent with epidemiological data. Early menarche has been correlated with reduced stature, and the mechanism is probably mediated through earlier exposure to estrogens resulting in earlier closure of the epiphyseal plates[8]. We therefore tested all

published common variants influencing height—44 independent loci—for association with age at menarche in our dataset[9]. Six of the alleles were also associated with menarcheal age ($P < 0.05$), with the strongest associations at *LIN28B* ($P = 0.0001$) and *PXMP3* ($P = 0.003$) (**Supplementary Table 3** online). We also tested the association of the newly identified menarche-associated variant, rs7759938, with measured height in our study population (**Supplementary Methods**) and found that it was associated with height, $P_{meta} = 0.0001$, in the same direction in all but one study; that is, the C allele was associated with reduction in age at menarche and also reduced stature. The published height SNP (rs314277) did not reach nominal significance with height in our study ($P_{meta} = 0.26$). These data suggest that some of the previously identified loci that influence adult height may also have a general role in adolescent growth.

At a given chronologic age, girls with earlier age at menarche tend to have greater body mass index (BMI) and adiposity than girls with a later age at menarche[10–12]. A marked secular decline in age at menarche occurred in Europe in the nineteenth and early twentieth centuries, which has been attributed to improved nutrition and health[1]. This trend may be continuing as a consequence of the obesity epidemic[13] and may involve a common metabolic response to the current nutritional environment[14] or be attributable, at least in part, to shared genetic influences or pleiotropy[15]. We therefore investigated the effect on menarcheal age of the ten currently known common gene variants associated with variation in BMI. Of these ten loci, eight showed an association in the direction consistent with epidemiological data ($P = 1.6 \times 10^{-6}$, based on Fisher's combined probability test: $-2 \times \text{sum}(\ln P)$ against χ^2 on (10×2) df), and five were nominally significant ($P < 0.05$) (**Supplementary Table 4** online). The two loci with the largest observed effects on BMI (*FTO* and *TMEM18*) also had the strongest evidence for association with menarcheal age ($P = 0.0008$ and 7.0×10^{-5}, respectively).

This study provides the first evidence for common genetic variants influencing normal variation in the timing of female sexual maturation. Our findings also indicate a genetic basis for the phenotypic associations between age at menarche and both height and BMI.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are indebted to the participants and staff of the studies for their important contributions. Full individual study acknowledgements are listed in the **Supplementary Methods**.

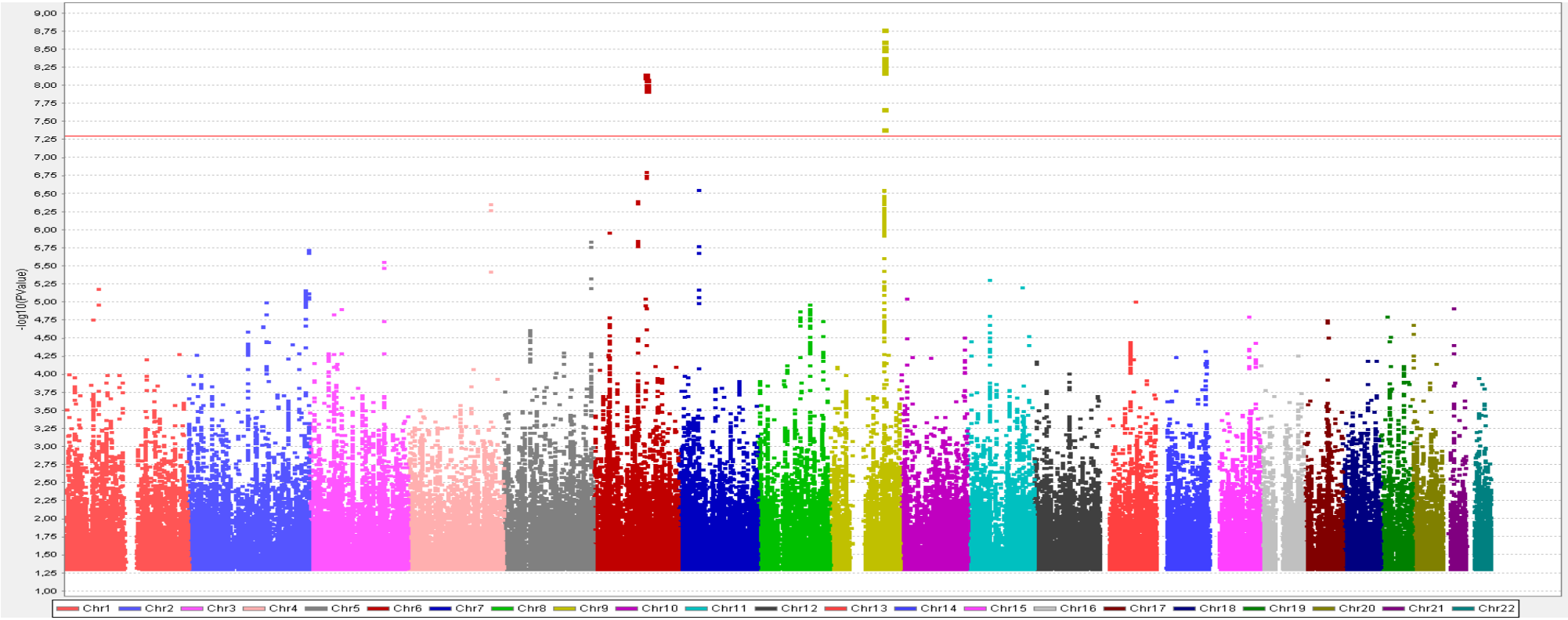
AUTHOR CONTRIBUTIONS

Statistical analyses: J.R.B.P., L.S., N.F., K.L.L., G.Z., P.F.M., A.V.S., T.A., K.E., V.G., F.R., N.S., T.T., M.N.W. and V.Z. Sample collection, preparation: N.F., A.V.S., T.A., S.B., E.B., L.C., G.E., L.F., A.R.F., V.G., A.H., S.G.W., T.B.H., T.D.S., E.W.D. and A.G.U. Genotyping: E.B., F.R., A.S., N.S. and A.G.U. Manuscript writing: J.R.B.P., L.S., P.F.M., A.V.S., T.A., G.E., V.G., E.A.S., A.M., E.W.D. and A.G.U. Review and revision of the manuscript: J.R.B.P., L.S., N.F., K.L.L., G.Z., P.F.M., A.V.S., T.A., S.B., E.B., L.C., G.E., K.E., L.F., A.R.F., M.G., V.G., A.H., D.K., D.P.K., L.J.L., J.v.M., M.A.N., F.R., A.R.S., A.S., N.S., T.T., J.A.V., M.N.W., S.G.W., V.Z., E.A.S., T.B.H., A.M., T.D.S., E.W.D., A.G.U. and J.M.M. Intellectual input: J.R.B.P., L.S., N.F., K.L.L., P.F.M., E.B., G.E., A.R.F., V.G., D.K., D.P.K., J.A.V., S.G.W., A.M., A.G.U. and J.M.M. Study design: J.R.B.P., L.S., K.L.L., G.Z., A.V.S., T.A., S.B., G.E., L.F., V.G., A.H., D.K., D.P.K., A.M., T.D.S., A.G.U. and J.M.M.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

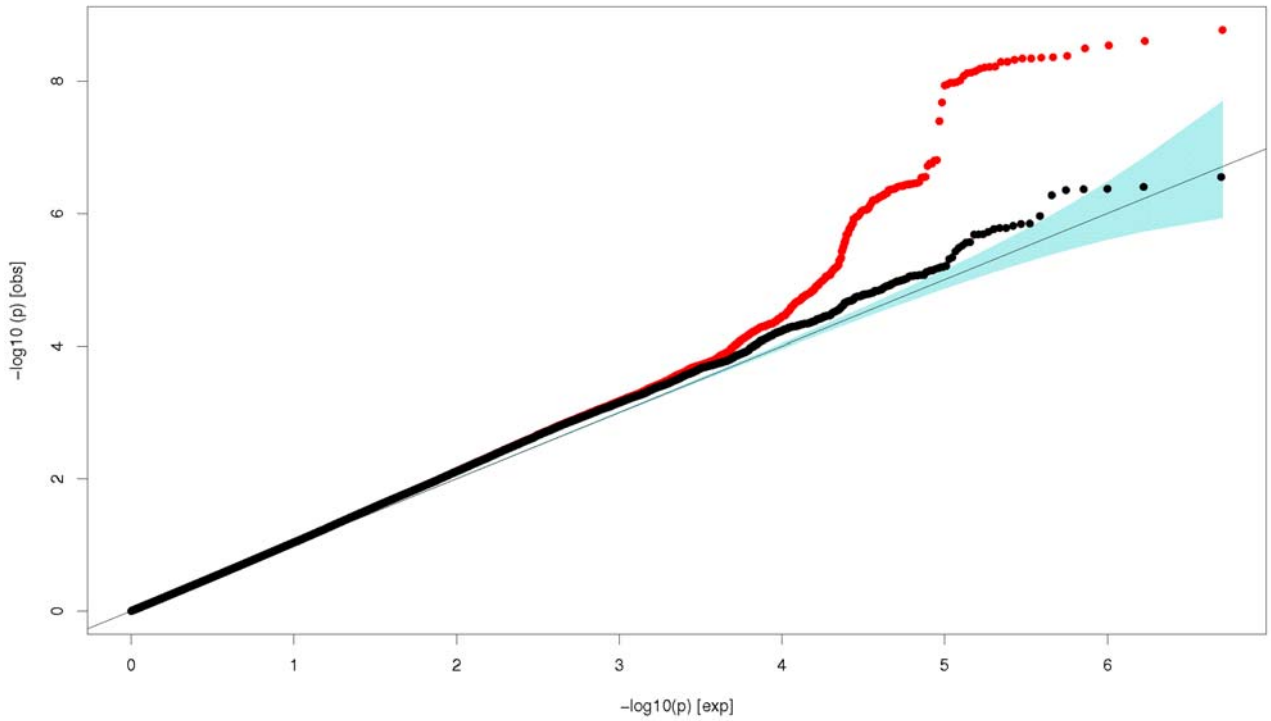
1. Marshall, W.A. & Tanner, J.M. *Arch. Dis. Child.* **44**, 291–303 (1969).
2. Towne, B. *et al. Am. J. Phys. Anthropol.* **128**, 210–219 (2005).
3. Snieder, H., MacGregor, A.J. & Spector, T.D. *J. Clin. Endocrinol. Metab.* **83**, 1875–1880 (1998).
4. Anderson, C.A. *et al. J. Clin. Endocrinol. Metab.* **93**, 3965–3970 (2008).
5. Gajdos, Z.K.Z. *et al. J. Clin. Endocrinol. Metab.* **93**, 4290–4298 (2008).
6. Bosetti, C., Tavani, A., Negri, E., Trichopoulos, D. & La Vecchia, C. *J. Clin. Epidemiol.* **54**, 902–906 (2001).
7. Lettre, G. *et al. Nat. Genet.* **40**, 584–591 (2008).
8. Onland-Moret, N.C. *et al. Am. J. Epidemiol.* **162**, 623–632 (2005).
9. Weedon, M.N. & Frayling, T.M. *Trends Genet.* **24**, 595–603 (2008).
10. Anderson, S.E., Dallal, G.E. & Must, A. *Pediatrics* **111**, 844–850 (2003).
11. Garn, S., LaVelle, M., Rosenberg, K. & Hawthorne, V. *Am. J. Clin. Nutr.* **43**, 879–883 (1986).
12. van Lenthe, F., Kemper, C. & van Mechelen, W. *Am. J. Clin. Nutr.* **64**, 18–24 (1996).
13. Euling, S.Y. *et al. Pediatrics* **121**, S172–S191 (2008).
14. Harris, M.A., Prior, J.C. & Koehoorn, M. *J. Adolesc. Health* **43**, 548–554 (2008).
15. Wang, W., Zhao, L.-J., Liu, Y.-Z., Recker, R.R. & Deng, H.-W. *Int. J. Obes. (Lond.)* **30**, 1595–1600 (2006).

Supplementary Figure 1. . Manhattan plot of all association $-\log_{10}$ P values ordered by chromosome, line indicates the genome-wide significance threshold of 5×10^{-8}.



Nature Genetics: doi:10.1038/ng.386

Supplementary figure 2. A quantile-quantile plot of the 2.55 million imputed SNPs on age at menarche (red dots) and excluding the 2 SNPs which reached genome-wide significance and all SNPs within 500kb upstream and downstream of those top signals (black dots). 95% confidence interval is shown in pale blue.



Nature Genetics: doi:10.1038/ng.386

Supplementary Table 1: Cohort genotype and phenotype descriptions

				before imputation				after imputation			
Study	**N**	**mean age at menarche (SD)**	**Array**	**Callrate cut-off**	**MAF cut-off**	**HWE P-value cut-off**		**number of SNPs**	**Imputation Program**	**Analysis program**	**Lambda**
ARIC	4247	12.9 (1.54)	Affymetrix 6.0	0.95	0.01	1*10-5		2,423,704	MACH* v1.0.16	ProABEL	1.03
FHS	3801	12.8 (1.5)	Affymetrix 500K + Affymetrix 50K	0.97	0.01	1*10-6		2,529,104	MACH v1.0.15	R packages Kinship, LME	1.01
RSI	3175	13.5 (1.6)	Illumina HumanHap 550K	0.98	0.01	1*10-6		2,542,887	MACH	MACH2QTL	1.03
TwinsUK	2276	12.99(1.55)	Illumina HumanHap 317K	0.95, if MAF > 0.05 0.99, if 0.01 < MAF < 0.05	0.01	5.7*10-7		2,544,233	IMPUTE	GenABEL	1.01
AGES	1849	13.6(1.30)	Illumina Hu370CNV	0.97	0.01	1*10-6		2,532,729	MACH	ProbABEL	1.03
RSII	1000	13.3 (1.6)	Illumina HumanHap 550K	0.98	0.01	1*10-6		2,542,887	MACH	MACH2QTL	1.01
InCHIANTI	597	13.3 (1.5)	Illumina HumanHap 550K	0.98	0.01	1*10-4		2,565,135	IMPUTE	SNPTEST	1.04
HAPI Heart Study	565	13.1 (1.3)	Affymetrix 500K	0.95	0.01	1×10^{-6}		2,543,014	MACH	ITSNBN (in-house)	1.04

Imputation Backbone (NCBI build): phased CEU haplotypes, HapMap release 21 (build 35)

Nature Genetics: doi:10.1038/ng.386

Supplementary Table 2: All SNP associations for age at menarche with $P < 5x10^{-8}$

Chr	Position	SNP	Effect Allele	Other Allele	Effect Freq	Effect size (Years)	SE	P value	N	Direction	Imputation QC	Directly genotyped
9	108006909	rs2090409	A	C	0.3142	-0.0967	0.0161	1.687E-09	17510	-------+	1 (0.95 - 1.06)	--------
9	108029220	rs1516890	A	G	0.479	-0.0821	0.0138	2.483E-09	17510	--------	0.97 (0.89 - 1.01)	--+-----
9	107981330	rs10156597	A	T	0.6868	0.0955	0.0161	2.893E-09	17510	+++++++-	1 (0.95 - 1.06)	--------
9	107971553	rs2138628	A	T	0.3136	-0.0952	0.0161	3.204E-09	17510	-------+	1 (0.93 - 1.06)	--------
9	107976495	rs7861820	T	C	0.5225	0.0794	0.0135	4.163E-09	17510	++++++++	0.99 (0.93 - 1.02)	++--+++-
9	107953878	rs2008393	A	G	0.6886	0.0963	0.0164	4.351E-09	17510	+++++++-	0.99 (0.87 - 1.06)	--------
9	107952875	rs1516881	C	G	0.6887	0.0964	0.0164	4.421E-09	17510	+++++++-	0.98 (0.87 - 1.06)	--------
9	107957087	rs12352703	T	G	0.3139	-0.0948	0.0162	4.549E-09	17510	-------+	0.99 (0.9 - 1.06)	--+-----
9	107952233	rs9299121	A	T	0.3115	-0.0964	0.0164	4.554E-09	17510	-------+	0.98 (0.86 - 1.06)	--------
9	107947088	rs1516883	A	G	0.3117	-0.0963	0.0164	4.772E-09	17510	-------+	0.98 (0.86 - 1.05)	--------
9	107960041	rs10453225	T	G	0.3162	-0.0938	0.016	5.096E-09	17510	-------+	0.99 (0.91 - 1.06)	--------
9	107952732	rs1516882	A	G	0.6893	0.096	0.0164	5.139E-09	17510	+++++++-	0.98 (0.87 - 1.06)	++--+---
9	107963278	rs2417687	A	G	0.5225	0.0793	0.0136	6.029E-09	17510	++++++++	0.99 (0.96 - 1.01)	--++---+
9	107956532	rs12686569	T	G	0.3115	-0.0952	0.0164	6.111E-09	17510	-------+	0.98 (0.88 - 1.06)	+++-+---
9	107956893	rs2222133	T	C	0.688	0.095	0.0164	6.192E-09	17510	+++++++-	0.99 (0.89 - 1.06)	++--+---
9	107955830	rs10978430	T	C	0.3119	-0.095	0.0164	6.517E-09	17510	-------+	0.99 (0.88 - 1.06)	--------
6	105485647	rs7759938	T	C	0.678	-0.0942	0.0163	7.081E-09	17510	--------	0.96 (0.9 - 1)	--------
6	105474309	rs9391253	A	T	0.6729	-0.0929	0.0161	7.425E-09	17510	--------	0.97 (0.9 - 1.01)	--------
6	105490671	rs2095812	C	G	0.673	-0.0927	0.016	7.545E-09	17510	--------	0.97 (0.92 - 1.02)	--------
6	105504111	rs395962	T	G	0.3267	0.0924	0.016	8.341E-09	17510	++++++++	0.97 (0.93 - 1.02)	--------
6	105499438	rs314263	T	C	0.6732	-0.092	0.016	9.793E-09	17510	--------	0.97 (0.93 - 1.02)	++--+---
6	105514692	rs314276	A	C	0.3299	0.0907	0.0158	1.035E-08	17510	++++++++	0.99 (0.97 - 1.02)	--++---+
6	105516741	rs167539	A	C	0.6699	-0.0906	0.0158	1.056E-08	17510	--------	0.99 (0.97 - 1.02)	--------
6	105550751	rs369065	T	C	0.6687	-0.0909	0.0159	1.056E-08	17510	--------	0.98 (0.93 - 1.02)	+++-+---
6	105524671	rs314268	A	G	0.6696	-0.0905	0.0158	1.118E-08	17510	--------	0.99 (0.97 - 1.02)	++--+---
6	105568575	rs314273	T	G	0.3315	0.0908	0.0159	1.162E-08	17510	++++++++	0.98 (0.92 - 1.01)	--------
9	107947467	rs9409084	A	G	0.6018	0.0936	0.0167	2.105E-08	13263	++?++++-	0.96 (0.86 - 1.02)	--?-----
9	108097268	rs966523	C	G	0.3951	-0.0817	0.0149	4.041E-08	17510	--------	0.98 (0.94 - 1)	--++---+

Meta - analysis performed after adjusting each study by its genomic control inflation factor. Alleles based on forward strand, positions on NCBI build 36. Effect Freq is calculated as the weighted allele frequency across all studies for effect allele. Impute QC is the min-max (avg) imputation quality

Nature Genetics: doi:10.1038/ng.386

score across all cohorts. Direction of effect and whether the SNP was directly genotyped (+) or imputed (-) is given for each study in following order: RSI, RSII, ARIC, FHS, InChianti, Twins, AGES, Hapi Heart study, ? indicates where data were not available.

Nature Genetics: doi:10.1038/ng.386

Supplementary table 3: Effects of 44 height loci on age at menarche

Gene	SNP	Chr	Position	Height Increasing Allele	Menarche Increasing Allele	Effect (Years)	SE	P value	Direction
LIN28B	rs314277	6	105514355	A	A	0.0963	0.0252	0.0001	++++++++
PXMP3	rs7846385	8	78322734	C	T	0.0493	0.0167	0.003	+-++++++
C6orf173	rs4549631	6	127008001	C	T	0.0299	0.0133	0.02	+-++++++
GNA12	rs798544	7	2729628	C	C	-0.035	0.0164	0.03	--------
DYM	rs8099594	18	45245158	A	A	0.032	0.0156	0.04	+-++++++
SCMH1	rs6686842	1	41303458	T	C	-0.0281	0.014	0.04	------++
ZBTB38	rs6440003	3	142576899	A	A	0.0264	0.0141	0.06	-+++++++
CABLES1	rs4800148	18	18978326	A	G	-0.0345	0.0196	0.08	++----+-
ADAMSTSL3	rs2562784	15	82077496	G	A	0.0319	0.0191	0.10	+++++++-
ANAPC13	rs10935120	3	135715782	G	G	-0.0248	0.0157	0.11	----++-+
DNM3	rs678962	1	170456512	G	T	0.0293	0.0188	0.12	-++--+-+
SOCS2	rs11107116	12	92502635	T	T	0.027	0.0185	0.15	-+++--++
HLA Class III	rs185819	6	32158045	T	C	-0.0152	0.0134	0.26	+-+-----
ACAN	rs8041863	15	87160693	A	T	-0.0155	0.0142	0.27	-+---+--
PLAG1	rs10958476	8	57258362	C	C	-0.023	0.0212	0.28	++--+---
TBX2	rs757608	17	56852059	A	G	-0.016	0.0158	0.31	-+--++-+
PPARD	rs4713858	6	35510763	G	G	-0.0254	0.0251	0.31	+-?--++-
HMGA2	rs1042725	12	64644614	C	C	-0.0137	0.0137	0.31	---+--++
HMGA1	rs1776897	6	34302989	G	T	0.0329	0.0332	0.32	+-?++++-
Histone cluster 1	rs10946808	6	26341366	A	A	0.0161	0.0169	0.34	+++--+-+
DLEU7	rs3116602	13	50009356	T	G	-0.0178	0.0191	0.35	-+---+-+
Histone cluster 2	rs11205277	1	148159496	G	A	0.0136	0.0151	0.37	-++++++-
ADAMSTS17	rs4533267	15	98603794	A	G	-0.0135	0.0166	0.41	--+--+-+
ZNF462	rs4743034	9	108672174	A	A	0.0133	0.0188	0.48	++--++--
BMP6	rs12198986	6	7665058	A	A	0.0084	0.0138	0.54	+-+--+-+
PPARD	rs2814993	6	34726871	A	G	-0.0136	0.0226	0.55	--+-+-++
HLA Class III	rs2844479	6	31680935	A	C	-0.0095	0.0159	0.55	+---+-+-
EFEMP1	rs3791679	2	55950396	A	G	-0.0106	0.0183	0.56	-+++----
FBLN5	rs8007661	14	91529711	C	T	0.0102	0.0192	0.59	++--+++-
GDF5	rs6060369	20	33370575	C	T	0.0077	0.0149	0.60	-+++--+-
NCAPG	rs16896068	4	17553938	G	G	-0.0109	0.0217	0.61	--+-+-++
PTCH1	rs10512248	9	97299524	G	G	-0.0076	0.016	0.63	+++-++--
IHH	rs6724465	2	219652090	G	G	-0.0125	0.0267	0.64	-++---+-

Nature Genetics: doi:10.1038/ng.386

GPR126	rs4896582	6	142745570	G	G	-0.0069	0.0167	0.68	--++--+-
HHIP	rs1812175	4	145794294	G	G	-0.0086	0.0211	0.68	--+-+-+-
CDK6	rs2282978	7	92102346	C	C	-0.0041	0.0156	0.79	++---+++
ZNF678	rs1390401	1	225864573	A	G	-0.005	0.0199	0.80	-+++--++
NOG	rs4794665	17	52205328	A	G	-0.0024	0.0133	0.86	+--++--+
RNF135	rs3760318	17	26271841	G	G	-0.0021	0.0151	0.89	++---+--
BMP2	rs967417	20	6568893	G	G	-0.0014	0.0136	0.92	+---+---
SPAG17	rs12735613	1	118685496	G	A	0.0015	0.0182	0.93	++-+-+-+
DOT1L	rs12986413	19	2121954	T	A	0.0011	0.0136	0.93	+-++---+
PLAG1	rs9650315	8	57318152	G	G	-0.0017	0.0242	0.94	-+--+-+-
TSEN15	rs2274432	1	182287568	A	G	-0.0004	0.0155	0.98	-++++-+-

Effect sizes in years, P-values corrected by individual study genomic control inflation factors. Alleles based on forward strand, positions on NCBI build 36. Impute QC is the min-max (avg) imputation quality score across all cohorts. Direction of effect is given for each study in following order: RSI, RSII, ARIC, FHS, InChianti, Twins, AGES, Hapi Heart study, ? indicates where data were not available.

Nature Genetics: doi:10.1038/ng.386

Supplementary table 4: Effects of 10 BMI genes on age at menarche

SNP	Chr	Position	Nearby Gene	BMI Increasing allele	Menarche P	N	Effect (Years)	SE	Directions
rs6548238	2	624905	TMEM18	C	7x10-5	13263	-0.1	0.02	++?+++++
rs9939609	16	52378028	FTO	A	0.0008	17510	-0.05	0.01	-+--+---
rs4074134	11	27603861	BDNF	C	0.006	17510	-0.05	0.02	++++++++
rs2815752	1	72524461	NEGR1	A	0.02	17510	-0.03	0.01	-------+
rs10938397	4	45023455	GNPDA2	G	0.04	17510	-0.03	0.02	-+++-++-
rs11084753	19	39013977	KCTD15	G	0.21	17510	-0.02	0.02	++++-+--
rs7498665	16	28790742	SH2B1	G	0.41	17510	-0.01	0.02	++++--++
rs7647305	3	187316984	SFRS10	C	0.48	17510	-0.01	0.02	++++---+
rs17782313	18	56002077	MC4R	C	0.82	17510	0.004	0.02	++--+-+-
rs10838738	11	47619625	MTCH2	G	0.96	17510	0.0009	0.02	--+-++-+

Effect sizes in years, based on BMI increasing allele. P-values corrected by individual study genomic control inflation factors. Alleles based on forward strand, positions on NCBI build 36. Impute QC is the min-max (avg) imputation quality score across all cohorts.
Direction of effect is given for each study in following order: RSI, RSII, ARIC, FHS, InChianti, Twins, AGES, Hapi Heart study? indicates where data were not available.

Nature Genetics: doi:10.1038/ng.386

Supplementary Methods

Phenotype Definition

AGES: A reproductive history questionnaire was administered to all women at the entry into the AGES Reykjavik study (2002-2006). The question asked to determine age at menarche was: "*At what age did your menstrual periods begin?*" Women answering between 9 and 17 years were included in the analyses. Height was measured at their first visit to the Reykjavik Study (1967 – 1997) when subjects were a mean age of approximately 50 years.

ARIC: A reproductive history questionnaire was administered to all women at the baseline (1987-1989) visit. The following question was asked to determine age at menarche: "*At approximately what age were you when your menstrual periods started?*" Responses were given to the closest whole year value (11, 12, 13, etc). Mean age at menarche among ARIC White women was 12.88 (SD 1.54). Height was obtained during the baseline clinic visit by trained technicians.

FHS: At the second Offspring examination (1979 to 1982) and at the first Third Generation examination (2002 to 2005), women were asked, "*Age at start of menses*" and "*How old were you when you had your first menstrual period (menses)?"* respectively. The self-reported age at first period was recorded. Offspring women were asked again about menarche at the time of participation in the Framingham Osteoporosis Study (1996 – 2001: *"About how old were you when you had your first menstrual period?"*). If menarche data were missing from Offspring examination two, the self-reported data from the Osteoporosis examination was used (n=214). There were 1777 Offspring Cohort and 2024 Third Generation women who reported an age at menarche between 9 and 17 years with genotyping available. The mean age at menarche was 12.8 years (SD 1.5 years) in the combined Offspring and Third Generation women in the sample. Height was measured by trained technicians at the first Offspring and Third Generation examinations.

HAPI Heart Study: During the baseline visit, women were given a reproductive history questionnaire. The self reported age at first period was recorded from the question "*How old were you when you had your first menstrual period?*"

InCHIANTI: During baseline visit women were asked "*How old were you when you had your first menstrual period?*"

RSI and RSII: At the first RSI interview (1989-1993) and at the first RSII interview (200-2001) women were asked "*How old were you when you had your first menstrual period?*". Self-reported age at menarche was available for 3,175 (RSI) and 1,000 (RSII) women. Mean age at menarche for the RSI cohort was 13.5 (SD 1.6 years) and 13.3 (SD 1.5 years) for RSII.

TwinsUK: Data on age at menarche was obtained by self-administered questionnaire. All females from the TwinsUK cohort were asked a question *"how old were you when you had your first menstrual period?"* There were 5523 female twins reporting their age at menarche. Of them, 2276 females of European descent (458 MZ pairs, 548 DZ pairs, and 264 singletons) had genotyping data available and included in the analysis. The mean age at menarche was 12.99 years with SD=1.55 (range 9-17).

Nature Genetics: doi:10.1038/ng.386

Genotyping and Imputation

There were four different genotyping platform used by the eight cohort studies: Illumina Human CNV 370 (AGES)/HumanHap 317K (TwinsUK), the Affymetrix Genome-Wide Human SNP Array 6.0 (ARIC), the Affymetrix 500K mapping array (HAPI) and the Affymetrix 500K in combination with the 50K supplemental array (FHS) and the Illumina Infinium II Human Hap 550 SNP chip array (InCHIANTI, RSI, RSII). Each study performed genotyping quality control checks based on duplicate sample genotyping, SNP call rate, Hardy-Weinberg equilibrium, Mendelian inconsistencies, and sex mismatch, and principle components methods were used to evaluate the presence of population stratification (details provided in **Supplementary Table 1**). Because there were only about 55,000 overlapping SNPs from the four genotyping platforms, each study imputed 2.5 million HapMap SNPs for each participant using currently available imputation methods. InCHIANTI and TwinsUK used IMPUTE (http://www.stats.ox.ac.uk/~marchini/software/gwas/impute) and all other cohorts used the MACH algorithm (http://www.sph.umich.edu/csg/ abecasis/MaCH/). All studies imputed the genotype "dosage" (0, 1, 2) for the expected number of minor alleles. Imputation quality was determined by either the r^2 value produced by MACH or calculated empirical variance divided by the expected variance (oevar) and for SNPTEST the 'proper info' output variable was used to determine imputation quality. SNP imputation methods and quality control procedures for each cohort are included in **Supplementary Table 1**.

Statistical Analysis

AGES: Analysis was preformed using linear regression against the imputed genotype dosage with the ProbABEL package. Birthyear was included as a covariate.

ARIC: Population stratification was estimated using principal component methods (EIGENSTRAT)(1), after removing few related individuals. Two principal components were significantly associated with age at menarche in linear regression models (alpha=0.05) and so they were included, along with year of birth and study center, as covariates in the genetic analyses. We used linear regression models and assumed additive genetic effects to study the association of imputed and genotyped SNPs (dosage data) and age of menarche. The analyses were implemented in the ProbABEL package from the ABEL set of programs (http://mga.bionet.nsc.ru/ yurii/ABEL/)(2).

FHS: SNP weights for 10 principal components (PCs) were inferred using a maximal set of independent individuals; the PCs for the remaining individuals were computed using the SNP weights obtained from the unrelated set of individuals. The first PC (PC1) was significantly associated with age at menarche ($P<0.01$), and therefore was included as a covariate in all SNP association analyses. In addition, we adjusted for birth cohort by decade. Linear mixed effects models were used to account for familial correlations. Each SNP was tested for association with age at menarche using an additive genetic model.

HAPI Heart Study: Analysis was performed using in house developed software. In brief, we performed a measured genotype approach utilizing a t-test of the beta coefficient for the SNP variable. We included birth year as a fixed covariates in the model and a polygenic component modeled as a random effect to account for the full 13-14 generation pedigree of the Amish. A total of 338,598 autosomal SNPs were used for imputation after applying filters: (1) not in HapMap, (2) frequency < 0.01, (3) Hardy-Weinberg p-value < 1×10^{-6}, and (4) missingness > 0.05.

Nature Genetics: doi:10.1038/ng.386

InCHIANTI: Analysis performed using linear regression allele dosage in SNPTEST (http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest). Birthyear was included as a covariate.

RSI and RSII: Adjusted linear regression analysis was done using MACH2QTL (http://www.sph.umich.edu/csg/ abecasis/MaCH/), birthyear was included as a covariate.

TwinsUK: Because of the relatedness in the TwinsUK cohort, we utilized the GenABEL software package(2) which is designed for GWAS analysis of family-based data by incorporating pair-wise kinship matrix calculated using genotyping data in the polygenic model to correct relatedness and hidden population stratification. The score test implemented in the software was used to test the association between a given SNP and the age at menarche with adjustment for birth-year as a covariate.

Meta-analysis: Menarche GWA

Inverse variance meta-analysis of the 8 studies was performed using the latest version of METAL (http://www.sph.umich.edu/csg/abecasis/Metal/index.html). A SNP within a study was not included if the minor allele frequency (MAF) was < 1% or imputation quality score was < 0.4 for SNPTEST or < 0.3 for MACH in that study. Genomic control was applied to the meta-analysis in METAL to correct for relatedness and population stratification (http://www.sph.umich.edu/csg/abecasis/metal) (**Supplementary Table 1)**. The meta-analysis included 2,551,160 autosomal, QC'd, SNPs and 17,510 samples.

Meta-analysis: Height SNPs

Association statistics for measured adult height were calculated for rs2090409, rs314277 and rs7759938 in the same study samples used for the age at menarche genome wide analysis. The total number with measured height of the 17,510 menarche samples was 16,371. The fixed-effects weighted meta-analysis was performed in METAL.

Acknowledgements

AGES: The Age, Gene/Environment Susceptibility Reykjavik Study is funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The authors thank the participants and the staff at the IHA for their contribution. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

ARIC: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022. The authors thank the staff and participants of the ARIC study for their important contributions.

Nature Genetics: doi:10.1038/ng.386

FHS: This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially 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). A portion of this research utilized 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. The Framingham Heart Study phenotype-genotype analyses were supported by the National Institute of Aging (Genetics of Reproductive Life Period and Health Outcomes, R21AG032598). R01 AR/AG 41398 funded some of the menarche data collection.

HAPI Heart Study: This work was supported by the National Institute of Health [grant number U01 HL72515]; the University of Maryland General Clinical Research Center [grant number M01 RR 16500); the Clinical Nutrition Research Unit of Maryland [grant number P30 DK072488].

InCHIANTI: The InCHIANTI study baseline (1998-2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336); the InCHIANTI Follow-up 1 (2001-2003) was funded by the U.S. National Institute on Aging (Contracts: N.1-AG-1-1 and N.1-AG-1-2111); the InCHIANTI Follow-ups 2 and 3 studies (2004-2010) were financed by the U.S. National Institute on Aging (Contract: N01-AG-5-0002);supported in part by the Intramural research program of the National Institute on Aging, National Institutes of Health, Baltimore, Maryland.

RSI and RSII: The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810, and funding from the European Commision (HEALTH-F2-2008-201865, GEFOS; HEALTH-F2-2008-35627, TREAT-OA) . The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. We thank Pascal Arp, Mila Jhamai, Dr Michael Moorhouse, Marijn Verkerk, and Sander Bervoets for their help in creating the GWAS database.

TwinsUK: TwinsUK is supported by the Wellcome Trust from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. We thank the staff from the TwinsUK, the DNA Collections and Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation; Quality Control of the Twins UK cohort for genotyping (in particular Amy Chaney, Radhi Ravindrarajah, Douglas Simpkin, Cliff Hinds, and Thomas Dibling); Paul Martin and Simon Potter of the DNA and Genotyping Informatics teams for

Nature Genetics: doi:10.1038/ng.386

data handling; Le Centre National de Génotypage, France, led by Mark Lathrop, for genotyping; Duke University, North Carolina, USA, led by David Goldstein, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki, led by Aarno Palotie.

Administrative: The Longevity Consortium, funded by the National Institute of Aging, grant number U19 AG023122, provided administrative resources for this work.

References

[1] 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 Aug;38(8):904-9.

[2] Aulchenko Y.S., Ripke S., Isaacs A., van Duijn C.M. GenABEL: an R package for genome-wide association analysis. *Bioinformatics.* (2007) 23(10):1294-6

[3] Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, Sigurdsson G, Thorgeirsson G, Aspelund T, Garcia ME, Cotch MF, Hoffman HJ, Gudnason V. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. Am J Epidemiol. 2007 May 1;165(9):1076-87.

[4] The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol. 1989 Apr;129(4):687-702.

[5] Dawber TR, Meadors GF, Moore FE Jr. (1951) Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health; **41**(3):279-281

[6] Dawber TR, Kannel WB, Lyell LP. (1963) An approach to longitudinal studies in a community: the Framingham Study. Ann N Y Acad Sci; **107**:539-556

[7] Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. (1975) The Framingham Offspring Study. Design and preliminary data. Prev Med; **4**(4):518-525

[8] Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. (1979) An investigation of coronaty heart disease in families. The Framingham Offspring Study. Am J Epidemiol; **110**(3):281-290

[9] 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 Natiaonal Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol; **165**(11):1328-1335

[10] Mitchell BD, McArdle PF, Shen H, Rampersaud E, Pollin TI, Bielak LF, Jaquish C, Douglas JA, Roy-Gagnon MH, Sack P, Naglieri R, Hines S, Horenstein RB, Chang YP, Post W, Ryan KA, Brereton NH, Pakyz RE, Sorkin J, Damcott CM, O'Connell JR, Mangano C, Corretti M, Vogel R, Herzog W, Weir MR, Peyser PA, Shuldiner AR. (2008) The genetic

Nature Genetics: doi:10.1038/ng.386

response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. Am Heart J; **155**(5):823-8

[11] Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, Guralnik JM. (2000). Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. J Am Geriatr Soc; **48**(12):1618-25

[12] Epstein MP, Duren WL, Boehnke M (2000) Improved inference of relationship for pairs of individuals. Am J Hum Genet; **67**:1219–1231

[13] Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC. (2007). The Rotterdam Study: objectives and design update. Eur J Epidemiol; **22**(11):819-2

[14] Andrew, T. et al. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* **4**, 464-77 (2001).

[15] Spector, T.D. & Williams, F.M. The UK Adult Twin Registry (TwinsUK). *Twin ResHum Genet* **9**, 899-906 (2006).

Supplementary note

Study populations

AGES- Reykjavik Study. The Reykjavik Study cohort originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik in 1967 (3). A total of 19,381 people participated in the Reykjavik Study examination, a 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was invited to participate in all subsequent examinations, while one group was designated as a control group and was not included in examinations until 1991. Other groups were invited to participate in specific examinations of the study. Between 2002 and 2006, the AGES-Reykjavik Study re-examined 5764 survivors of the original Reykjavik Study. Successful genotyping was available for 1849 AGES women participants who were eligible for this study. The AGES-Reykjavik Study GWAS was approved by the National Bioethics Committee and the Data Protection Authority and also was covered under the MedStar Institutional Review Board. All subjects provided written informed consent.

ARIC: The ARIC study is a multi-center prospective investigation of atherosclerotic disease in a bi-racial population (4). White and African American men and women aged 45-64 years at baseline were recruited from 4 communities: Forsyth County, North Carolina; Jackson, Mississippi; suburban areas of Minneapolis, Minnesota; and Washington County, Maryland. A total of 15,792 individuals participated in the baseline examination in 1987-1989, with four follow-up examinations in approximate 3-year intervals, during 1990-1992, 1993-1995, and 1996-1998. Only White women with genotype data and age at menarche between 9 and 17 years of age were included in this analysis (N=4247). This study was approved by the institutional review board at each field center, and this analysis was approved by the University of North Carolina at Chapel Hill School of Public Health Institutional Review Board on research involving human subjects. All subjects provided written informed consent.

Nature Genetics: doi:10.1038/ng.386

FHS: The Original Cohort of the Framingham Heart Study was enrolled in 1948 to study determinants of cardiovascular disease and other major illnesses (5,6). In 1971, Offspring of the Original Cohort participants and Offspring spouses including 2641 women (mean age 36 years) were enrolled into the Framingham Offspring Study. Offspring participants have been examined approximately every 4 years (7,8). From 2002 to 2005, 4095 adults including 2641 women (mean age 40 years) with at least one parent in the Offspring cohort were enrolled in the Framingham Third Generation cohort (9). The Framingham Heart Study examinations were approved by the institutional review board at Boston University Medical Center. All participants provided written informed consent.

HAPI Heart Study: The Heredity and Phenotype Intervention (HAPI) Heart Study was initiated in 2002. Participants of the HAPI Heart Study comprised adults from the Old Order Amish community of Lancaster County, PA, who were recruited over a three-year period. Study participants were included if they were aged 20 years and older and considered to be relatively healthy based on exclusion criteria of severe hypertension (blood pressure > 180/105 mm Hg), malignancy, and kidney, liver or untreated thyroid disease. The study aims and recruitment details, including ascertainment criteria, have been described previously (10). Physical examinations were conducted at the Amish Research Clinic in Strasburg, PA and a reproductive health questionnaire was completed by female participants. Women presenting pregnant or within 6 months postpartum were excluded from the study.

InCHIANTI: The InCHIANTI study is a population-based epidemiological study aimed at evaluating factors that influence mobility in the older population living in the Chianti region of Tuscany, Italy. Details of the study have been previously reported (11). Briefly, 1616 residents were selected from the population registry of Greve in Chianti (a rural area: 11 709 residents with 19.3% of the population greater than 65 years of age) and Bagno a Ripoli (Antella village near Florence; 4704 inhabitants, with 20.3% greater than 65 years of age). The participation rate was 90% (n= 1453) and participants ranged between 21–102 years of age. The study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review. There were 85 parent-offspring pairs, 6 sib-pairs and 2 halfsibling pairs documented. We investigated any further familial relationships using IBD of 10,000 random SNPs using RELPAIR and uncovered 1 parent-offspring, 79 siblings and 13 half-sibling (12). We utilized the correct family structure inferred from genetic data for all analyses.

RSI and RSII: Rotterdam Study I and II, ongoing prospective population-based cohort studies, focus on chronic disabling conditions of the elderly in the Netherlands. In summary, men and women aged 55 years or older, living in Ommoord, a suburb of Rotterdam, the Netherlands, were invited to participate (13).

TwinsUK: *The TwinsUK cohort* consisted of a group of twins ascertained to study the heritability and genetics of age-related diseases (www.twinsUK.ac.uk). These unselected twins were recruited from the general population through national media campaigns in the UK and shown to be comparable to age-matched population singletons in terms of disease-related and lifestyle characteristics (14,15).

Nature Genetics: doi:10.1038/ng.386

A Study of the Relational Aspects of the Culture of Academic Medicine

Linda Pololi, MBBS, MRCP, Peter Conrad, PhD, Sharon Knight, PhD, RN, and Phyllis Carr, MD

Abstract

Purpose
The impact of medical school culture on medical students has been well studied, but little documentation exists regarding how medical faculty experience the culture in which they work. In an ongoing project, the National Initiative on Gender, Culture and Leadership in Medicine, the authors are investigating how the existing culture of academic medical institutions supports all faculty members' ability to function at their highest potential.

Method
The authors conducted a qualitative study of faculty in five disparate U.S. medical schools. Faculty in different career stages and diverse specialties were interviewed regarding their perceptions and experiences in academic medicine. Analysis was inductive and data driven.

Results
Relational aspects of the culture emerged as a central theme for both genders across all career categories. Positive relationships were most evident with patients and learners. Negative relational attributes among faculty and leadership included disconnection, competitive individualism, undervaluing of humanistic qualities, deprecation, disrespect, and the erosion of trust.

Conclusions
The data suggest that serious problems exist in the relational culture and that such problems may affect medical faculty vitality, professionalism, and general productivity and are linked to retention. Efforts to create and support trusting relationships in medical schools might enhance all faculty members' efforts to optimally contribute to the clinical, education, and research missions of academic medicine. Future work will document the outcomes of a five-school collaboration to facilitate change in the culture to support the productivity of all medical faculty.

Acad Med. 2009; 84:106–114.

A central task of medical schools is to help students, faculty, and medical practitioners learn how to form caring, healing relationships with patients and their communities and with each other. The educational environments, or the culture or milieu of work, reinforce learning, teaching, and the practicing agenda or belie the very intent behind our work.[1] Given the relevance of relational issues to health care, we describe our findings regarding the relational aspects of the culture of academic medicine.

A rich literature on relationships between physicians and patients provides evidence of the importance of trusting relationships for enhancing patient care and clinical outcomes.[2–5] In medical education, effective relationship formation and trust is pivotal in facilitating learning[6,7] and is helpful for interdisciplinary clinical partnerships and multidisciplinary research collaboration. In hospitals, positive interpersonal relationships have been linked to enhanced nurse and physician performance.[8,9] More recently, relationship-centered health care has been linked to quality of care and organizational performance.[10–12] In a report to the Association of American Medical Colleges (AAMC), Inui[13] "acknowledges the importance of the many relationships between individuals and positions in academic medical centers that embody our culture and affect any strategic plan we might devise or implement. In our organizations many of these relationships are hierarchical in nature and must be in play for any systematic change to go forward. . . . Other key relationships that may express and shape professional values and behaviors as well as medical organizational change include those among peers (clinical, research, education) and others." Beyond medicine, relational trust is a hallmark of effective education in middle schools[14]; in management and sociology studies, positive workplace relationships have been shown to benefit careers,[15] and in business schools, relationships between faculty have been shown to impact professional life beyond career benefits.[16]

Respected analysts have expressed concern about a conflict between traditional professional values and the commercialism of medicine.[17–21] Others have concluded that the development and funding of new knowledge in the biomedical and clinical sciences, and pressure to increase clinical "productivity," are given priority, with less attention devoted to social activism, humanistic concerns, or faculty development, and with educational excellence a less prestigious individual accomplishment.[22–24] To our knowledge, there has been little study of how current challenges in academic health centers

Dr. Pololi is senior scientist, Brandeis University, resident scholar, Women's Studies Research Center, and director, National Initiative of Gender, Culture and Leadership in Medicine: C - Change, Brandeis University, Waltham, Massachusetts.

Dr. Conrad is Harry Coplan Professor of Social Sciences and chair, Health, Science, Society and Policy Program, Brandeis University, Waltham, Massachusetts.

Dr. Knight is professor of health education and promotion and associate dean, East Carolina University, College of Health and Human Performance, Greenville, North Carolina.

Dr. Carr is professor of medicine and associate dean of students, Boston University School of Medicine, Boston, Massachusetts.

Correspondence should be addressed to Dr. Pololi, National Initiative on Gender, Culture and Leadership in Medicine: C - Change, Brandeis University, 415 South Street, Mailstop 088, Waltham, MA 02454-9110; telephone: (781) 736-8132; fax: (781) 736-8135; e-mail: (lpololi@brandeis.edu).

(AHCs) affect the experience and relationships of medical faculty.

Although the focus of our overall research project has been on culture change to support the advancement of women and underrepresented minority (URM) faculty in academic medicine, this report is restricted to a central emergent theme, the relational aspects of the culture of AHCs, as reflected in our interviews with faculty. Little work has addressed cultural changes that would create environments more appropriate for taking full advantage of women and URM faculty's potential contributions, although one report did suggest that cultural and organizational issues in academic medicine contribute to women's lack of advancement[25]; outside medicine, a body of literature describes organizational issues for women in the workplace.[26,27]

The purposes of this article are therefore (1) to present insights into relational aspects of the culture of academic medicine at select institutions from the perspective of medical faculty, and (2) to consider the implications of these insights for all faculty, with special attention to women and URM faculty. Our findings are part of a larger set of qualitative data associated with the National Initiative on Gender, Culture and Leadership in Medicine (informally known as C - Change, which is short for Culture-Change).[28] This initiative, made possible by the Josiah Macy, Jr. Foundation, comprises a partnership of five medical schools and Brandeis University, collectively engaged in action research to address the imperative of developing women and URM faculty members' full potential and leadership in academic medicine in the United States.

Method

We selected the five C - Change medical schools to represent the diverse organizational characteristics of the 126 medical schools in the U.S. at the time of our study (i.e., public/private, National Institutes of Health [NIH] research intensive, primary care/community orientation). Schools were selected from each of the AAMC-designated regions. One school each was selected from the Western, Southern, and Central regions and two from the Northeastern region, which has the largest concentration of medical schools. Included were two public and three private schools. Aggregate statistics regarding women and URM faculty in these schools were almost identical to national statistics. At the time of this study (2006–2007), nationally, 32% of medical faculty were women, and, in participating schools, 35% were women. Nationally, women made up 16% of full professors, and, in participating schools, 17% were women. Nationally, 7% of faculty were from URM groups (3% African American/black and 4% Hispanic), and, in participating schools, 6% were from URM groups (2% African American/Black and 4% Hispanic).

Table 1

Gender and Career Stage (n = 96) of Faculty From a Five-Medical-School Study of the Experiences of Faculty (2006–2007)

Career stage (total =96)	Female faculty	Male faculty
Early (20)	13%	7%
Plateau (23)	15%	9%
Leadership (29)	16%	14%
Left academic medicine (24)	11%	14%

Participant criteria

We selected medical faculty from the five C - Change medical schools using stratified purposeful and chain sampling strategies[29] according to medical school site, gender, race/ethnicity, department/discipline, and career status. We used stratified purposeful sampling to capture variations of experience or perspective that may occur among individuals at different career stages. The PI (L.P.) sent an e-mail invitation to participate to potential interviewees together with information about the purpose of the study. We did not offer any compensation. Prior to interviews, we secured written informed consent. Interviewers reiterated the purpose of the study and ensured confidentiality and anonymity at the beginning of each interview. IRB approval was obtained.

Participants were research scientists, medical and surgical subspecialists, and generalist medical faculty who held doctoral degrees (84% MD/DO, 16% PhD) and represented a wide diversity of subspecialties. We invited 170 faculty members in four career stages to participate: (1) early career, that is, those who had been faculty members for two to five years, (2) "plateaued," that is, those who had not advanced as expected in rank and responsibility and who had been faculty members for 10 or more years, (3) faculty in leadership roles such as deans, departmental chairs, and center directors (identified as "senior" in the quotations below), and (4) former faculty who had left academic medicine ("departed"). We divided interviewees equally among the four groups, but with fewer participants in the early-career stage because we reached data saturation in this category early in the study (Table 1). We interviewed similar numbers of faculty from each of the five schools.

Data collection and analysis

The four coauthors conducted the one-on-one interviews. We audio-recorded and transcribed verbatim all interviews, which were typically one hour in length. The semistructured interview guide consisted of open-ended questions focused on choice of medicine as a career, aspirations of faculty, energizing aspects of their careers, barriers to advancement, interdisciplinary collaboration, leadership, power, values alignment, and work–family integration. We developed the guide through a pilot series of interviews. List 1 shows some of the questions.

List 1

Selected Questions From the Interview Guide

- What is it about your work that energizes you?
- When have you felt most successful in your work?
- What has been your sense of being a part of your department/university/institution?
- What has been difficult or frustrating in your work?
- How are your aspirations for yourself in academic medicine being fulfilled?

We analyzed aggregated data by repeatedly reading masked interview transcripts to develop understanding and interpret meaning. After identifying and applying codes to the more than 4,000 pages of transcribed narrative data, we stored coded data using Atlas.ti software. Our analysis involved data reduction or condensation, from which we identified patterns and themes emergent in the coded data. We used an inductive and data-driven analysis process, in line with grounded theory.[30,31] To verify our conclusions, we returned to the transcripts, reevaluating the findings by review among the four coauthors to develop intersubjective consensus.

Results

Sample selection

Of the 170 faculty who were invited to participate, 8 individuals refused (usually because of time constraints), 54 did not respond, and we were unable to schedule with 12 potential interviewees. A total of 96 faculty participated, for an acceptance rate of 56%. Fifteen percent (16) of interviews were in person, and the remainder (80) were by telephone. It was more difficult to identify male "plateau" faculty than similar-stage women faculty members. Overall, those in leadership were more likely to agree to participate, and we found it most difficult to secure interviews with early-career faculty (Table 1). Women (55%) and URM faculty were oversampled (17% African American/black, 4% Hispanic/Latino, and 79% Caucasian/white), as were generalists (20%) (defined as general internal medicine, family medicine, and general pediatrics).

Relational aspects of the culture

Relational aspects of the culture emerged as a central theme in the data, with no appreciable gender differences noted. Quotes in different categories are from different respondents. Relational comments tended to be spontaneously mentioned rather than elicited by interviewers, except in responses to the question about institutional support of interdisciplinary collaboration. Relatively few women and men described positive relational attributes with colleagues. Where positive relationships with their colleagues were described, there was often a sense that these relationships assumed important protective or buffering effects against the dysfunctional aspects in the culture. A senior woman said,

> I wouldn't say it's an overly socially warm place but I guess it is welcoming. People are nice here, it's easy to interact. It is a place that's relatively free of bias, at least racial bias. Women haven't done any better here than they have in other academic places but it's still a nice environment and I've been able to do all the things I've wanted.

She went on to say later in the interview,

> You have to position yourself so that you are part of the decision making. You can't expect and wait to be the beneficiary of benevolence . . . the recipient, and so I guess that's been my guiding principle; this is a marketplace environment. Even though we like to think we live the life of the mind, it in fact is a marketplace and you have to have something to be able to bring to the table that gives you equality.

A woman early in her career said,

> We have a small group of junior faculty who by all of the turmoil—we have been thrown together into each other's laps and that has been a wonderful thing because we can say "Oh my gosh, this project or line of thinking isn't working out. There's basically something wrong there and I can't figure out what it is." If I were to say that to a senior colleague, I would put at risk their evaluation of me because I had made such a grave mistake. So identifying those colleagues can help you sort out the problems you have at work.

When this individual was asked how her aspirations in medicine were being fulfilled, she replied,

> I guess with establishing relationships in the workplace of trust, relationships in which I can howl my failures without fear of retribution in some way, so that part is terrific. I mean, to work in an environment doing something that you like to do and having colleagues around you who are for the most part supportive or a number of them, that's very rewarding. I still don't know if I'm going to succeed . . . so that's a scary or uncomfortable position.

Other faculty also commented on how any closeness with a colleague would provide some counterbalance to the negative aspects of the culture. A senior man said,

> I feel [a sense of connectedness] when I'm working or talking doctor to doctor and we're discussing a mutual patient. That's where there's a real sense of family on the individual level. And the flipside of that is the administrative side which forces all of the life out of everything you do.

Respondents particularly valued their research collaborators. A senior female scientist who had left academic medicine said, "I felt very little of a sense of belonging except to my own research group, which felt like a team with a wonderful mix of people."

Positive relationships with learners and patients

Numerous faculty spoke of positive and valued relationships with students and residents, and with patients. Male and female faculty found particularly rewarding their interactions with physicians-in-training, as illustrated below:

> I'm most successful in my work when I'm actually at the bedside. Yeah. When I'm at the bedside with a learner. I feel like I'm giving out. I'm giving in a way that you can only give if you're with the patient. You're with a learner. That's when I'm most gratified. (Senior male in obstetrics)

> And so to see the residents get excited about something that I'm excited about, and that's good for the community and good for people who really don't have access to care. I think right now that's what really gets me, so it is teaching, But it's teaching, not so much the kind of didactic or the typical ward attending kind of teaching, it's sort of the broader teaching about the world of medicine in community. (Plateaued female generalist)

> The sense of belonging I had was really at its highest when I was with learners. I never really felt a particular kinship to other faculty. I think it was really pretty much at times when I was really immersed and surrounded by students that I felt like I was a member of the university. (Senior woman in pathology)

Faculty also referred to their relationships with patients when we asked about what energized them in their work. For example:

> I really do like the oneness of working with patients and getting to know them; establishing rapport and hopefully getting them to open up to . . . establishing the trust and all. (Senior man)

> I take care of frail elders. That really energizes me in a way that I can't explain. I love these people. Even after I leave a day here, I can go to the nursing home even though I'm tired, and it will reenergize me. (Midcareer woman)

Collaboration

Some respondents described collaboration with a few colleagues where it was seen as a

very positive activity that enhances the environment, research, and education:

> Part of it is a network of colleagues—I mentioned locally, but also nationally. So I have a couple of folks whose values I share and we get together sporadically to trade stories. They tend to be in leadership roles in other institutions and we support each other in trying to affect these big complex institutions and train a group of people to carry on the work. (Male senior generalist)

> I really try to make a collaborative environment where everybody feels that they're part of the decision making and part of effective change . . . it's a much better environment if there's collaboration and support. (Woman leader)

> It's great when it happens, but when it happens, it's because you're sitting with somebody at lunch . . .. It's because you're sitting there and talking about your work and you think, "Oh, I work on that. Let's do something together". I've gotten a grant by doing that. . . ." You go to meetings; you go to each other's seminars. That's how it happens. And so several of us collaborate on that, but it's not because anybody told us to. It's because we found each other. So that does work. Collaboration is the best thing. (Senior female scientist)

Dilemmas in relationships

Disconnection. In contrast, many faculty members described feeling isolated and lacked supportive relationships. Narratives documented personal disconnection and separation rather than relationship formation among colleagues. Respondents perceived that the environmental norms and structures did not value or support relationships and did not facilitate their formation. The troubling relational themes that emerged in the data were evident for both men and women and were expressed by faculty in all career stages.

A male subspecialist early in his career commented,

> I couldn't pick out anybody that I corresponded with by e-mail or letters out of a line-up. I knew very few people in different divisions. It was very much an isolated situation. Go to your clinic, do your thing, go back to your office, go to the medical suite, do your procedures, go back to the office. . . .

Comments from two faculty (midcareer and departed) illustrate their awareness of barriers to relationship formation:

> I realized that I valued relationships and interpersonal behavior that the institution did not. (Female)

> So a lot of what women do to make a culture more nurturing—help the people to grow as human beings, to become, in my judgment, better clinicians—is not something the institution values. They can do it if they want, they get all kinds of laudatory praise, but they do it on their own time. (Female)

We heard these themes from many faculty in early or midcareer stages, but those in leadership also commented on their disconnection with colleagues. A number of faculty commented that leaders felt distant from them.

Competitive individualism. Interviewees described an intensely individualistic and competitive environment where rewards are usually accorded to individual contributions. Respondents perceived that individuals and institutions tend to function on behalf of their own self-interests. It was accepted that a stressful, competitive environment is necessary to promote scientific progress and achievement. A senior male faculty said,

> You're encouraged to be a reductionist in your thinking, to get your niche to get to be successful, which again I don't have a problem with that in general. You are encouraged to be single minded, self-indulgent, selfish; the first question out of people's mouths is, "Well what is this going to do for me?" "What paper do I get out of it, where do I go on the paper, who's looking out for me?" and all that stuff. And it's just like there should be enough to go around.

He went on to comment,

> But I think what it breeds—and this gets into the heart of the academic culture certainly at its lower and midlevels—is an unpleasant place in a lot of ways, as people are scrambling up over one another trying to find their way and find their niche and find their grants and so forth. . . . I don't like what it does to people. And I think very nice, thoughtful people become very selfish and self-indulgent because they're pushed to get the grant.

A number of faculty found individual self-promotion distasteful. A woman noted,

> She [her supervisor] said "you have to brag, you really do." And that's very difficult for many people because it's not the nature of some of our cultures, experience, and maybe just family culture, too.

A senior man who left academic medicine said,

> I wasn't driven by the self-promotion that I think has to come on in academic institutions. It's all about getting new grants—"I have more than you" and "I'm the expert in this." So it's a little bit of an unreal, self-promoting kind of environment.

The expectation of personal overextension was often expressed and may be another result of the competitive environment. The culture described by respondents was often linked to having other adverse effects on faculty. The competitiveness of climbing the ladder in academic medicine was related to faculty members assuming aggressive, self-seeking, and uncollegial behaviors not previously evident. A number of faculty suggested that dealing with this environment brought about changes in outlooks and behaviors, both in colleagues and themselves. A senior woman who left academic medicine noted,

> And there were colleagues of mine to whom power meant a great deal and I watched them become people I didn't like as they dealt with this hostility and grabbed for the power, and they achieved a great deal and I don't take it away from them, but in the course of it, they lost their humanity. They became people I could no longer respect. They became dishonest and manipulative.

Another senior woman who stayed in academic medicine pointed to the creation of a "toxic" environment:

> I never felt like I belonged. . . . The environment that I was in was quite toxic in an interpersonal way. . . . You learn to become extremely aggressive and obnoxious people.

Undervaluing humanistic qualities. Numerous faculty spoke of not being recognized as people beyond their professional roles at work. There was a lack of attention to what individual faculty were themselves feeling, with no invitation or expectation to express personal emotions or to talk about important personal issues either related to work or to their personal lives. Faculty described the environment as having a "dehumanizing" effect on them by only recognizing the work aspect of faculty. The culture seemed to reduce the

qualities in faculty that make them able to meet human needs, be compassionate, and show sensitivity to others. A midcareer female medical subspecialist said,

> Nobody cares what makes me tick here. I'm completely invisible—as a human—as a person. A nonprofessional person. It just seems like I go through most of my day with nobody recognizing who I think I am. Or acknowledges me in any—in any complex sense . . . or me as a unique individual. I just appear to be what I represent.

Another female faculty member commented,

> Check your humanity at the door, that was how it felt. Any sign of . . . this is gonna sound harsh, but . . . any tendency towards kindness was viewed as weakness.

Several interviewees felt that this situation had the effect of preventing them from being fully themselves in their work life and that they only selectively brought aspects of themselves and their thinking to their professional lives:

> One consequence of this is a dehumanizing effect on the faculty, where an individual is not able to bring his or her feelings authentically into the workplace. (Senior woman physician)

Deprecation and disrespect. Interviewees gave little indication of medical schools cultivating an appreciative culture, but rather one of finding fault. Researchers, educators, and clinicians spoke of feeling disrespected or of not being valued as faculty who have contributed to the medical school's successes. A female medical subspecialist commented on the common expectation of finding fault:

> People tend to defend their territory, defend and assume that you're attacking them. This is an environment where the assumption is that people are trying to think ill of you. Or are trying to find the moment where you slip up. Who wants to work in that kind of environment?

A male medical subspecialist told of an experience:

> I have a relatively new [supervisor]; after he'd been here a year, he called me into the office and said that he had reviewed everything that the department had ever done, all our publications and in his mind—this is a quote—"We'd never done anything important in the history of the department." [laughs] I sat there and said, "Really?" I said "Then could you perhaps clarify to me what counts as important?" He said, "Yes, publishing in [two elite medical and science journals]." I said that "within my department of 100 faculty, I doubt there is anybody who even reads those journals, much less publishes in them."

Another dimension of disrespect emerged as disloyalty, as faculty also perceived that their larger organizations are not loyal to the faculty. A senior female basic science faculty member commented,

> I think what they've done recently, not to me as much, but to faculty who have always had grants and who are now having trouble getting them. . . . And now they're turning around and if people can't get grants, they're making them feel bad, making them feel kind of worthless. I think that's not nice. And instead of saying, "Good job, you've done a good job. You've gotten grants for 20 years and you've been a good teacher or you've been a decent teacher." And now [they say instead], "You don't have a grant; now you're worthless."

Numerous faculty members commented on not feeling recognized by the medical school for their contributions, as this example from an early-career woman physician illustrates:

> We're not rewarded by the medical school at all. We're not recognized. A few people each year might be recognized, but for the ongoing day-to-day grind, we're not recognized by the medical school for our efforts.

A midlevel male clinical faculty member articulated a common theme where respect was associated with receiving grants and disrespect with teaching activities:

> I've seen it everywhere I've ever been, so it's not unique to this university. What the university had was this hierarchy of needs that began with your ability to support yourself with grants, [and] with teaching at the bottom, and it was very explicit. Personal relationships were defined by distrust and disdain if you didn't get a grant.

A senior woman remembered how she adapted, but at a cost:

> My assumption would be that a lot of their behavior was from a place of insecurity in which they learned some really powerfully negative pushback behavior that I learned too.

Erosion of trust. In the interviews, we found instances where faculty were unwilling to say what they believed for fear of retaliation. Likewise, faculty feared being penalized for discussing home problems. A female plateaued faculty member explained her sense of not being able to express herself for fear of losing her job:

> Well, I think the hardest thing for me was to be in a department where you couldn't express yourself [your opinion] without feeling that you were jeopardizing your career. The hardest thing was that I wasn't honest to myself sometimes and because I was afraid earlier on that I would lose my job—I would get kicked out of the department. Although I don't know if that would have happened, but it did happen to other people. There were people in our department who lost their jobs over their being expressive. Their lives were made absolutely miserable.

A midcareer female faculty member described how she overextended herself because of her fear that if she was not seen to be doing this, she would risk losing her job:

> Early on, when I was doing purely clinical, in my division, I was bringing in more money than anyone and part of it is because I would be working until 10:00 at night and just thinking, "These people may fire me. I've got to do all this work. You know, I can't ever refuse anything." So I was really, really, really killing myself and, of course, getting older in age and feeling more and more tired.

Respondents revealed breaches of academic integrity that seemed to be tolerated or even expected in the environment. Some examples follow from women:

> I work on projects, where if I present any kind of tantalizing evidence someone down the hall will go and do those experiments and just scoop you and just essentially take all your ideas and everything and just run with it and because they're bigger and they're faster. It's also stealing of ideas. You know, you send a grant to have a colleague look it over and lo and behold, your data end up in their grant and things like that.
>
> What he had a wonderful ability to do is to take all of my hard work and give himself credit for it.
>
> My chairman asked me to take over the new faculty, and so he would give me these things that I was supposed to tell them we were going to do for them. And I would say to him, "We don't do this for people. How can I tell people this is available when you and I both know we don't follow through on that?" He said "Well, you have to because that's the only

> way they'll come here." Well, I don't lie. That's not what I do.

The following quotations are from male leaders:

> I basically value being completely honest. You know, I'm asking for this, and this is why I'm asking for it, and this is what I'd like to do with it. Honesty is not always either rewarded or reciprocated in academic medicine.

> I think there are plenty of people that will try to maneuver or get things done or get decisions on their behalf by not being fully forthright or honest about what the issues are. I have this with students as well as faculty and administrators. So you have to be very careful in the academic environment to ask the right questions and look for the right motivations, or the wrong motivations, as they are in some cases.

The theme of dishonesty emerged in the educational enterprise, too. A female midcareer faculty member said,

> We would tell students they were going to get excellent teaching, but they kept increasing the number of patients that the doctors had to see, and I watched the education of the students falling off everyone's radar. I said, "You know, we're lying to the people who are doing our evaluations, we're putting things on paper that we don't do, and we're not being fair to the students." It was like, "We have to make money so the students are going to have to suck it up." And you know, what happened was that the people who were the best teachers ended up leaving over and over again.

Discussion

As illustrated at the beginning of the Results section, we did hear very positive comments about teaching, where effective relationships with students were evident and prized, and relationships with patients emphasized trust and caring. Some faculty also spoke of supportive collaborative relationships with close colleagues. However, negative perceptions of relational experiences were articulated in the majority of all interviews, despite the fact that the questions posed were open-ended and purposefully sought accounts of positive experiences in the tradition of Appreciative Inquiry[32] and did not request accounts of negative relational issues. Even in this context, fundamental aspects discussed by faculty of the experience of academic medical culture were a sense of disconnection and an erosion of trusting relationships with colleagues and supervisors.

Our data suggest that serious problems exist in the relational culture and that these can affect faculty vitality, professionalism, and productivity and are linked to retention. These aspects of the culture may undermine the goals of medical institutions and are antithetical to fostering superior patient care, biomedical research, and educational excellence. At the very least, they make medical schools much less supportive and positive workplaces for professional work.

This study was conducted in only five schools; the ability to generalize insights from our qualitative findings is being assessed by us through a national quantitative survey of medical faculty. However, the themes we report were generally consistent across the faculty we interviewed. In earlier pilot studies when the PI interviewed a national sample of 22 faculty, she found similar results (unpublished).

Alignment with the findings of other researchers

Our findings align with those of others; a recent survey study of four U.S. medical schools found elevated rates of depression and job dissatisfaction, especially among younger faculty.[33] These authors note, "Current medical students are being taught by faculty who are increasingly stressed and dispirited. . . . The majority of faculty respondents indicated that their initial job expectations were not being realized, they were not the contributors they used to be, and that their productivity was decreasing. Significant numbers of the faculty felt unsupported."[33] Additional evidence pointing to dysfunction in the culture are high levels of physician dissatisfaction[34] and faculty burnout in 37% to 47% of academic faculty,[35] although burnout was found to be uncommon in deans.[36]

Women physicians have 1.6 times the risk of burnout compared with male colleagues,[37] and the suicide rate in women physicians is twice that of other working women.[38] Women may be more sensitive reactors to the milieu of AHCs, and, together with URM faculty, they may be on the leading edge of a reaction to the perceived challenges of the environment of academic medicine (the "canary in the coal mine").

Relationships in health care

A lack of positive relational attributes may also be found in nonmedical workplaces, but we and others believe that such a lack has particular significance for medical settings because physicians must be skilled in forming trusting relationships with their patients to effectively address the biological, psychological, and social impact of illness.[39] Substantial evidence links relational deficiency with adverse health care outcomes. Physicians who are self-aware of their own responses and feelings when they are with patients are more effective and more satisfied in providing patient care,[40] and physicians' humanism correlates with patient satisfaction and adherence to medical advice.[41] Beach et al[10] recognized the quality of relationships as central to health care and articulated core principles: relationships in health care ought to include the personhood of the participants, affect and emotion are important components of relationships, and all health care relationships occur in the context of reciprocal influence. Safran et al[11] extend these concepts and propose a model of relationship-centered organizations. It has been shown that chronic disconnection results in diminished energy and creativity and precludes growth-fostering relationships.[42–44] Our data also suggest that negative relational attributes are barriers to faculty vitality, creativity, and satisfaction. Disconnection and emotional detachment in the culture can be viewed as a parallel to ineffective communication between doctor and patient, as well as influencing organizational performance.[11] Continuity of relationship is emerging in new trends in medical education such as the Cambridge Hospital initiative[45] and a current Carnegie Foundation study.[46]

Linkage to professionalism

Cohen and colleagues,[23] reflecting on a new guide to medical professionalism,[47] noted recently that "institutional and organizational settings of contemporary medical practice pose significant impediments to achieving several of the responsibilities to be assumed by physicians." They suggest that these structural barriers to professionalism may be beyond the control of physicians.

Although we agree with their social structural analysis, we also ask whether there are additional barriers to professionalism in the culture and relationships within academic medicine.

In the past year, new accreditation standards mandate "interpersonal and communication skills and professionalism" for residency training.[48] Moreover, "medical schools (including faculty) must ensure that the learning environment for medical students promotes the development of explicit and appropriate professional attributes in their medical students."[49] These standards reinforce the American Board of Internal Medicine Project Professionalism[50] and the AAMC's Medical School Objectives Project,[51] and they direct attention to effective relationship formation among medical faculty.

Latent culture

Medical students tend to lose their humanistic and altruistic attitudes during their medical school years.[52–60] Current practices may, in fact, be barriers to physicians-in-training developing compassion and competence, and they may contribute to unprofessional behaviors.[61,62] A substantial literature describes the informal or "hidden curriculum" for medical students where students experience behaviors and attitudes, embodied in the organizational approach, that contrast with the school's espoused mission.[63] In effect, students undergo tacit social conditioning[58,64,65] and learn certain informal norms and values that may be at odds with humanistic and ethical comportment. Many attributes of the culture in medical schools[66] have been thought to be barriers to learning, team building, and compassionate care.

Just as the "hidden curriculum" has been linked to unwarranted stress for students, as well as to lapses in their professional and ethical behaviors,[59,67–69] we postulate that faculty who experience the latent culture described in our findings may exhibit similar feelings and behaviors. Furthermore, medical faculty are teachers to be observed and emulated by medical students, and, in this latent culture, they may pass on the norms and culture they experience for themselves. Part of the responsibility of medical schools is to teach medical students to be humanistic, socially responsible, and compassionate[70] and to provide learning experiences that nurture trainees' self-awareness and emotional development. Lack of modeling of effective relational practices by faculty would logically impact student learning as well as patient care.[71]

A structure that rewards individual achievement, self-promotion, and being lead author on publications, rather than collaborative efforts, may seriously impede interdisciplinary and collaborative work in the biomedical sciences. The NIH have recently recommended interdisciplinary collaboration between scientists as a priority for scientific discovery and translation of new knowledge to clinical outcomes.[72] For example, teams of researchers that include practicing physicians as well as laboratory scientists are more likely to produce advances in scientific knowledge that can be translated and applied to improved patient care. High levels of ethical misconduct reported in federally funded faculty researchers[73] may also be partially attributed to the culture in which faculty may find themselves working.

Why faculty stay in academic medicine

Similarly to students in *Boys in White*[74] who turned their medical school experience into something functional as they learned the practice of medicine, our faculty may draw on more positive aspects of their experience in AHCs to buffer them from dysfunctions in the system. The qualities and rewards felt from their relationships with students and patients, the excitement of the intellectual challenge of medicine, an altruistic social contract, and the few close relationships that they do have with colleagues buffer, protect, and support faculty in their contributions to health care, education, and research, and decrease the likelihood of members of this critical group leaving academic medicine. Our data suggest that behaviors promoting relationship formation can mitigate stress and may help prevent burnout. Our experience of working with medical faculty strongly suggests that most faculty wish to have trusted colleagues and that most desire connection and relationship formation.[64,65,75–78]

Collaboration involves forming relationships, developing understanding of the perspectives of others, and learning effective patterns of interpersonal communication. To be successful in the long term, arriving at some enjoyment of working with other team members is helpful. Essential are the cognitive contributions and expertise of team members, but their emotions affect not only their own work but the work of the group. So, in any workplace, supporting both the emotional and intellectual well-being of workers will be vital for optimal work.

Recommendations

"*The Human Condition of Healthcare Professionals* aptly expresses the everyday essence and deeper meaning of medical work and the impact on the men and women who have chosen it."[39] A logical response to the findings of our study would be for medical schools to make efforts to instigate and support practices that encourage relationship formation among faculty and leaders. Supporting connection in trusting relationships and the human condition of health professionals would facilitate a core change in the medical school culture and contribute to realizing the potential of all faculty, including women and members of URM groups. We suggest that enhancing relational practices in medical schools would result in improved communication and collaborative efforts in patient care, research, education, and administration, and a more satisfied and energized faculty. This would allow the institution to avail itself of both women's and men's potential contributions and skills. Similarly, in his recent AAMC presidential address, Kirch suggested that low faculty personal morale is caused by an imbalance within our institutions and recommended that "we spend time explicitly assessing and building the right kind of culture."[79]

The eventual improvements achieved by the C - Change Initiative should benefit all faculty in academic medicine and enhance the value of the nation's substantial investment in health care delivery and workforce development. In addressing what are national problems in medical schools, the deans of the five C - Change schools have taken a leadership stance and courageous approach to having their faculty confidentially interviewed. Additionally, the deans are committed to being personally engaged in a collaborative Learning Action Network to explore methods for making

changes in the cultures of the C - Change schools. Efforts and outcomes of the C - Change Initiative will be disseminated as the larger project more finely hones the issues. Efforts to create and support trusting relationships in medical schools are likely to enhance all faculty members' efforts to optimally contribute to the clinical, education, and research missions of academic medicine.

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 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 Health Care Research and Quality; and the Centers for Disease Control and Prevention and Health Resources and Services administration.

The authors thank Jordan Cohen, David Kern, Milton Kahne, Michael Goldstein, Athanasius Anagnostou, Tom Inui, and Hilda Kahne for generously offering their helpful suggestions on earlier drafts of this manuscript. Thanks as well to Marianne McPherson and Meg Lovejoy for assistance with data coding. The authors are indebted to the medical faculty who generously shared their experiences in the interviews.

References

1 Tresolini CP; The Pew Fetzer Task Force on Psychosocial Education. Health Professions Education and Relationship-Centered Care: Report of the Pew-Fetzer Task Force on Advancing Psychosocial Education. San Francisco, Calif: Pew Health Commission; 1994:1–76.
2 Kaplan SH, Greenfield S, Ware JE Jr. Assessing the effects of physician patient interactions on the outcome of chronic disease. Med Care. 1989;27:S110–S127.
3 Williams GC, Deci EL, Ryan RM. Building healthcare partnerships by supporting autonomy: Promoting maintained behavior change and positive health outcomes. In: Suchman AL, Botelho RJ, Hinton-Walker P, eds. Partnerships in Healthcare: Transforming Relational Process. Rochester, NY: University of Rochester Press; 1998:67–88.
4 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centers. Ann Intern Med. 1986;104:410–418.
5 Fugelli P. Trust in general practice: The James Mackenzie lecture, 2000. Br J Gen Pract. 2001;51:575–579.
6 Rogers C. The Interpersonal Relationship in the Facilitation of Learning. Freedom to Learn for the 80's. Columbus, Ohio: Bell & Howell Co; 1983:119–134.
7 Daloz L. Guiding the Journey of Adult Learners. San Francisco, Calif: Jossey-Bass; 1999.
8 Revans RW. The hospital as a human system. Bull N Y Acad Med. 1996;73:418–429.
9 Spickard A, Gabbe SG, Christensen J. Mid-career burnout in generalist and specialist physicians. JAMA. 2002;288:1447–1450.
10 Beach MC, Inui TS; Relationship-Centered Care Research Network. Relationship-centered care: A constructive reframing. J Gen Intern Med. 2006;21(1 suppl):S3–S8.
11 Safran DG, Miller W, Beckman H. Organizational dimensions of relationship-centered care: Theory, evidence and practice. J Gen Intern Med. 2006;21(1 suppl):S9–S15.
12 Suchman AL, Deci EL, McDaniel SH, Beckman HB. Relationship-centered administration: A case study in a community hospital department of medicine. In: Quill TE, Frankel RM, McDaniel SH, eds. The Biopsychosocial Approach: Past, Present, and Future. Rochester, NY: University of Rochester Press; 2003.
13 Inui TS. A Flag in the Wind: Educating for Professionalism in Medicine. Washington, DC: Association of American Medical Colleges; 2003.
14 Bryk AS, Schneider BL. Trust in Schools: A Core Resource for School Reform. Educ Leadersh. 2003;60:40.
15 Brass DJ. Men's and women's networks: A study of interaction patterns and influence in an organization. Acad Manage J. 1985;28: 327–343.
16 Gersick CJG, Bartunek JM, Dutton JE. Learning from academia: The importance of relationships in professional life. Acad Manage J. 2000;43:1026–1044.
17 Relman AS. Education to defend professional values in the new corporate age. Acad Med. 1998;73:1229–1233.
18 Relman AS. A Second Opinion: Rescuing America's Health Care. New York, NY: PublicAffairs; 2007.
19 Relman AS. Medical professionalism in a commercialized health care market. JAMA. 2007;298:2668–2670.
20 Swick HM. Academic medicine must deal with the clash of business and professional values. Acad Med. 1998;73:751–755.
21 McArthur JH. The two cultures and the health care revolution. JAMA. 1997;277:985–989.
22 Lewis IJ, Sheps CG. The Sick Citadel: The American Academic Medical Center and the Public Interest. Cambridge, Mass: Oelgeschlager, Gunn & Hain; 1983.
23 Cohen JJ, Cruess S, Davidson C. Alliance between society and medicine: The public's stake in medical professionalism. JAMA. 2007;298:670–667.
24 Ludmerer KM. Learning to Heal the Development of American Medical Education. New York, NY: Basic Books, Inc.; 1985.
25 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.
26 Ely R, Myerson D. Theories of gender in organizations: A new approach to organizational analysis and changes. Res Organ Behav. 2000;22:103–151.
27 Fletcher JK. Disappearing Acts: Gender, Power, and Relational Practice at Work. Cambridge, Mass: MIT Press; 1999.
28 National Initiative on Gender, Culture and Leadership in Medicine: C - Change. Available at: (http://cchange.brandeis.edu). Accessed September 14, 2008.
29 Biernacki P, Waldorf D. Snowball sampling: Problems and techniques of chain referral sampling. Sociol Methods Res. 1981;10:141–163.
30 Glaser B, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago, Ill: Aldine; 1967.
31 Charmaz K. Constructing Grounded Theory. Thousand Oaks, Calif: Sage Publications; 2006.
32 Cooperrider DL, Sorensen PF, Whitney D, Yaeger TF, eds. Appreciative Inquiry: Rethinking Human Organizations Toward a Positive Theory of Change. Champaign, Ill: Stipes Publishing; 2000.
33 Schindler BA, Novack DH, Cohen DG, et al. The impact of the changing health care environment on the health and well-being of faculty at four medical schools. Acad Med. 2006;81:27–34.
34 Zuger A. Dissatisfaction with medical practice. N Engl J Med. 2004;350:69–75.
35 Shanafelt T, Sloan J, Haberman T. The well-being of physicians. Am J Med. 2003;114: 513–519.
36 Gabbe SG, Webb LE, Moore DF, Harrell FE, Spickard A, Powell R. Burnout in medical school deans: An uncommon problem. Acad Med. 2008;83:476–482.
37 McMurray JE, Linzer M, Konrad TR, Douglas J, Shugerman R, Nelson KJ. The work lives of women physicians results from the physician work life study. The SGIM Career Satisfaction Study Group. J Gen Intern Med. 2000;15: 372–380.
38 Peterson MR, Burnett CA. The suicide mortality of working physicians and dentists. Occup Med. 2007;58:25–29.
39 Fox RC. Essays in Medical Sociology. The Human Condition of Health Professionals. New Brunswick, NJ: Transaction; 1988:572–587.
40 Novack DH, Suchman AL, Clark W, Epstein RM, Najberg E, Kaplan C. Calibrating the physician personal awareness and effective patient care. JAMA. 1997;278:502.
41 Hauck FR, Zyzanski SJ, Alemagno SA, Medalie JH. Patient perceptions of humanism in physicians: Effects on positive health behaviors. Fam Med. 1990;22:447–452.
42 Jordan JV, Kaplan AG, Miller JB, Stiver IP, Surrey JL. Women's Growth in Connection: Writings From the Stone Center. New York, NY: Guilford Press; 1991.
43 Jordan JV, ed. Women's Growth in Diversity: More Writings From the Stone Center. New York, NY: Guilford Press; 1997.
44 Miller JB, Stiver IP. The Healing Connection: How Women Form Relationships in Therapy and in Life. Boston, Mass: Beacon Press; 1997.
45 Hirsh DA, Ogur B, Thibault G, Cox M. Continuity as an "organizing" principle for clinical education reform. N Engl J Med. 2006;356:858–866.
46 Cooke M, Irby DM. Carnegie Foundation for the Advancement of Teaching. Personal communication of report in preparation.
47 ABIM Foundation; ACP-ASIM Foundation; European Federation of Internal Medicine. Medical professionalism in the new

millennium: A physician charter. Ann Intern Med. 2002;136:243–246.
48 Accreditation Council on Graduate Medical Education (ACGME) Competencies. Available at: (http://www.umm.edu/gme/core_comp.htm). Accessed September 30, 2008.
49 Association of American Medical Colleges. 2007 LCME Standard MS-31-A. Available at: (http://www.lcme.org/standard.htm). Accessed September 24, 2008.
50 The American Board of Internal Medicine. Project Professionalism. Available at: (http://www.abim.org/pdf/publications/professionalism.pdf). Accessed September 30, 2008.
51 Association of American Medical Colleges. Medical School Objectives Projects. Available at: (http://www.aamc.org/meded/msop/start.htm). Accessed September 24, 2008.
52 Coombs RH. Mastering Medicine: Professional Socialization in Medical School. New York, NY: Free Press; 1978.
53 McKegney CP. Medical education: A neglectful and abusive family system. Fam Med. 1989;21:452–457.
54 Hundert EM, Hafferty F, Christakis D. Characteristics of the informal curriculum and trainees' ethical choices. Acad Med. 1996;71:624–642.
55 Hojat M, Mangione S, Nasca T, et al. An empirical study of decline in empathy in medical school. Med Educ. 2004;38:934–941.
56 Crandall SJ, Volk RJ, Loemker V. Medical students' attitudes toward providing care for the underserved: Are we training socially responsible physicians? JAMA. 1993;269:2519–2523.
57 Woloschuk W, Harasym PH, Temple W. Attitude change during medical school: A cohort study. Med Educ. 2004;38:522–534.
58 Coulehan J, Williams JD. Vanquishing virtue: The impact of medical education. Acad Med. 2001;76:598–605.
59 Dyrbye LN, Thomas MR, Shanafelt TD. Medical student distress: Causes, consequences, and proposed solutions. Mayo Clin Proc. 2005;80:1613–1622.
60 Newton BW, Barber L, Clardy J, Cleveland E, O'Sullivan P. Is there a hardening of the heart during medical school? Acad Med. 2008;83:244–249.
61 Dans PE. Self-reported cheating by students at one medical school. Acad Med. 1996;71(1 suppl):S70–S72.
62 Rennie SC, Rudland JR. Differences in medical students' attitudes to academic misconduct and reported behavior across the years—A questionnaire study. J Med Ethics. 2003;29:97–102.
63 Hafferty FW, Franks R. The hidden curriculum, ethics teaching and the structure of medical education. Acad Med. 1994;69:861–871.
64 Pololi L, Frankel RM, Clay M, Jobe AC. One year's experience with a program to facilitate personal and professional development in medical students using reflection groups. Educ Health. 2001;14:36–39.
65 Pololi L, Frankel RM. Small group teaching emphasizing reflection can positively influence medical students' values. Acad Med. 2001;76:1172–1173.
66 Kassebaum DG, Cutler ER. On the culture of student abuse in medical school. Acad Med. 1998;73:1149–1158.
67 Snyder BR. The Hidden Curriculum. New York, NY: Alfred A Knopf; 1971.
68 Feudtner C, Christakis CA. Do clinical clerks suffer ethical erosion? Students' perceptions of their clinical environment and personal development. Acad Med. 1994;69:670–679.
69 Wear D. Professional development of medical students: Problems and promises. Acad Med. 1997;72:1056–1062.
70 Cuff PA, Vanselow NA. Improving Medical Education: Enhancing the Behavioral and Social Science Content of Medical School Curricula [IOM report]. Washington, DC: The National Academies Press; 2004.
71 Haidet P, Stein H. The role of student–teacher relationships in the formation of physicians: The hidden curriculum as process. J Gen Intern Med. 2006;21:S16–S20.
72 National Institutes of Health. NIH Roadmap for Medical Research. Available at: (http://nihroadmap.nih.gov/overview.asp). Accessed September 30, 2008.
73 Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. Nature. 2005;435:737–738.
74 Becker HS, Geer B, Huges EC, Strauss AL. Boys in White. Chicago, Ill: University of Chicago Press; 1961.
75 Pololi L, Frankel R. Humanizing medical education through faculty development: Linking self awareness with teaching skills. Med Educ. 2005;39:154–162.
76 Pololi L, Clay M, Lipkin M, Hewson M, Kaplan C, Frankel R. Reflections on integrating theories of adult education into a medical school faculty development course. Med Teach. 2001;23:276–283.
77 Pololi L, Dennis K, Mitchell JA. A needs assessment of medical school faculty: Caring for the caretakers. J Contin Educ Health Prof. 2003;23:21–29.
78 Pololi L, Knight S, Frankel R. Helping medical school faculty realize their dreams: An innovative collaborative mentoring program. Acad Med. 2002;77:377–384.
79 Kirch D. AAMC President's Address 2007. Culture and the Courage to Change. Available at: (http://aamc.org/newsroom/pressrel/2007/071104a.htm). Accessed September 24, 2008.

The Culture of Academic Medicine: Faculty Perceptions of the Lack of Alignment Between Individual and Institutional Values

Linda Pololi, MBBS, MRCP[1], David E. Kern, MD, MPH[2], Phyllis Carr, MD[3], Peter Conrad, PhD[1], and Sharon Knight, PhD, RN[4]

[1]Women's Studies Research Center, Brandeis University, Waltham, MA, USA; [2]School of Medicine, Johns Hopkins University, Baltimore, MD, USA; [3]Boston University School of Medicine, Boston, MA, USA; [4]East Carolina University, Greenville, NC, USA.

BACKGROUND: Energized, talented faculty are essential to achieving the missions of academic medical centers (AMCs) in education, research and health care. The alignment of individuals' values with workplace experiences are linked to meaningfulness of work and productivity.

OBJECTIVE: To determine faculty values and their alignment with institutional values.

DESIGN: A qualitative hypothesis-generating interview study to understand the professional experiences of faculty and organizational approach in five AMCs that were nationally representative in regional and organizational characteristics. Analysis was inductive and data driven.

PARTICIPANTS: Using stratified, purposeful sampling, we interviewed 96 male and female faculty at different career stages (early career, plateaued, senior faculty and those who had left academic medicine) and diverse specialties (generalists, medical and surgical subspecialists, and research scientists).

APPROACH: Dominant themes that emerged from the data.

RESULTS: Faculty described values relating to excellence in clinical care, community service (including care for the underserved and disadvantaged), teaching, intellectual rigor/freedom and discovery, all values that mirror the stated missions of AMCs. However, many faculty also described behaviors that led them to conclude that their AMCs, in practice, undervalued excellence in clinical care, and their social and educational missions. Themes were seen across gender, career stage, race and discipline, except that female leaders appeared more likely than male leaders to identify incongruence of individual values and organizational practices.

CONCLUSIONS: In this study of five diverse medical schools, faculty values were well aligned with stated institutional missions; however, many perceived that institutional behaviors were not always aligned with individual faculty values.

KEY WORDS: medical faculty values; institutional values; career stage.
J Gen Intern Med 24(12):1289–95
DOI: 10.1007/s11606-009-1131-5

Prior Presentation: SGIM 32nd Annual Meeting, 2009: Plenary Research Presentation.
Received May 26, 2009
Revised July 21, 2009
Accepted September 9, 2009
Published online October 16, 2009

> Schools are the sanctuaries of our personal and civic values and incubators of intellect and integrity. The values that mark our community are the values most likely to be learned by our students.
> E. Grady Bogue[1]

Energized, creative and compassionate faculty are essential to achieving the tripartite mission of medical schools to train physicians, advance knowledge through research and provide high quality care to the communities they serve. High levels of faculty dissatisfaction,[2] attrition[3] and burnout[4,5] have been documented, but little research has focused on faculty values, their alignment with institutional values and the relationship of these to faculty's work experience.

An important correlate of job satisfaction and optimal performance is the meaningfulness of one's work.[5–6] Our and Wright's prior research on medical faculty suggested that values serve as motivators and that alignment of values with work may impact function and success.[7–9] Research in academic medicine has tended to focus on faculty satisfaction.[2,10] There has been a dearth of studies on the impact of values congruence. Values are beliefs or ideals about what is good or desirable and act as guiding principles for choices, attitudes and behaviors.[11–15] For example, a person who holds honesty as a prioritized value is less likely to cheat on tests than a person who prioritizes other values. Typically, an individual's values are acquired through interaction with family, peers and social systems.[16] They tend to be fairly stable over a lifetime.[17,18]

During in depth interviews with faculty, we identified deeply held professional values as expressed through being optimally energized in work, and explored their relationship to faculty's work experience and perceptions of institutional values.

METHODS

Setting and Participants

This study included faculty from five US medical schools engaged in an action research project: the National Initiative on Gender, Culture and Leadership in Medicine (C - Change).[19] The Initiative promotes an organizational culture in academic medicine that helps all faculty realize their potential.

The five schools were representative of different regions and organizational characteristics of medical schools, e.g., public vs. private ownership and NIH research intensive vs. primary care focus. Although the project addresses the needs of all faculty, it spotlights women, under-represented minority (URM) and generalist faculty. The project received IRB approval, and all participants gave written informed consent.

We selected equal numbers of medical faculty from each of the five C - Change schools through purposeful and chain sampling strategies.[20] Interviewees were invited to participate in a study of faculty experiences in academic medicine. Participants were stratified by gender, race/ethnicity, department/discipline and career stage ("early career," i.e., faculty for 2 to 5 years; "plateaued," i.e., faculty for >10 years who had not advanced as expected in rank and responsibility; faculty in leadership roles such as deans, departmental chairs and center directors (identified as "senior" in the quotations below), and former faculty who had left for a career outside academic medicine). Men and women interviewed were approximately equally divided among the career stages, but with fewer male early career participants since we reached data saturation in this category early in the study.

Data Collection

In 2006–2007, four research team members conducted 1-h semi-structured interviews (15% in person, 85% by telephone), which were audio-recorded and transcribed verbatim. Interview questions focused on aspirations of a career in medicine, energizing aspects of their work, barriers to advancement, leadership, power, values and work-family integration. The questions were open-ended, non-leading and unbiased in wording to permit the respondent to describe what was personally meaningful and salient. The interview guide was based on a pilot national series of 21 interviews of faculty conducted by the PI. Questions that specifically addressed values included**:** When have you felt most successful in your work? What do you see as valued at your institution? How do your personal values align or conflict with what you experience in academic medicine?

Data Analysis

We used an inductive and data-driven, grounded theory process of analysis.[21,22] The multidisciplinary research team identified codes for units of meaning in the masked transcripts. With coding consensus and aided by two research assistants, the team coded the 4,000 pages of narrative data. We stored and organized coded data utilizing Atlas.ti software, and identified patterns and themes emergent in the data. Data relevant to values were derived from the entire set of interviews. The example quotations in this paper illustrate dominant themes related to values.

RESULTS

Participant Characteristics

One hundred seventy faculty were invited to participate: 8 refused (usually due to time constraints), 54 did not respond, and we were unable to schedule interviews with 12 individuals, leaving 96 completed interviews. Participants were research scientists, medical and surgical sub-specialists, and generalists, 84% MD or DO, and 16% PhD. Women composed 55%, African Americans/Blacks 17%, Hispanics/Latinos 4% and generalists 20% (general internal medicine, family medicine and general pediatrics). On analysis, themes were not unique to any career stage, gender, race, school or discipline. Quotes were chosen to include a mix of gender, basic scientists, subspecialists and generalists. Quotes are identified by gender and career stage; discipline and race are not identified to protect anonymity.

Energizing Aspects of Work that Reflect Faculty Values

Dominant themes in the interviews, which delineated energizing aspects of their careers, reflected the faculty's core values in their work. Clinical caring, social mission, teaching, intellectual rigor, discovery and self-direction were themes that emerged from all categories of faculty.

Clinical Caring. Physicians often described how clinical care was energizing for them and how rewarded they feel when they help a patient. This sense of reward extended beyond curing disease and treating medical problems to "caring for people," ameliorating chronic disease, and getting to know and build trusting relationships with their patients.

> The thing that most energizes me is the one-on-one work with complicated children and their families…it's the clinical work that energizes me most. (male, early-career)

> I still think probably the most satisfying thing in my work is when I feel I've helped a patient in some way, that I've helped someone deal with a serious illness or helped them recover. (female, plateaued)

> I really do like the oneness of working with patients and getting to know them. Establishing rapport and getting them to open up to you and likewise, establishing trust. (male, left academic medicine)

> I take care of the frail elders. That really energizes me in a way that I can't explain. I love these people. I love my nursing home residents. Even after I leave a day here, I can go to the nursing home even though I'm tired and it will re-energize me. …you are more often than not looking at caring over curing, because these people are at a point where you are not going to cure anything, but you can provide them and their families with a lot of care. (female, senior)

Social Mission. Faculty valued highly the social mission of medicine to care for the underserved and disadvantaged who

would not otherwise be able to afford care. They expressed the desire to serve the community and address issues of diversity.

> I think that academic medicine aligns very closely with what I think is correct and ethical, in the sense that I can provide care to patients and I never have to ask anybody whether they have insurance or not. I don't ever have to ask anybody if they can pay. If you come to the hospital and you're sick, I'm going to take care of you. (female, early-career)

> …my values are to provide health care, education for the community at large as well as the underserved community, and to pay particular attention to diversity in people and thinking. (female, senior)

Responsibility for Medical Education. Excellence in teaching was held as an essential value by many faculty. Faculty frequently commented on the personal meaningfulness of their role as teacher and medical educator. Numerous faculty described teaching as one of the reasons for choosing to work in academic medicine:

> Part of it is teaching others how to become a scientist or a physician, and you're perpetuating this wonderful field. (female, senior)

> I like the mission of what we're doing. I mean, it's really fun being around our young students who are trying to figure what their place in the world is going to be and how to do a good job of it and so I find that inspiring. (male, senior)

> And I get my biggest fire from taking young minds and helping shape them into physicians–high quality physicians. I'm most successful in my work when I'm actually at the bedside. When I'm at the bedside with a learner. I feel like I'm giving out. I'm giving in a way that you can only give if you're with the patient. That's when I'm most gratified. (male, senior)

Intellectual Rigor, Discovery and Self-Direction. High level intellectual stimulation, pursuit of the advancement of knowledge through research and intellectual autonomy were identified by participants as highly valued and integral to their roles as faculty.

> I think academic medicine still does pride intellectual advancement, and an eagerness for exploration, and new knowledge; that fits with what I find rewarding. I value the flexibility and freedom to pursue my own intellectual ends. (male, senior)

Many spoke of the excitement of scientific inquiry—of constructing research questions and of scientific work that reveals new insights—of how this new understanding adds to the global understanding of life and disease—and of their great satisfaction in seeing their own discovery translated into clinical application. Along with this came the gratification of having one's ideas and new knowledge receive the external recognition and accolade of being published in scientific journals. Faculty saw this as a legacy of their own hard work, intellect and contribution to the biomedical sciences.

> I had my own lab,…there was an unbelievably driving passion to answer questions in a way that I would be adding to the information that would make children's lives better. (female, senior)

> I think the few moments that I had when I was realizing that I was figuring something out that people hadn't figured out before; I was seeing data that no one had ever seen before because my experiments had generated it! That was pretty exciting. (female, left academic medicine)

Others spoke of science as serving the social mission:

> My research also opened the door to the whole question of how we use race in medicine. So there was a social part and there's a science part, and I guess I would have to say that that's probably been the highest point. (female, senior)

Other aspects of work that were described as energizing were external recognition for accomplishments, relationships with trainees, being involved in student graduation ceremonies and other university functions, taking leadership roles and accomplishing policy change.

Non-Alignment of Faculty and Perceived Institutional Values

In contrast to the highly valued and energizing aspects of life in medical school, some disturbing and deeply felt issues emerged from the data that suggested that faculty often found themselves in a conflict situation where their own individual values were not aligned with the behaviors and expectations of the institution in which they worked.

Three predominant themes emerged in the data with respect to the non-alignment of faculty and institutional values: a sense of institutional betrayal of the public trust by academic medicine, values conflict with the institutional culture regarding ethical issues and discomfort with the expectation of self-promotion.

Public Trust. Many faculty were disheartened by their perception that academic medicine is at times betraying the public trust and that it has lost its social mission. Faculty voiced this as a major reason contributing to their dissatisfaction with or departure from academic medicine.

> So while we have this emerging technology and the ability to treat patients, we have no sense of social purpose or social policy. (male, early career)

Excellence in clinical care. Embedded in the sense of betrayal of the public trust were faculty perceptions that clinical care

was not adequately supported or valued by the institution. Even though the institutional mission stated excellence in patient care as a priority, faculty perceived a failure to provide support for this mission and noted that clinical excellence was not rewarded:

> Publications and being invited to speak at other institutions, getting a lot of grants; that is valued higher than patient care. If you were to ask somebody, "who is most accomplished?"—those people are not necessarily the ones most adept at patient care. (female, early career)

> I think that everybody has to re-examine what it is to be in academic medicine—and it really came to light to me about a year ago, when I was on a search committee for a division chief in another department. We were there in a circle interviewing, talking, and every single candidate that came in talked about how they needed to protect their faculty from clinical work. Yet on the other hand, patients come here expecting the most experienced and the most savvy clinicians because it's a big university academic medical center. (male, left academic medicine)

> In my discussions with the dean, he is always talking about "no money—no mission," and I understand what he is saying: if we can't keep the doors open financially, then we won't take care of any sick folks, we won't train any medical students. But my point to him was that if we lose sight of what we are here for, we have no reason to keep our doors open. I think the focus is too much on the bottom line, to the point where we talk about giving up what makes us physicians in the first place. (female, plateaued)

Community responsibility. Similarly, faculty voiced the notion that the medical school was "*hypocritical*" in its responsibility to its local community.

> But in an academic institution that doesn't value community, culture, partnerships, collaboration, I wouldn't have wanted to stay there. I would not—that's where I was going and what I valued. It was really a dead end. …I think academic institutions are still about me, the individual. They're not necessarily about community and collaboration. I think if academic medicine can figure this out, that they can change how they promote people, value collaboration, value community partnership, true partnership and what I mean by that is being able to share resources, share authority with the community, to share data, to take as long as it takes to develop a study, as long as it takes to help communities choose issues that are important to the community, which is the antithesis of the individual faculty member, isn't it? (male, left academic medicine)

> I truly believe in health care for everybody, that it's a basic human right, and that it's our goal to organize resources in our society to make sure that that can happen, and that does not happen in academic medicine. (male, left academic medicine)

Excellence in education. Faculty frequently commented on the institution's inadequate provision of support for education, of not rewarding and recognizing those faculty who excel in this area. They described a devaluing of efforts devoted to medical education.

> I run the resident clinic. We're not rewarded by the Medical School at all. Few people each year might be recognized, but for the ongoing day-to-day grind, we're not recognized by the Medical School for our efforts…We basically provide free labor, [for education] you might call it, for the school…people stay because they feel a dedication to education….(female, plateaued)

A related theme was that institutions seemed to be self-serving and self-perpetuating rather than serving of their constituents. This was likened to a corporate culture.

> I think universities are looking much more like corporations than they used to. I hear the same kind of business speak stuff at the university as I heard in the private, for-profit company that I worked for. In fact, I was noting that there were fewer differences than ever before, so I think it's become a business. And I don't think that's where education should be, and I also don't think that's where health care should be. (female, left academic medicine)

Conflict with Perceived Institutional Culture Regarding Ethical Issues. A number of interviewees described experiences of unethical and fraudulent behavior that they believed were condoned by senior faculty. These instances were described as examples of lack of alignment of their own values with those of the institution or leadership. A male leader described a situation of major unethical use of funding. At an executive level discussion about the event, he remembered the following:

> We sat in that room for quite a while and after about a half an hour when people were hemming and hawing, I said, "Isn't it pretty clear what we have to do? This man has to be fired." And, literally, a senior administrator of the institution said—the words actually came out of his mouth: "We can't do that, he's one of us." (male, senior)

This was an example of the unethical use of funds and the concept of being a 'club' member as more important than public trust or integrity. Other faculty described situations where the institution created or tolerated a situation where individuals could be motivated to be unethical in research, by placing greater emphasis and value on the funding amount and quantity of research, rather than the integrity of the research.

> Two of his research assistants, young women, came in and talked to me yesterday that they couldn't sleep for 2 weeks because they believe the person who is directly supervising them is fraudulently creating data for a research project. (male, senior)

> I think my personal values don't align terribly well with academic medicine, interestingly. Unfortunately, I find

> that sometimes I feel that studies are done for the sake of doing the study. When I actually look at it and say, "What value is that going to provide either to patients or to our knowledge or anything like that." I wonder, is it just because it's another paper? (female, early career)

A woman described times when she had been expected to lie by her supervisor and how she found this unacceptable:

> My chairman asked me to take over the new faculty, and he would give me these things that I was supposed to tell them we were going to do for them and I would say to him, "we don't do this for people. How can I tell people this is available when you and I both know we don't follow through on that?" He said, "Well, you have to because that's the only way they'll come here." Well, I don't lie. That's not what I do. Or, I said "we're lying to the people who are doing our school evaluations, we're putting things on paper that we do that we don't do, and we're not being fair to the students. The students think they're getting a good deal because they don't know any better, but you and I both know what we're doing here." And it was sort of like "well, we have to make money so the students are going to have to suck it up." (female, left academic medicine)

Another faculty member described her experiences with her supervisor:

> Like the kind of leaders I saw in my early career, including my division chief. There's no way I want to be in a situation where I have to be deceitful in order to get people to do what they need to do. Or where I would have to work on scheming and cover-ups, as a way of doing my job. And that's what I think he felt he had to do—hide money, lie about money or at least cook the books a little bit. And not be concerned about a student's career, a fellow's career, because you had financial obligations to meet. (female, plateaued)

Twelve of 16 female leaders interviewed mentioned a lack of alignment of their own values and practices they observed in their organizations, and 5 of 12 male leaders commented similarly. Both male and female leaders stated that they only act in accordance with their own values.

Self-Promotion. Another area where participants were concerned was that they perceived self-promotion as necessary for survival and success. Some faculty commented on how distasteful they found being required "*to brag*" about themselves and that they found this behavior to be out of line with a personal value of being humble and more dedicated to achieving good than to personal aggrandizement. The following are illustrative:

> She said you have to brag. She said you have to brag, you really do. And that's very difficult, and I think that's difficult for many people because it's not the nature of some of our cultures, experience and maybe just family culture, too. It's not there for some of us. (female, senior)

> I wasn't driven by the self-promotion that I think has to come on in academic institutions. So that self-promotion becomes more important than the work from an academic, non-tenured faculty member. All of a sudden, they lose sight of why they're doing what they're doing. It's all about getting new grants, it's a club of 'I have more than you and I'm the expert in this.' So it's a little bit of unreal, self-promoting kind of environment. (male, left academic medicine)

DISCUSSION

Our research adds to the literature on the culture of academic medicine by more comprehensively and explicitly identifying faculty values, and faculty perceptions of the lack of alignment of their own and perceived institutional values. Faculty reported being most energized when they were engaged in clinical caring, the social mission of medicine to provide excellent care for all patients regardless of means, teaching, intellectual stimulation and advancement of knowledge. These valued activities aligned with the stated values of their five institutions. The faculty members we interviewed often inferred the values of their institution by observing behaviors and actions. They reported a significant lack of alignment between their own and perceived institutional values. In particular, numerous faculty perceived a lack of attention to the social mission of providing care for all people and to the community, a lack of prioritization of excellence in clinical care, a devaluing of educational roles, questionable ethical behavior among leadership or management, and the necessity for self-promoting behavior to achieve success. Values incongruence was associated with dissatisfaction, demoralization and sometimes with intent to leave their institution or academic medicine. Several quotations were from former faculty, but the same perspectives were expressed within all career categories.

Others have written about the link between authenticity and productivity. Authenticity reflects acting in accordance with one's values, preferences and needs, as opposed to acting merely to please others, to attain rewards or avoid punishments.[6,23,24] Faculty are more likely to instill a passion for medicine in their students or conduct stellar research if they are working on something that they are personally passionate about and that is aligned with their values.[6] The contrasting state of 'burnout' results in lesser performance and 'depersonalization' or the absence of bringing one's personal self to work.[24] Literature from other fields suggests that institutions need the ideas, self-expression, questioning and creativity that comes from empowering employees.[25,26]

Faculty values aligned well with the stated missions of most medical schools: clinical care, education and research. However, faculty based their perceptions of institutional values on observed behaviors rather than mission statements. An organization achieves congruence when its espoused principles and actions are aligned; our faculty frequently reported the lack of such congruence. Outside medicine, Waterman,[27] found that nine companies that practiced according to their values outperformed the Dow Jones industrial average by 350%. Collins also found that

organizations were most successful where their values were embodied in the fabric of the organization, in its systems, practices, process and rewards.[28]

Our results align with the research findings[29] that women value consistently more than men benevolence and universalism (understanding, appreciation, tolerance and protection for the welfare of all people), and female physicians are more motivated by helping others than males.[8] This would suggest that lack of alignment of values as shown in our data may contribute to women's lack of advancement in academic medicine. Another study from our interview data set shows that URM faculty report a call to serving their own underserved communities.[30] The latter data suggest that values incongruence may be one factor contributing to the difficulty of academic medicine in recruiting and retaining URM faculty members.[31,32]

One limitation of our study is that our data were drawn from just five medical schools, but the schools were chosen to be representative of the nation in regional and organizational characteristics, and their faculty demographics at the time of this study were almost identical to national means. However, for a qualitative hypothesis-generating study, the large number of respondents and multi-institutional sample are strengths. The responses may not be representative of the responses of all faculty, but do come from a diverse group of faculty in terms of gender, discipline, career stage and race within each of the five disparate medical schools. The themes we heard were dominant in the data and evident across all career stages of faculty. To assess for generalizability, the findings from this study are being tested in a nationally representative survey of faculty from 26 medical schools. Another strength of the study is that it involved a research team from different disciplines and used accepted and rigorous approaches to hypothesis- generating qualitative research. The carefully maintained confidentiality and anonymity of interviews protected respondents and probably increased their willingness to be frank. Many interviewees commented that they were grateful to have the opportunity to express their views. However, this strength is also a limitation as it prevented us from analyzing data by subgroups, which might have jeopardized the anonymity of participants.

By identifying faculty's deeply held values as expressed when they feel most vital and successful, we hope that this study will increase medical school leaders' awareness and promote congruence between individual values and institutional values, assisting the realization of the full potential and contributions of a diverse faculty. What may be more challenging are faculty perceptions that the stated social and educational missions of academic medical centers, which are well aligned with their individual values, are not fully congruent with institutional behaviors. The findings of this study should encourage academic medical institutions to address these faculty perceptions of the culture.

Acknowledgements: *The authors gratefully acknowledge the critical funding support of the Josiah Macy, Jr. Foundation, Brandeis University Women's Studies Research Center, and Wanda Jones and Anna Kindermann for leading the interagency agreement to provide supplemental support by the Office of Public Health and Science, Office on Women's Health and Office on Minority Health; the National Institutes of Health, Office on 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.*

The authors thank Meg Lovejoy and Marianne MacPherson, who participated in 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.

Conflict of Interest: *None of the authors have any conflict of interest to disclose. The authors had access to all the study data, take responsibility for the accuracy of the analysis, and had authority over manuscript preparation and the decision to submit the manuscript for publication. The funders had no involvement in the design of the study, the collection of data, analysis and interpretation of the data, and the decision to approve publication of the finished manuscript. Dr. Pololi had full access to all of the data in this study and takes responsibility for the integrity of the data and the accuracy of the data analysis.*

Specific contributions from each author: *Pololi: conception, design, data collection, analysis and interpretation, drafting the article, final approval*
Kern: interpretation, drafting the article, final approval
Carr: design, data collection, analysis and interpretation, drafting the article, final approval
Conrad: design, data collection, analysis and interpretation, drafting the article, final approval
Knight: design, data collection, analysis and interpretation, drafting the article, final approval

Corresponding Author: *Linda Pololi, MBBS, MRCP: Brandeis University, Mailstop 088, Women's Studies Research Center, Waltham, MA 02454-9110, USA (e-mail: lpololi@brandeis.edu, URL: http://cchange.brandeis.edu).*

REFERENCES

1. **Bogue EG.** An agenda of common caring: the call for community in higher education. In: McDonald W, ed. Creating Campus Community: In Search of Ernest Boyer's Legacy. 1st ed. San Fransisco: Jossey-Bass; 2002;1–19.
2. **Schindler BA, Novack DH, Cohen DG, Yager J, Wang D, Shaheen NJ, Guze P, Wilkerson L, Drossman D.** The impact of the changing health care environment on the health and well-being of faculty at four medical schools. Acad Med. 2006;81:27–34.
3. **Alexander H, Lang J.** The Long-Term Retention and Attrition of US Medical School Faculty. Analysis in Brief, 8(4). Washington DC: Association of American Medical Colleges; 2008.
4. **Shanafelt TD, Sloan JA, Haberman TM.** The well-being of physicians. American J Med. 2003;114:513–9.
5. **Shanafelt TD, West CP, Sloan JA, Novotny PJ, Poland GA, Menaker R, Rummans TA, Dyrbye LN.** Career fit and burnout among academic faculty. Arch Intern Med. 2009;169(10):990–5.
6. **Kahn WA.** To be fully there: psychological presence at work. Hum Relat. 1992;45:321–49.
7. **Pololi LH, Dennis K, Winn GM, Mitchell J.** A needs assessment of medical school faculty: caring for the caretakers. J Contin Educ Health Prof. 2003;23:21–9.
8. **Wright SM, Beasley BW.** Motivating factors for academic physicians within departments of medicine. Mayo Clin Proc. 2004;79:1145–50.
9. **Pololi LH, Knight SM, Dennis K, Frankel RM.** Helping medical school faculty realize their dreams: an innovative, collaborative mentoring program. Acad Med. 2002;77:377–84.
10. **Bunton SA.** US Medical School Faculty Job Satisfaction. Analysis in Brief:8(5). Washington DC: Association of American Medical Colleges; 2008.
11. **Schwartz SH.** Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In: Zanna MP, ed. Advances in Experimental Social Psychology. Vol. 1. New York: Academic Press; 1992;1–65.
12. **Halstead JM, Taylor MJ.** Learning and teaching about values: a review of recent research. Camb J Educ. 2000;30(2):169–202.
13. **Pendelton D, King J.** Values and leadership. BMJ. 2002;325:1352–5.
14. **Pololi L.** Career development for academic medicine–a nine step strategy. BMJ Careers. 2006;322(7535):38–9.

15. **Rokeach M.** A theory of organization and change within value-attitude systems. J Soc Issues. 1968;24(1):13–33.
16. **Feather NT.** Values in Education and Society. New York: Free Press; 1975.
17. **Rokeach M.** The Nature of Human Values. New York: Free Press; 1973.
18. **Rokeach M.** Understanding Human Values: Individual and Societal. New York: Free Press; 1979.
19. National Initiative on Gender, Culture and Leadership in Medicine: C - Change. Available at: http://cchange.brandeis.edu accessed Aug 29, 2009.
20. **Biernacki P, Waldorf D.** Snowball sampling: problems and techniques of chain referral sampling. Sociol Methods Res. 1981;10:141–63.
21. **Glaser BG, Strauss AL.** The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine; 1967.
22. **Charmaz K.** Constructing Grounded Theory: a Practical Guide Through Qualitative Analysis. Thousand Oaks: Sage Publications; 2006.
23. **Goldman BM, Kernis MH.** The role of authenticity in healthy psychological functioning and subjective well-being. Ann Am Psychother Assoc. 2002;5(6):18–20.
24. **Maslach C.** Burnout: the Cost of Caring. Englewood Cliffs: Prentice-Hall; 1982.
25. **Van Maanen J, Schein E.** Toward a theory of organizational socialization. In: Shaw B, ed. Research in Organizational Behavior. Greenwich: JAI Press; 1979;209–64.
26. **Weisbord MR.** Productive Workplaces: Organizing and Managing for Dignity, Meaning and Community. San Francisco: Jossey-Bass; 1987.
27. **Waterman RH.** Adhocracy: the Power to Change. New York: Norton; 1992.
28. **Collins JC, Porras JI.** Built to Last: Successful Habits of Visionary Companies. New York: Harper Business; 1994.
29. **Schwartz SH, Rubel T.** Sex differences in value priorities: Cross-cultural and multimethod studies. J Pers Soc Psychol. 2005;89(6):1010–28.
30. **Pololi L, Cooper L, Carr P.** Race, Disadvantage, and Faculty Experiences in Academic Medicine, submitted for publication.
31. Association of American Medical Colleges. Faculty Roster 2008. Available at: http://www.aamc.org/data/facultyroster/ accessed Aug 29, 2009.
32. **Castillo-Page L.** Diversity in Medical Education: Facts & Figures 2008. Washington: Association of American Medical Colleges; 2008.

JOURNAL OF WOMEN'S HEALTH
Volume 18, Number 10, 2009

DOI: 10.1089/jwh.2008.1260

Original Article

The Heart Truth Professional Education Campaign on Women and Heart Disease: Needs Assessment and Evaluation Results

Janet Pregler, M.D.,[1] Karen M. Freund, M.D., M.P.H.,[2] Mary Kleinman,[3] Maureen G. Phipps, M.D., M.P.H.,[4] Rose S. Fife, M.D., M.P.H.,[5] Becky Gams, R.N., M.S., C.N.P.,[6] Ana E. Núñez, M.D.,[7] Margaret R. Seaver, M.D., M.P.H.,[8] Cathy J. Lazarus, M.D.,[9] Nancy C. Raymond, M.D.,[6] Joan Briller, M.D.,[3] Sebastian Uijtdehaage, M.D.,[10] Cindy S. Moskovic, M.S.W.,[1] Gretchen Guiton, Ph.D.,[11] Michele David, M.D., M.B.A., M.P.H.,[2] Geralde V. Gabeau, M.P.H.,[2] Stacie Geller, Ph.D.,[12] Kelli Meekma,[13] Christopher Moore,[14] Candace Robertson, M.P.H.,[7] and Gloria Sarto, M.D., Ph.D.[15]

Abstract

Background: Heart disease is the leading cause of death for women in the United States. Research has identified that women are less likely than men to receive medical interventions for the prevention and treatment of heart disease.

Methods and Results: As part of a campaign to educate healthcare professionals, 1245 healthcare professionals in 11 states attended a structured 1-hour continuing medical education (CME) program based on the 2004 AHA Evidence-Based Guidelines for Cardiovascular Disease Prevention in Women and completed a pretest and posttest evaluation. We identified significant knowledge deficits in the pretest: 45% of attendees would initially recommend lifestyle changes alone, rather than statin therapy, for women diagnosed with coronary artery disease (CAD); 38% identified statin therapy as less effective in women compared with men for preventing CAD events; 27% identified Asian American women at low risk (rather than high risk) for type 2 diabetes mellitus (DM); and 21% identified processed meat (rather than baked goods) as the principal dietary source of trans fatty acids. Overall, healthcare professionals answered 5.1 of 8 knowledge questions correctly in the pretest, improving to 6.8 questions in the posttest ($p < 0.001$). Family physicians, obstetrician/gynecologists, general internists, nurse practitioners/physician assistants, and registered nurses all statistically significantly improved knowledge and self-assessed skills and attitudes as measured by the posttest.

Conclusions: Significant knowledge deficits are apparent in a cross-section of healthcare providers attending a CME lecture on women and heart disease. A 1-hour presentation was successful in improving knowledge and self-assessed skills and attitudes among primary care physicians, nurse practitioners, physician assistants, and registered nurses.

[1]Iris Cantor-UCLA Women's Health Center, Los Angeles, California.
[2]Boston University Center of Excellence in Women's Health, Boston, Massachusetts.
[3]University of Illinois at Chicago, Illinois.
[4]Brown University/Women & Infants Hospital, Providence, Rhode Island.
[5]Indiana University School of Medicine, Indianapolis, Indiana.
[6]Deborah E. Powell Center for Women's Health, Chicago, Illinois.
[7]Drexel University College of Medicine, Philadelphia, Pennsylvania.
[8]Boston University School of Medicine, Boston, Massachusetts.
[9]Chicago Medical School, Rosalind Franklin University of Medicine and Science, Chicago, Illinois.
[10]David Geffen School of Medicine, Los Angeles, California.
[11]University of Colorado, School of Medicine, Boulder, Colorado.
[12]UIC Center for Research on Women and Gender, University of Illinois, Chicago, Illinois.
[13]Center for Women's Health Research, University of Wisconsin, Madison, Wisconsin.
[14]Christiana Care Health Initiatives, Newark, Delaware.
[15]University of Wisconsin School of Medicine, Madison, Wisconsin.

Introduction

The Heart Truth Professional Education Campaign was developed by the National Heart, Lung, and Blood Institute (NHLBI) in conjunction with the U.S. Department of Health and Human Services Office on Women's Health (DHHS/OWH), the American Heart Association (AHA), and other partner organizations to address the lack of heart disease awareness among women. This campaign was based on recommendations from experts who convened in 2001 to develop a national action plan to reduce heart disease in women.[1] The action plan recommended that women speak with their healthcare provider about specific risks that women have for cardiovascular disease (CVD); therefore, a provider education campaign component was commissioned by the NHLBI through the U.S. DHHS/OWH in 2003.

Lack of healthcare provider knowledge of the guidelines for prevention of CVD in women has been identified as a barrier to reducing morbidity and mortality in this population. Historically, studies have shown that women receive less cholesterol screening, less lipid-lowering therapies, less use of heparin, beta-blockers and aspirin during myocardial infarction (MI), and fewer referrals to cardiac rehabilitation compared with men. These studies suggest that women are less likely than men to receive guideline-recommended care across the spectrum of CVD prevention and treatment, including those within the scope of primary care.[2–5] There is further evidence that minority women may be less likely to receive appropriate treatment than white women even when access to care is similar or identical.[6–10] We conducted the following study to assess healthcare providers' baseline knowledge of CVD and to determine the effects of a structured educational program on provider knowledge and self-assessed skills.

Materials and Methods

The Heart Truth Professional Education Campaign was developed by a consortium of federally designated National Centers of Excellence in Women's Health, based at academic medical centers, and National Community Centers of Excellence in Women's Health, based in community health centers, community hospitals, and other community-based organizations providing healthcare. The Heart Truth Professional Education Campaign educational objectives were based on the 2004 Guidelines for the Prevention of Cardiovascular Disease in Women developed by the AHA in consultation with NHLBI and other stakeholders.[11] Educational materials development was informed by an advisory panel including experts from multiple medical associations and federal expert panels, as well as focus groups of women consumers who commented on information they wanted their healthcare providers to know and communicate.

Provider education materials included continuing medical education (CME) lectures, additional slide resources, web-based CME modules posted on Medscape®, standardized patient and problem-based learning materials for medical and nursing students, printed guides to web-based resources for diagnosis, prevention, and treatment of CVD in women, a bibliography, and a static website providing access to the materials. Versions of these materials (updated to reflect the 2007 AHA/NHLBI Evidence-Based Guidelines for the Prevention of Cardiovascular Disease in Women) can be accessed at www.womenshealth.gov/hearttruth/.

The CME lecture component was developed as a 1-hour lecture. Lecturers selected a slide set from three standardized sets of materials, all with the same 31 essential slides addressing guidelines for the prevention of heart disease in women (Appendix). The essential slides were supplemented with 15–20 slides that provided additional information on guidelines topics. At each lecture, the presenter chose which supplemental slides to include based on the presenter's assessment of the audience's knowledge base and interests. This format was based on expert recommendations that presentations should be available to address the interests of three main groups: (1) healthcare professionals who do not themselves prescribe medications for heart disease (expanded information on counseling and lifestyle information), (2) healthcare professionals judged to have average interest and experience in preventing, diagnosing and treating heart disease (expanded information most relevant to basic guidelines), and (3) healthcare professionals judged to have above-average interest and experience in preventing, diagnosing and treating heart disease (expanded information on recent controversies, e.g., folic acid use, high-sensitivity C-reactive protein (hsCRP) testing). Lecture materials included annotated discussion of the information on each slide and references to support facts presented. Presenters were instructed not to deviate from the lecture format in order to standardize the material presented.

The CME lecture was delivered by a local or regional expert on heart disease and women. The choice of qualifications for the speaker was made based on the expected composition of the audience. These expert speakers included generalist and specialist physicians, advanced practice nurses, and registered nurses with cardiovascular expertise. CME venues included grand rounds presentations and other clinic and hospital-based CME programs as well as free-standing CME programs. Sixty-eight presentations were made in 11 states (Arizona, Connecticut, Delaware, Massachusetts, Minnesota, Missouri, Ohio, Pennsylvania, Rhode Island, Vermont, and Wisconsin) during an 18-month period from July 2005 through December 2006.

Participants were prompted by the presenter and a slide to complete a pretest form before the lecture and a posttest form after the lecture. The pretest-posttest evaluation questions were formulated by experts and pilot tested by groups of healthcare providers of varying background and expertise. The pretest form contained five demographic questions, five self-assessment of knowledge and preparedness (skills) questions, and eight multiple-choice knowledge assessment questions. The posttest contained four questions soliciting feedback on the program in addition to the original five self-assessment of knowledge and preparedness (skills) questions and eight multiple choice knowledge assessment questions from the pretest. Knowledge questions were drawn from content addressing the following curricular objectives: risk stratification, lifestyle modification, pharmacotherapy, and implications of race and ethnicity in CVD prevention in women.

Following the lecture, participants were asked to complete a pretest-posttest evaluation form containing information about the research portion of the evaluation. This form was completed anonymously, and participants were given the option to exclude their information from being used for research. The research portion of the project received Institu-

tional Review Board approval or exemption from all participating institutions. Statistical analyses were carried out with SPSS version 15.0 (SPSS, Chicago, IL).

Results

Of the 2155 healthcare professionals attending the CME lectures, 1285 (59.6%) completed and returned the pretest-posttest form. No information is available about the 40% of CME attendees who did not return the pretest-posttest form. One CME lecture was excluded from our data analysis because no attendance information or pretest forms were obtained. Practice characteristics of providers are shown in Table 1. Subgroup analysis was performed for family physicians, general internists, obstetrician/gynecologists, nurse practitioners/physician assistants, and registered nurses. Nurse practitioners and physician assistants are grouped for purposes of reporting because no significant differences were found between these two groups.

Knowledge assessment

Table 2 shows the pretest and posttest results for all healthcare professionals. Subgroups included family physicians, obstetrician/gynecologists, general internists, nurse practitioners/physician assistants, and registered nurses. A mixed-model ANOVA demonstrated a significant interaction effect between subgroups and a gain in knowledge as well as baseline differences.

Overall, healthcare professionals answered 5.1 of 8.0 knowledge questions correctly in the pretest. Family physicians were significantly more knowledgeable than obstetrician/gynecologists at pretest ($p < 0.001$). Registered nurses were significantly less knowledgeable than other groups ($p < 0.001$). Healthcare professionals improved their knowledge scores at posttest to 6.8 of 8.0 questions; this is a 28% increase over the pretest knowledge scores ($p < 0.001$) (Fig. 1). The increase in posttest over pretest scores was significantly improved for each of the subgroups ($p < 0.001$). At posttest, there was one statistically significant difference between subgroups: registered nurses were significantly less knowledgeable than family physicians, obstetrician/gynecologists, and nurse practitioners/physician assistants ($p < 0.001$).

The pretest revealed significant knowledge deficits (defined as $\geq$20% of attendees endorsing a specific wrong answer) before the CME lecture. These deficits included the following: 45% of attendees would initially recommend lifestyle changes alone (rather than statin therapy) for women diagnosed with CVD; 38% identified statin therapy as less effective in women compared with men for preventing CVD events; 27% identified Asian American women as having a lower risk (rather than higher risk) for type 2 diabetes mellitus (DM); and 21% identified processed meat (rather than baked goods) as the principal dietary source of trans fatty acids.

At posttest, there was a significant improvement in the knowledge deficits for all questions; however, 30% of attendees continued to endorse lifestyle changes alone initially, rather than statin therapy, for women diagnosed with CVD, and 22% identified statin therapy as less effective in women compared with men for preventing CVD events.

Prior to the CME lecture, attendees who indicated they treat patients with heart disease in clinical practice were significantly more knowledgeable than attendees who indicated they do not treat patients with heart disease ($p < 0.001$). After the CME lecture, there was no significant difference in knowledge between these two subgroups.

Table 1. Practice Characteristics of Lecture Attendees: The Heart Truth Professional Education Campaign Continuing Medical Education

Specialty	*Number*	*%*
Family physician	162	13
General internist	131	10
Obstetrician/gynecologist	151	12
Cardiologist	9	0.7
Other M.D. or D.O.[a]	94	7
Nurse midwife	29	2
Nurse practitioner	151	12
Physician assistant	25	2
Registered nurse	346	27
Other	181	14
Missing	6	0.5
Weekly clinical activity of attendees		
0–25 patients		37
26–50 patients		23
51–75 patients		19
76–99 patients		13
>99 patients		8
Percentage of patients who are women		
0–25%		9
26–50%		28
51–75%		31
75–99%		24
100%		9
Total	1285	100

[a]M.D., Doctor of Medicine; D.O., Doctor of Osteopathic Medicine.

Self-assessment of knowledge

Self-assessment of knowledge was judged on a 5-point Likert scale from Not at all knowledgeable to Very knowledgeable for two issues: (1) knowledge about current approaches to smoking cessation, exercise, weight management, and diet to reduce risk for CVD in women and (2) knowledge about goals for major risk factor interventions to prevent CVD in women, including goals for management of blood pressure, lipids, and DM. Mean results for the pretest and posttest are presented in Table 3. All subgroups showed a highly significant improvement in self-assessed knowledge after the CME session ($p < 0.001$). At pretest, family physicians rated themselves as more knowledgeable than the other subgroups. This difference disappeared at posttest, with the exception of registered nurses; they continued to rate themselves as less knowledgeable than other groups despite their improvement after the CME session.

Self-assessment of preparedness (skills)

Self-assessment of skills was judged on a 5-point Likert scale from Not at all prepared to Completely prepared for three issues: (1) preparedness to assess and stratify women into high, intermediate, low, and optimal risk categories for heart disease, (2) preparedness to counsel women about the

Table 2. Pretest-Posttest Knowledge Scores by Specialty: % Correct by Learning Objective: The Heart Truth Professional Education Campaign

	All		*General internist*		*Family practitioner*		*OB/GYN*[a]		*PA/NP*		*RN*	
Objective	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
Identify factors that place women at high risk (>20% over 10 years) for CVD event	89.3	94.3	91.3	92.1	91.6	98.1	85.6	95.2	88.3	96.6	91.2	94.0
Identify correct information about use of statins in women to prevent CVD events	30.2	64.6	50.8	77.0	55.5	72.9	28.1	69.9	39.0	70.7	13.6	55.3
Identify that African Americans, Latinas, and Asian Americans are not at low risk of type 2 DM	59.4	89.7	68.3	86.5	66.5	92.9	54.8	89.0	66.3	95.1	58.6	90.0
Identify primary dietary source of trans fatty acids	62.6	86.8	55.6	69.8	67.7	89.7	55.5	87.0	74.6	91.2	58.0	88.5
Identify that postmenopausal hormone therapy and antioxidant vitamin supplements are not recommended for CVD prevention	64.3	91.6	89.7	96.8	87.1	92.9	78.0	96.6	77.6	95.1	41.7	88.8
Identify that African Americans are most likely to die from heart disease among women of all races and ethnicities	76.5	96.3	81.7	96.8	82.6	96.1	84.9	96.6	81.5	96.6	70.1	97.3
Identify facts related to smoking cessation in women	55.9	75.6	59.5	75.4	65.2	78.7	64.4	81.5	56.6	79.0	47.4	67.4
Identify facts about treatment of HTN in women	62.3	82.3	74.6	85.7	78.1	81.3	61.6	85.6	65.4	83.9	56.5	84.0

[a]OB/GYN, obstetrician/gynecologist; PA, NP, Physicians assistant/nurse practitioner; RN, registered nurse; CVD, cardiovascular diseases; DM, diabetes mellitus; HTN, hypertension.

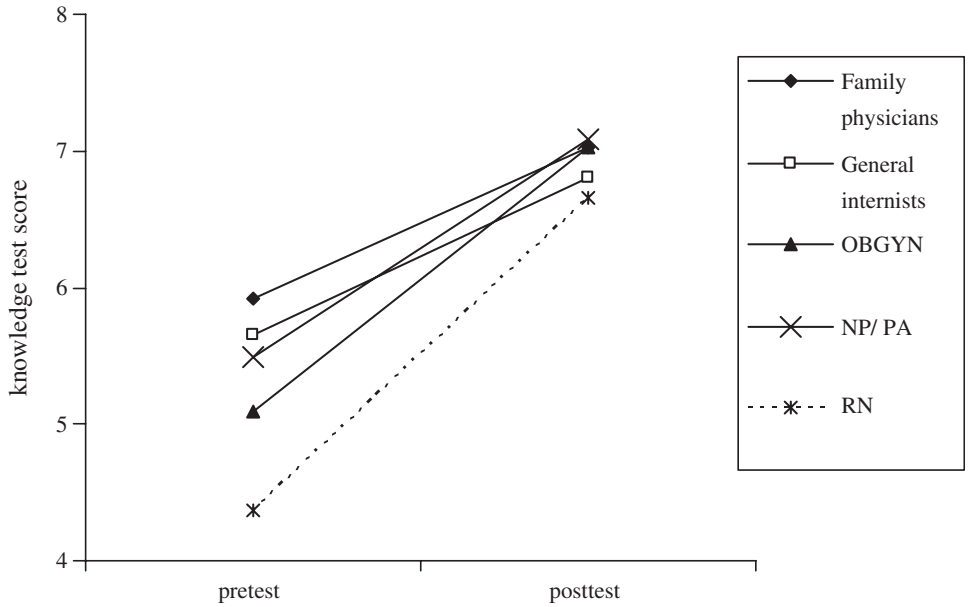


FIG. 1. Knowledge test scores on the pretest and posttest: The Heart Truth Professional Education Campaign. Knowledge test scores on the pretest and posttest for family physicians, general internists, obstetrician/gynecologists, nurse practitioners/physicians assistants, and registered nurses. The test scores indicate the number of questions answered correctly out of 8.0 questions. All subgroups showed a significant improvement in knowledge scores between pretest and posttest ($p < 0.001$).

Table 3. Pretest-Posttest Self-Assessment of Knowledge and Preparedness: The Heart Truth Professional Education Campaign[a]

		n	*Pretest* M[b]	*Pretest* *SD*	*Posttest* *M*	*Posttest* *SD*
Self-assessment of knowledge question						
How knowledgeable are you about current approaches to smoking cessation, exercise, weight management and diet to reduce risk for CVD in women?	Family physician	155	3.9	±0.6	4.1	±0.7
	General internist	121	3.7	±0.8	4.0	±0.7
	OB/GYN	143	3.4	±0.8	3.8	±0.7
	NP/PA	200	3.8	±0.8	4.1	±0.7
	RN	324	3.2	±0.9	3.7	±0.8
	Total	943	3.5	±0.9	3.9	±0.8
How knowledgeable are you about goals for major risk factor interventions to prevent CVD in women?	Family physician	152	4.0	±0.6	4.3	±0.6
	General internist	120	3.8	±0.9	4.3	±0.7
	OB/GYN	141	3.1	±0.8	3.8	±0.7
	NP/PA	198	3.6	±1.0	4.0	±0.8
	RN	316	3.0	±0.9	3.5	±0.9
	Total	927	3.4	±1.0	3.9	±0.8
Self-assessment of preparedness question						
How prepared are you to assess and stratify women into high, intermediate, lower, and optimal risk categories for CVD?	Family physician	155	3.6	±0.8	4.1	±0.7
	General internist	123	3.5	±0.8	4.0	±0.7
	OB/GYN	142	2.9	±0.9	3.7	±0.7
	NP/PA	198	3.2	±1.0	3.8	±0.8
	RN	320	2.5	±1.0	3.4	±0.8
	Total	938	3.0	±1.0	3.7	±0.8
How prepared are you to counsel a woman who asks about the use of hormone therapy, antioxidant supplements, or aspirin to reduce coronary disease event risk with up-to-date information?	Family physician	149	3.5	±0.8	4.1	±0.7
	General internist	121	3.4	±0.9	4.1	±0.7
	OB/GYN	141	3.1	±1.0	3.9	±0.8
	NP/PA	197	3.0	±1.0	3.9	±0.8
	RN	316	2.2	±1.0	3.4	±0.9
	Total	924	2.9	±1.0	3.8	±0.9
How prepared are you to prevent, evaluate, and treat heart disease in women of diverse racial and ethnic backgrounds?	Family physician	152	3.2	±0.8	4.0	±0.7
	General internist	120	3.3	±0.9	4.0	±0.7
	OB/GYN	141	2.4	±0.9	3.4	±0.9
	NP/PA	198	2.7	±1.0	3.6	±0.9
	RN	316	2.2	±0.9	3.2	±1.0
	Total	927	2.6	±1.0	3.6	±0.9

[a]Mean scores are based on a 5-point Likert scale, with 1 defined as Not at all knowledgeable, and 5 defined as Very knowledgeable.
[b]M, mean; SD, standard deviation; CVD, cardiovascular diseases; OB/GYN, obstetrician/gynecologist; RN, registered nurse; NP/PA, nurse practioner/physician assistant.

use of hormone therapy, antioxidant supplements, and aspirin to reduce CVD event risk with up to date information, and (3) preparedness to prevent, evaluate, and treat heart disease in women of diverse racial and ethnic backgrounds. Mean results for the pretest and posttest are presented in Table 3. Family physicians and general internists felt better prepared to stratify women into risk categories and to prevent, evaluate, and treat heart disease before and after the CME session compared with the other groups. All groups, however, showed a significant posttest improvement ($p < 0.001$) on all three items.

Program feedback

Attendees ranked the program highly. On a 5-point Likert scale with 1 defined as Strongly disagree and 5 defined as Strongly agree, the mean score was 4.4 (±0.6) that the information was credible, 4.0 (±0.8) that the information presented would change how they treated their patients, 4.3 (±0.7) that they felt more knowledgeable about heart disease in women, and 4.2 (±0.8) that the information presented will help them better care for patients of diverse racial and ethnic backgrounds.

Discussion

A standardized 1-hour didactic session with evidence-based materials tailored toward the presenter's assessment of audience needs improved healthcare providers' knowledge of gender-specific prevention and treatment issues in CVD. The session also increased providers' self-assessed efficacy in addressing heart disease in women. The educational program was widely accepted by a range of healthcare professionals, including primary care physicians of various specialties, nurses, and nurse practitioners and physician assistants. Despite knowledge gains, significant knowledge gaps persisted after CME training.

CVD remains the most common reason for death in women in the United States. To reduce the incidence of CVD in women, efforts have focused on educating healthcare consumers about risk and supporting gender-specific research about CVD. Knowledge gaps identified in this study

demonstrate that specific strategies to disseminate knowledge of CVD in women to healthcare professionals are urgently needed.

The baseline knowledge assessment revealed some interesting gaps in knowledge. For example, a third or more of respondents were unaware that statins are recommended for all women with known CVD and that statins have specific benefits for women. Few studies have systematically assessed gender-specific gaps in provider knowledge. In one study, 300 primary care physicians, 100 obstetrician/gynecologists, and 100 cardiologists were asked to respond to experimental cases of CVD in women and men; this study found that women in the experimental cases were more likely to be inappropriately assigned to low-risk categories than men. Women who were inappropriately identified as low risk were less likely to receive appropriate recommendations for lifestyle change and preventive pharmacotherapy.[12] More information is needed about gender-specific gaps in provider knowledge to better understand how to tailor gender-specific information in CME programs to ensure that women receive evidence-based care from their providers.

Healthcare providers in our study demonstrated knowledge of the greater prevalence of risk factors for CVD in the African American population; however, they had less knowledge about risk factors in Asian American women. Other studies have shown racial disparities in care even after controlling for socioeconomic status (SES), including one study that found cholesterol-lowering drugs are less commonly used after MI by African Americans than by whites, even after controlling for socioeconomic status.[9] Another study assessing racial disparities in clinical evaluation of CVD practice found that when presented with identical case histories deemed by experts to warrant further cardiac testing, physicians were significantly less likely to suggest cardiac evaluation for an African American woman than for either a white or African American man or a white woman. In this study, clinicians evaluated identical case histories presented by videotaped actors, so that the scenario varied only by the apparent race and gender of the patient.[10]

Providers' lack of knowledge about trans fatty acids may reflect the timing of the intervention. CME presentations occurred before the announcement of proposed government regulations to reduce dietary trans fatty acids, which were publicized at the end of 2006.[13] However, lack of knowledge about dietary trans fatty acids may also accurately reflect healthcare providers' current lack of knowledge and training about dietary issues.

A limitation of the study was the use of self-assessment to measure providers' preparedness and skills. Providers self-selected to attend the program; this may have resulted in a group that was more knowledgeable or less knowledgeable than providers as a whole. The effectiveness of the program might be different if it were made mandatory or if presenters were more or less expert in the topic area or skilled as educators. Because presentations were made by different presenters and the presenters were standardized only in core elements, some groups may have had repetitions in materials or more effective verbal explanations than others. The decision to test knowledge gained using an identical pretest and posttest could have introduced bias, as participants likely had heightened interest in hearing information about questions initially posed.

Because of the short-term follow-up period, the study did not reassess provider knowledge after a period of time had lapsed to demonstrate that the information learned at the CME lecture was retained. The study lacks data on actual clinical practice as pretest and whether the practice of the providers changes after the CME program. Efforts to study long-term effects of The Heart Truth CME programs are underway.

Conclusions

Significant knowledge deficits are apparent in a cross-section of healthcare providers attending a CME lecture on women and heart disease. A 1-hour presentation was successful in improving knowledge and self-assessed skills and attitudes among primary care physicians, nurse practitioners/physician assistants, and registered nurses.

Acknowledgments

We thank the following organizations for their support and advice in development of the educational materials: the American Academy of Physician Assistants, the American College of Cardiology, the American College of Family Physicians, the American College of Obstetricians and Gynecologists, the American College of Physicians, the American Heart Association, the Association of Black Cardiologists, the Asian and Pacific Islander American Health Forum, the Association of Women's Health, Obstetric, and Neonatal Nurses, the National Hispanic Medical Association, the National Medical Association, and the Department of Health and Human Services Office on Women's Health Minority Women's Health Panel of Experts. The listing of these organizations does not imply their endorsement of this article. We further wish to acknowledge the work and support of the following individuals who contributed to the dissemination portion of The Heart Truth Professional Education Campaign: Eileen P. Newman, M.S., R.D., Ann M. Taubenheim, Ph.D., M.S.N, Gladys Ruiz, R. Ann Abercrombie, M.L.S., Kathy W. Beenen, M.B.A, Rachna Chaudhari, M.P.H., Mary Dortenzo, M.S.N., N.P.C, Jo Jean Elenes, Denise Flinn, Gina Gilliland, R.N.C., WHNP, Mandy Herleth, Dixie Horning, Anna L. Kindermann, J.D., Katherine Kolb, R.N., B.S.N, M.S., Jennifer Lennon, Terry Long, Melissa McNeil, M.D., Celisse Morris-Miller, M.S., Susan M. Nappi, Elizabeth O'Neill, B.A., Michele Ondeck, R.N., M.E.d., Abigail L. Reese, C.N.M., M.S.N., Tara Rizzo, M.P.H., Susan Samuels, R.N., B.S.N., Kimberly D. Sanders, Robert F. Schwartz, Stephanie Singer, Ying Studebaker, M.S., R.N, C.N.S., Elizabeth C. Thul, L.S.W., Margaret Trautz, Elisabeth von der Lohe, M.D., Esther Yasui, Erika Graves, and Robin Lam.

This work was supported by the Department of Health and Human Services Office on Women's Health (order number HHSP233200400916P). Additional support for manuscript preparation was provided by the Executive Advisory Board of the Iris Cantor-UCLA Women's Health Center.

The views expressed within are solely those of the authors and do not necessarily represent the views of the Department of Health and Human Services. The authors had full access to the data and take responsibility for its integrity. All authors have read and agree to the article as written.

All authors received funding support from the Department of Health and Human Services Office of Women's Health for

development or dissemination of the Heart Truth Professional Education Campaign. S.G. is the recipient of a DHHS grant, Advancing System Targets for HP2010 To Improve Cardiovascular Health of Women in Rural Illinois. C.R. and A.E.N. are recipients of a K-08 NHLBI Cultural Medical Education Curriculum grant. M.D. is the recipient of NIDA R01, Exercise and Nicotine Replacement for Female Smokers, receives honoraria as a member of the Massachusetts Public Health Council, and owns stock in Pfizer Inc.

Disclosure Statement

The authors have no conflicts of interest to report.

References

1. National Heart, Lung, and Blood Institute. The Heart Truth Campaign overview. Available at: www.nhlbi.nih.gov/health/hearttruth/press/nhlbi_04_campaign.htm Accessed August 1, 2007.
2. Chandra NC, Ziegelstein RC, Rogers WJ, et al. Observations of the treatment of women in the United States with myocardial infarction: A report from the National Registry of Myocardial Infarction-I. Arch Intern Med 1998;158:981–988.
3. Nohria A, Vaccarino V, Krumholz HM. Gender differences in coronary artery disease in women: Gender differences in mortality after myocardial infarction: Why women fare worse than men. Cardiol Clin 1998;16:45–57.
4. Scott LB, Allen JK. Providers' perceptions of factors affecting women's referral to outpatient cardiac rehabilitation programs: an exploratory study. J Cardiopulm Rehabil. 2004; 24:387–391.
5. O'Meara JG, Kardia SL, Armon JJ, Brown CA, Boerwinkle E, Turner ST. Ethnic and sex differences in the prevalence, treatment, and control of dyslipidemia among hypertensive adults in the GENOA study. Arch Intern Med 2004;164: 1313–1318.
6. Curtis LH, Al-Khatib SM, Shea AM, Hammill BG, Hernandez AF, Schulman KA. Sex differences in the use of implantable cardioverter-defibrillators for primary and secondary prevention of sudden cardiac death. JAMA 2007; 298:1517–1524.
7. Hernandez AF, Fonarow GC, Liang L, et al. Sex and racial differences in the use of implantable cardioverter-defibrillators among patients hospitalized with heart failure. JAMA 2007;298:1525–1532.
8. Hendrix KH, Riehle JE, Egan BM. Ethnic, gender, and age-related differences in treatment and control of dyslipidemia in hypertensive patients. Ethn Dis 2005;15:11–16.
9. Ayanian JZ, Landon BE, Landrum MB, Grana JR, McNeil BJ. Use of cholesterol-lowering therapy and related beliefs among middle-aged adults after myocardial infarction. J Gen Intern Med 2002:17:95–102.
10. Schulman KA, Berlin JA, Harless W, et al. The effect of race and sex on physicans' recommendations for cardiac catheterization. N Engl J Med 1999:340:618–626.
11. Mosca L, Appel LJ, Benjamin EJ, et al. Evidence-based guidelines for cardiovascular disease prevention in women. Circulation 2004;109:672–693.
12. Mosca L, Linfante AH, Benjamin EJ, et al. National study of physician awareness and adherence to cardiovascular disease prevention guidelines. Circulation 2005;111:499–510.
13. New York City passes trans fat ban. MSNBC.com. December 5, 2006. Available at www.msnbc.msn.com/id/16051436/ Accessed September 5, 2007.

Address correspondence to:
Janet Pregler, M.D.
Iris Cantor-UCLA Women's Health Center
911 Broxton Avenue 1st Floor
Los Angeles, CA, 90024

E-mail: japregler@mednet.ucla.edu

Appendix: Guidelines-Based Content for The Heart Truth CME Presentations

All CME presentations contained 31 identical core slides and an additional 15–20 slides that varied based on the knowledge and interests of the audience. This list reflects the information contained in the core slides.

Background

- Cardiovascular disease mortality trends for males and females
- Cardiovascular disease ranking for cause of death among women of different racial and ethnic groups
- Association between race/ethnicity and high risk for diabetes

Process

- Five-step approach to prevention: assessment of level of risk, lifestyle approaches for coronary artery disease prevention; treatment of hypertension, diabetes, lipid abnormalities; highest priority is for intervention in high-risk patients; avoid therapies that lack benefit or where risks outweigh benefits
- Risk stratification, including definition of metabolic syndrome

Interventions

- Smoking cessation and avoidance of environmental tobacco
- Encourage physical activity
- Weight maintenance/reduction goals
- Treatment of hypertension, including lifestyle approaches
- Optimal lipid levels; dietary and medication interventions to reach goals
- Recommended HbA1C target for diabetics
- Use of aspirin in high-risk, intermediate-risk, and lower-risk women
- Hormone therapy should not be given for primary or secondary prevention of heart disease
- Antioxidant supplements should not be given for primary or secondary prevention of heart disease

References

Mosca L, Appel LJ, Benjamin EJ, et al. Evidence-based guidelines for cardiovascular disease prevention in women. Circulation 2004; 109:672–695. Available at circ.ahajournals.org/cgi/content/short/109/5/672

The Heart Truth Professional Education Campaign. Available at www.nhlbi.gov/health/hearttruth/press/nhlbi_04_campaign.htm

AIDS Behav (2009) 13:310–317
DOI 10.1007/s10461-008-9383-3

ORIGINAL PAPER

Correlates of Any Condom Use Among Russian Narcology Patients Reporting Recent Unprotected Sex

Anita Raj · D. M. Cheng · E. M. Krupitsky · S. Coleman · C. Bridden · J. H. Samet

Published online: 10 April 2008

Abstract The purpose of this study was to assess whether HIV/sexually transmitted infection (STI) risk factors: risky sex (multiple sex partners and sex trade involvement), past HIV or STI diagnosis and substance use (at risk drinking and injection drug use) are associated with the outcome any condom use in the past 6 months among Russian narcology hospital patients. Participants ($N = 178$) included only those who reported unprotected sex in the past 6 months and were aged 18–55 years and 76% male. Any condom use in the past 6 months was reported by 55% of the sample. History of STIs was reported by 43% of participants; 15% were HIV-infected. Regression analyses adjusted for demographics demonstrated that those reporting multiple sex partners ($OR_{adj} = 4.2$, 95% CI = 2.0–8.7) and sex trade involvement ($OR_{adj} = 2.4$, 95% CI = 1.1–5.1) in the past 6 months had significantly higher odds of reporting any condom use in this same timeframe. HIV/STI and substance use were not associated with increased odds of condom use.

Keywords Condom use · HIV · Substance abuse

A. Raj (✉) · J. H. Samet
Department of Social and Behavioral Sciences, Boston University School of Public Health, 715 Albany Street Talbot Building, Boston, MA 02118, USA
e-mail: anitaraj@bu.edu

D. M. Cheng
Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

D. M. Cheng · C. Bridden · J. H. Samet
Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA

E. M. Krupitsky
St. Petersburg State Pavlov Medical University, St. Petersburg, Russian Federation

S. Coleman
Data Coordinating Center, Boston University School of Public Health, Boston, MA, USA

Introduction

HIV infection was rare in Eastern Europe in the mid-1990s, but its prevalence has been increasing without effective prevention efforts in the past decade. The Russian Federation (heretofore referred to as Russia) currently has an adult HIV prevalence rate of 1.1%; this country also has the largest number of HIV-infected individuals in all of Europe, 370,000 as of 2006 (UNAIDS 2007). At the heart of the Russian epidemic is the large number of young injection drug users (IDUs), primarily in urban centers (UNAIDS 2007). Approximately 2.5% of the adult population in Russia is an IDU (UNAIDS 2006b); 66% of Russians infected with HIV in 2005 and 2006 acquired the virus via injection drug use (UNAIDS 2006a; UNAIDS 2007). Notably, however, growing numbers of individuals in Russia are becoming infected via sexual transmission—from less than 10% in 2000 to more than 40% in 2005 (UNAIDS 2006a). Recent clinical and epidemiologic data now indicate that sexual transmission of HIV may be the most rapidly increasing of all HIV infection transmission risk behaviors in the region (Aral et al. 2005).

Sexual transmission of HIV infection within Russia is assumed to result from HIV-infected IDUs engaging in unprotected sex, bridging the epidemic to non-IDU populations. Studies with IDUs recruited from urban centers within Russia have found that the majority reports recent sexual activity, multiple partners and non-condom use,

particularly with steady sex partners (Rhodes et al. 2004; Somlai et al. 2002; Takacs et al. 2006). The epidemic has been propelled even further via the link between injection drug use and sex work, particularly for female IDUs. A substantial proportion of female IDUs (37%) report having engaged in sex work (Benotsch et al. 2004), and those engaging in sex work are more likely to report both risky injection drug use and a history of sexually transmitted infections (STIs), as compared with male IDUs or female IDUs reporting no history of sex work (Karapetyan et al. 2002; Platt et al. 2005). Nonetheless, awareness of HIV risk among IDUs is increasing (UNAIDS 2006b), and HIV risk perceptions among IDUs in Russia are actually greater than that seen in the United States (US) (Ksobiech et al. 2005).

Recent evidence from Russia raises the possibility that HIV may be spreading beyond IDUs and sex workers and reaching those with problem alcohol behavior (Krupitsky et al. 2004; UNAIDS 2006a). Increased risk for HIV among Russians with unhealthy alcohol use would affect a substantial proportion of the Russian population, as Russia has one of the highest per capita use of alcohol in the world (Nemtsov 2000; World Health Organization 2004). Research with a representative sample of Russian adolescents and adults found that one-third of sexually active individuals used alcohol prior to their last sex (Vannappagari 2004); additional research from Russia documents pervasive perceptions of at risk episodic drinking prior to sex and unprotected sex as social norms (World Health Organization 2005). Although Russian studies have not examined whether drinking risky amounts of alcohol increases likelihood of risky sexual practices, this may be the case given study findings from the US documenting that those reporting drinking risky amounts of alcohol are more likely to report a history of STI, sex with multiple partners, sex trade involvement and unprotected sex (Kalichman et al. 2007; Markos 2005; Raj et al. 2007; Rasch et al. 2000; Weinhardt and Carey 2000).

In summary, current research indicates that IDUs and risky drinkers in Russia are experiencing notable rates of HIV/STI and report notable HIV risks, including multiple partners, sex trade involvement and unprotected sex (Benotsch et al. 2004; Krupitsky et al. 2004; Platt et al. 2005; Rhodes et al. 2004; Somlai et al. 2002; Takacs et al. 2006; Karapetyan et al. 2002). Such research has not, however, examined associations between HIV/STI risk histories and condom use. Condom use in Russia generally is uncommon and primarily for pregnancy prevention when it does occur (Bobrova et al. 2005; Vannappagari 2004; World Health Organization 2005); thus, increased likelihood of condom use among riskier Russian substance users cannot be assumed, although this has been shown to be the case among US substance users (Bogart et al. 2005; Kwiatkowski et al. 1999; Shlay et al. 2004). The current study seeks to build upon the growing body of work in the area of substance use and sexual risk in Russia by assessing whether history of HIV/STI and risky sexual and substance use behaviors are associated with increased likelihood of any condom use among Russian narcology patients reporting recent unprotected sex. Any versus proportion of condom use was examined to provide insight into which patients ever and never use condoms, among this sample of patients who have engaged in recent unprotected sex. Findings from this work can be used to inform the growing sexual prevention and intervention efforts in Eastern Europe and other regions in which substance use plays a central role in the HIV epidemic (UNAIDS 2006b).

Methods

Study Design & Subject Recruitment

Data for this research came from the Russian PREVENT (Partnership to Reduce the Epidemic Via Engagement in Narcology Treatment) study, a randomized controlled trial (RCT) of an HIV behavioral intervention in narcology hospital in-patients in Russia. The PREVENT study included alcohol and/or drug-dependent men and women recruited from two narcology hospitals in the vicinity of St. Petersburg, Russia: (a) the Leningrad Regional Center for Addictions (LRCA) and (b) the Medical Narcology Rehabilitation Center (MNRC). Narcology hospitals are a standard treatment setting for drug and alcohol dependent persons in Russia and Eastern Europe. Typically, the hospitalization is 3–4 weeks in length and involves patients undergoing detoxification and then receiving addiction treatment.

Participants for this study were recruited from October 2004 to April 2005. Trained physician research associates approached all patients after initial detoxification (3–7 days after program entry) and assessed them for study eligibility. Eligible participants were 18 years or older, reported unprotected vaginal or anal sex in the past 6 months, and had a primary diagnosis of alcohol or drug dependence. Additional study entry criteria were the following: abstinence from alcohol and other abusive substances for 48 h; willing to undergo HIV testing as per standard narcology hospital protocol if not known to be HIV-infected; willing and able to provide contact information for themselves as well as a relative or close friend through whom they could be contacted; residing within 150 km of St. Petersburg; and possessing a home telephone. Individuals who were not fluent in Russian or demonstrated severe cognitive impairment as assessed by the research associate's clinical judgment at recruitment were excluded from the study.

Overall, 329 individuals were approached and screened for participation in this study. Of these, 129 were excluded due to not meeting eligibility criteria; 70 of these 129 ineligible participants reported no unprotected anal or vaginal sex in the last 6 months; breakdown of participants reporting abstinence versus those reporting consistent condom use were not available. Of the 200 eligible participants identified, 19 (9.5%) refused participation, yielding a final sample size of $N = 181$. All eligible and willing subjects provided written informed consent prior to study enrollment.

Procedure

Subsequent to recruitment and eligibility assessment, all participants provided written informed consent and received their baseline survey, which assessed demographics, HIV risk behaviors, substance use behaviors and other key health indicators. At baseline, while subjects were in the narcology hospitals, survey data were collected in two ways: (a) a face-to-face interview with a staff trained in survey administration and not providing care to the participant and (b) a computerized survey—Audio Computer-Assisted Self Interviewing (ACASI) system. ACASI removes the interviewer and, therefore, allows additional privacy, minimizes literacy issues, encourages truth telling, and provides an identical recording of each question; using this system has been shown to enhance the quality of self-report behavioral assessments, to maintain confidentiality, and to provide an acceptable method for collecting self-reports of HIV risk behavior (Newman et al. 2002). All interviews were conducted in Russian, and participants were compensated US$5 for the baseline assessment. The current analyses include data collected at the baseline assessment. Given the focus of the current study on condom use in the past 3–6 months, female participants reporting exclusively female sex partners ($n = 3$) were excluded from analyses, yielding a final sample size of 178 subjects.

Measures

Independent Variables

The six main independent variables for this study were risky sexual behaviors (multiple sex partners and sex trade involvement), substance use behaviors (at risk drinking and injection drug use), history of STI, and HIV serostatus.

Risky sexual behaviors were assessed via survey items from the Risk Assessment Battery (Navaline et al. 1994); the RAB was chosen based on its previously demonstrated validity with Russian narcology patients (Krupitsky 2005). A single item asked participants the number of sex partners in the past 6 months; participants were defined as having multiple sex partners if they reported two or more partners in this time frame. Two additional RAB items were used to assess buying sex with money or drugs (buying sex) and selling sex for money or drugs (selling sex) in the past 6 months. Sex trade involvement was defined as either buying or selling sex within the past 6 months.

Alcohol consumption in the past 30 days was collected using a Timeline Followback (TLFB) approach (Dillon et al. 2005; Midanik et al. 1998; Vinson et al. 2003; Weinhardt et al. 1998), in which participants noted the number of drinks they had in each of the past 30 days prior to hospitalization. At Risk Drinking was defined as having five or more drinks per day for men and 4 or more for women, in the past 30 days, based on an NIAAA definition of at risk drinking (NIAAA 2005). Two RAB items assessed recent injection drug use. The first item asked whether the participant engaged in injection drug use in the past 6 months; the second asked, for those reporting yes on item one, whether they had shared needles or works when injecting drugs in the past 6 months. IDU was defined as engaging in any injection drug use in the past 6 months. Additional data collected included substance use diagnosis based on the narcology hospital record; staff psychiatrists provided diagnoses at intake as part of clinical care. Diagnoses were based on ICD-X criteria (the standard Russian diagnostic manual).

STI diagnosis was collected via self-report. Participants were asked via survey whether they had ever been diagnosed with syphilis, gonorrhea, chlamydia, genital warts, genital herpes, other STIs (defined to exclude HIV) or pelvic inflammatory disease (women only); history of STI was defined as having any STI diagnosis ever. HIV serostatus was determined by HIV test results documented in the narcology hospital record. All narcology hospital patients are tested for HIV at program entry unless they are already known to be HIV-infected; patient HIV serostatus was then noted in the patient's medical record.

Outcomes

The primary outcome of interest was any condom use (yes vs. no), among all partners; this was assessed via a single RAB item on frequency of condom use in the past 6 months. Secondary outcomes were any condom use with main partners (past 3 months) and any condom use with casual partners (past 3 months). A series of questions examined condom use with main partners and casual partners: the number of times engaged in vaginal or anal sex and condom protected vaginal or anal sex, with each type of partner in the past 3 months. A main sex partner was defined as "the person you have sex with most often and regularly and/or the person with whom you feel most

attached," and casual partners were defined as "people you have sex with less frequently and with whom you do not consider yourself in a steady relationship."

Confounders

Demographic data collected included age, gender, marital status, sex of partners, education and employment, assessed via single survey items. The potential confounding factors included in regression analyses were: age, gender, and marital status (currently married vs. single, divorced, or widowed).

Data Analysis

Descriptive statistics were used to characterize the study subjects at baseline. Logistic regression analyses adjusting for potential confounders were conducted to assess associations between risky sexual behaviors (multiple sex partners and sex trade involvement), substance use behaviors (at risk drinking and injection drug use), and STI/HIV diagnoses with the outcome any condom use. Separate models were fit for each potential predictor. Independent variables that were significant at an alpha level of 0.05 were included together in a final multivariable model that also adjusted for potential confounding factors. Analyses of the secondary outcomes any condom use by main partner were conducted in the subset of subjects reporting a main partner and any condom use by casual sex partners were conducted only in the subset reporting a casual partner. The secondary outcomes were analyzed using the same approach as that described for the primary outcome. To minimize the potential for collinearity, we assessed correlation between pairs of independent variables and verified that no pair of variables included in the same regression model was highly correlated (i.e., $r > 0.40$). Analyses were performed using SAS software (version 9.1; SAS Institute, Cary, NC).

Results

Sample Characteristics

Participants were aged 18–55 years (mean age 33.2 years) and predominantly male (75.8%). Half (50.6%) were unemployed; 6.2% had less than a high school education (Table 1). One-third of the sample (33.1%) was married, and 75.3% had a main sex partner. Nearly everyone (96.1%) identified as heterosexual; involvement with same sex partners was reported by 7.9% of participants.

Condom Use, STI/HIV Diagnosis, and Risky Sex and Substance Use Behaviors

Condom use in the past 6 months was reported by 55.1% of the participants. Notably, although those screened into the study were required to have engaged in at least one unprotected sex episode in the past 6 months, 2.8% of survey participants reported no unprotected sex in the past 6 months during the baseline assessment. Among those reporting sex with main partners ($n = 125$), 33.6% reported condom use with these partners, with only 5.6% reporting that it was consistent condom use. Similarly, among those reporting sex with casual partners ($n = 107$), 50.5% reported condom use with these partners, with 13.1% reporting that it was consistent condom use.

Almost half the sample (42.9) reported a history of STI diagnosis, and 14.6% of the sample was HIV-infected. The majority of the sample (69.1%) reported multiple sex partners in the past 6 months, with 26.4% of the sample reporting 4 or more sex partners in this timeframe. (Table 1) More than a quarter of the sample (27.0%) reported sex trade involvement, with 11.2% reporting selling of sex and 18.5% reporting purchase of sex.

Almost three-quarters of the sample (71.9%) reported alcohol use in the past 30 days; 64.0% reported at risk drinking in this same period. Past 6 months injection drug use was reported by 39.5% of the sample, with 77.1% of these IDUs reporting needle or works sharing in this same timeframe. These findings are consistent with subjects' clinical diagnoses, which indicate that 60.1% of this sample is alcohol-dependent, 31.5% heroin dependent, and 8.4% both alcohol and heroin dependent.

Associations of HIV/STI, Risky Sex and Substance Use Behaviors with Condom Use

In models adjusted for age, gender, and marital status, participants reporting multiple sex partners ($OR_{adj} = 4.2$, 95% CI = 2.0–8.7) and sex trade involvement ($OR_{adj} = 2.4$, 95% CI = 1.1–5.1) had a higher odds of reporting any condom use in the past 6 months. (Table 2.) In the final regression model including relevant demographics and both multiple sex partners and sex trade involvement, only multiple sex partners remained significantly associated with condom use ($OR_{adj} = 3.6$, 95% CI = 1.7–7.9).

Secondary analyses conducted to assess variables associated with condom use by type of sex partner, main or casual revealed that among those with a main sex partner, having multiple sex partners was significantly associated with condom use with the main partner (OR_{adj} 2.5, 95% CI = 1.0–6.0, $P = .04$). (Table 3.) Among those with a casual partner, the effect of having multiple sex partners

Table 1 Demographics, HIV/STI history, risky sexual and substance use behaviors among narcology hospital patients in St. Petersburg, Russia ($N = 178$)

Characteristic	% For total sample	% Reporting any condom use by characteristic	Chi-square (df), *P*-value
Sex			
Male	75.8	54.1	.22 (1), .64
Female	24.2	58.1	
Employment status			
Employed	49.4	46.6	5.04 (1), .02
Unemployed	50.6	63.3	
Level of education			
<High school education	6.2	54.6	.001 (1), .97
≥High school education	93.8	55.1	
Marital status			
Married	33.1	49.2	1.24 (1), .26
Single/divorced/widowed	66.9	58.0	
Sexual orientation			
Heterosexual/straight	96.1	54.4	.79 (1), .37
Gay/lesbian/bisexual	3.9	71.4	
HIV/STI history			
STI history			
Yes	42.7	51.3	.75 (1), .39
No	57.3	57.8	
HIV-serostatus			
Positive	14.6	61.5	.52 (1), .47
Negative	85.4	54.0	
Risky sexual behavior			
Multiple sex partners			
Yes	69.1	65.0	16.04 (1), < .0001
No	30.9	32.7	
Sex trade			
Yes	27.0	72.9	8.47 (1), .004
No	73.0	48.5	
Substance use			
At risk alcohol use			
Yes	64.0	49.0	3.13 (1), .08
No	36.0	62.2	
Injection drug use			
Yes	39.5	64.3	3.73 (1), .05
No	60.5	49.5	

could not be evaluated as only 3 of 107 (2.8%) subjects reported no multiple sex partners.

Discussion

Despite very high HIV/STI risk in this sample-43% with an STI history and 15% HIV-infected, almost half of the participants from this study of in-treatment substance users reported no condom use. Such findings are not indicative of Russian narcology patients as a whole, as the sample excluded those sexually abstaining and those using condoms consistently in the past 6 months, but the findings do indicate that among those patients engaging unprotected sex, many are not using condoms at all. Findings from the current study support previous research from non-substance using Russian samples which indicate low condom use in this population (Bobrova et al. 2005; Vannappagari 2004; World Health Organization 2005), but do indicate greater condom use among substance users. Nonetheless, greater efforts remain needed to promote condom use in Russia, particularly among at risk populations such as those in substance use treatment.

Table 2 Adjusted[a] logistic regression analyses to assess the associations between risky sex and substance use behaviors and STI/HIV diagnosis with any condom use in the past 6 months among narcology hospital patients in St. Petersburg, Russia ($N = 178$)

	OR_{adj} (95% CI)
Multiple sex partners	4.2 (2.0–8.7)
Yes	
No	
Sex trade	2.4 (1.1–5.1)
Yes	
No	
STI history	0.6 (0.3–1.1)
Yes	
No	
HIV-infected	0.9 (0.4–2.2)
Yes	
No	
At risk alcohol use	0.7 (0.3–1.7)
Yes	
No	
IDU history	1.0 (0.5–2.3)
Yes	
No	

[a] Adjusted analyses control for age, gender and marital status

Importantly, those participants with a history of HIV/STI diagnosis were no more likely to report condom use. Such findings may be a consequence of low HIV risk perceptions related to sexual behaviors, as they correspond to other Russian research indicating low HIV risk perceptions among STI clinic patients (Benotsch et al. 2004). Those reporting recent injection drug use, which for the majority involved sharing of needles, also were no more likely to use condoms. This finding was particularly striking given high rates of HIV among IDUs. A recent study from St. Petersburg, the site of this study, indicates that 30% of IDUs in the city are HIV-infected (Shaboltas et al. 2006); further, cross-national research has found higher HIV risk perceptions with regard to injection drug use and yet greater use of needle sharing among Russian compared to US IDUs (Ksobiech et al. 2005). Overall, these findings suggest that HIV risk among Russian substance users may be affecting injection drug use but not sexual behaviors. Greater focus on sexual risk and risk reduction among IDUs is likely needed to address the HIV epidemic in Russia, given recent findings from St. Petersburg IDUs that document higher rates of HIV among those reporting a greater number of partners as well as among females reporting sex trade involvement (Kozlov et al. 2006).

Condom use was found to be linked with other sexual practices in this study; specifically, multiple sex partners in the past year and sex trade involvement. These findings are consistent with those seen with US substance using samples (Bogart et al. 2005; Kwiatkowski et al. 1999; Shlay et al. 2004). Given previous evidence of condom use in Russia primarily being used for pregnancy rather than STI prevention (Bobrova et al. 2005; Vannappagari 2004),

Table 3 Adjusted[a] logistic regression analyses to assess the associations between risky sex and substance use behaviors and STI/HIV diagnosis with main sex partner condom use ($n = 125$) and with casual sex partner condom use ($n = 107$) in the past 3 months, among narcology hospital patients in St. Petersburg, Russia

	Any condom use with main partner (%)	OR_{adj} (95% CI)	Any condom use with casual partner (%)	OR_{adj} (95% CI)
Multiple sex partners		2.5 (1.0–6.0)		N/A[b]
Yes	41.0		51.9	
No	21.3		0	
Sex trade		1.2 (0.5–3.1)		1.1 (0.5–2.6)
Yes	42.9		55.8	
No	30.9		46.9	
STI history		1.7 (0.8–3.9)		0.5 (0.2–1.1)
Yes	42.9		45.8	
No	26.1		54.2	
HIV-infected		1.0 (0.3–3.2)		0.4 (0.1–1.5)
Yes	42.9		46.7	
No	32.4		51.1	
At risk alcohol use		0.4 (0.1–1.2)		1.6 (0.5–4.6)
Yes	22.0		44.4	
No	43.9		59.1	
IDU history		1.3 (0.5–3.2)		0.7 (0.3–2.0)
Yes	43.1		57.5	
No	27.4		47.0	

[a] Adjusted analyses control for age, gender and marital status

[b] Analysis not conducted, as only 2.8% (3/107) of those reporting sex with casual sex partners in the past 3 months reported no multiple sex partners

motivation for use in these contexts is unclear. Regardless, exploratory analyses further indicate that the association between multiple sex partners and condom use holds true for main partners. Although, this association may also be the case for casual relationships, it could not be established in the current study due to too few participants reporting only one casual sex partner.

While the current findings contribute to our growing understanding of the HIV epidemic in Russia, this study has several limitations. The cross-sectional design reflecting a single point in time limits the ability to establish causality. Reliance on self-report for behavioral risk variables potentially results in social desirability and recall biases. Recall bias was minimized by using short timeframes for behavior assessment (i.e., past 1–6 months prior to hospitalization). Social desirability bias was mitigated by use of the ACASI technology to assess risk behavior; such a bias typically results in an underestimate rather than overestimate of risk behavior and STI prevalence (Newman et al. 2002).

Conclusion and Implications

Among substance dependent patients in Russia, condom use is low and does not appear to be associated with some of the important risk factors for HIV infection. The fact that condom use behaviors are affected in the context of steady relationships when the substance-using individual has multiple other sex partners, suggests that some sexual risk reduction education is reaching this population and could be further reinforced with proper intervention. Further, this study demonstrates that narcology treatment settings are an important venue in which HIV intervention could occur. Overall, this study demonstrates that sexual risk among alcohol and drug dependent patients in Russia specifically is an important HIV prevention issue meriting research and clinical attention.

Acknowledgements We would like to thank the interventionists and research associates at the Russian narcology hospitals for their contributions to the study. These individuals received compensation from the NIH grant for their contributions. Grant support for this study came from the National Institute on Alcohol Abuse and Alcoholism (NIAAA), NIH: R21-AA014821 (HIV Prevention Partnership in Russian Alcohol Treatment) and K24-AA015674 (Impact of Alcohol Use on HIV Infection—In US and Russia).

References

Aral, S. O., St. Lawrence, J. S., Dyatlov, R., & Kozlov, A. (2005). Commercial sex work, drug use, and sexually transmitted infections in St. Petersburg, Russia. *Social Science and Medicine, 60*, 2181–2190.

Benotsch, E. G., Somlai, A. M., Pinkerton, S. D., Kelly, J. A., Ostrovski, D., Gore-Felton C., & Kozlov A. P. (2004). Drug use and sexual risk behaviours among female Russian IDUs who exchange sex for money or drugs. *International Journal of STD and AIDS, 15*, 343–347.

Bobrova, N., Sergeev, O., Grechukhina, T., & Kapiga S. (2005). Social-cognitive predictors of consistent condom use among young people in Moscow. *Perspectives on Sexual and Reproductive Health, 37*, 174–178.

Bogart, L. M., Kral, A. H., Scott, A., Anderson, R., Flynn, N., Gilbert, M. L., & Bluthenthal, R. N. (2005). Sexual risk among injection drug users recruited from syringe exchange programs in California. *Sexually Transmitted Diseases, 32*, 27–34.

Dillon, F. R., Turner, C. W., Robbins, M. S., & Szapocznik, J. (2005). Concordance among biological, interview, and self-report measures of drug use among African American and Hispanic adolescents referred for drug abuse treatment. *Psychology of Addictive Behaviors, 19*, 404–413.

Kalichman, S. C., Simbayi, L. C., Jooste, S., & Cain, D. (2007). Frequency, quantity, and contextual use of alcohol among sexually transmitted infection clinic patients in Cape Town, South Africa. *American Journal of Drug and Alcohol Abuse, 33*, 687–698.

Karapetyan, A. F., Sokolovsky, Y. V., Araviyskaya, E. R., Zvartau, E. E., Ostrovsky, D. V., & Hagan, H. (2002). Syphilis among intravenous drug-using population: Epidemiological situation in St Petersburg, Russia. *International Journal of STD and AIDS, 13*, 618–623.

Koslov, A. P, Shaboltas, A. V., Toussova, O. V., Verevochkin, S. V., Masse, B. R., Purdue, T., Beauchamp, H., Sheldon, W., Miller, W. C., Heimer, R., Ryder, R. W., & Hoffman, I. F. (2006). HIV incidence and factors associated with HIV acquisition among injection drug users in St. Petersburg, Russia. *AIDS, 20*(6), 901–906.

Krupitsky, E. M. (2005). Risk assessment battery: Validation of the instrument in Russia. Uchenye Zapiski S.-Peterburgskogo Gosudarstvennogo Universiteta im. Pavlova *Proceedings of St. Petersburg Pavlov State Medical University, 10*(Suppl. 2), 46–55.

Krupitsky, E. M., Zvartau, E., Karandashova, G., Horton, N. J., Schoolwerth, K. R., Bryant K., & Samet, J. H. (2004). The onset of HIV infection in the Leningrad region of Russia: a focus on drug and alcohol dependence. *HIV Medicine, 5*, 30–33.

Ksobiech, K., Somlai, A. M., Kelly, J. A., Gore-Felton, C., Benotsch, E., McAuliffe, T., Hackl, K., Ostrovski, D., & Kozlov, A. P. (2005). Demographic characteristics, treatment history, drug risk behaviors, and condom use attitudes for U.S. and Russian injection drug users: the need for targeted sexual risk behavior interventions. *AIDS and Behavior, 9*, 111–120.

Kwiatkowski, C. F., Stober, D. R., Booth, R. E., & Zhang, Y. (1999). Predictors of increased condom use following HIV intervention with heterosexually active drug users. *Drug and Alcohol Dependence, 54*, 57–62.

Markos, A. R. (2005). Alcohol and sexual behaviour. *International Journal of STD and AIDS, 16*, 123–127.

Midanik, L. T., Hines, A. M., Barrett, D. C., Paul, J. P., Crosby, G. M., & Stall, R. D. (1998). Self-reports of alcohol use, drug use and sexual behavior: Expanding the timeline follow-back technique. *Journal of Studies in Alcohol, 59*, 681–689.

Navaline, H. A., Snider, E. C., Petro, C. J., Tobin, D., Metzger, D., Alterman, A. I., & Woody, G. E. (1994). Preparations for AIDS vaccine trials. An automated version of the Risk Assessment Battery (RAB): Enhancing the assessment of risk behaviors. *AIDS Research and Human Retroviruses, 10*(Suppl. 2), S281–S283.

Nemtsov, A. V. (2000). Estimates of total alcohol consumption in Russia, 1980–1994. *Drug and Alcohol Dependence, 58*, 133–42.

Newman, J. C., Des Jarlais D. C., Turner, C. F., Gribble, J., Cooley, P., & Paone, D. (2002). The differential effects of face-to-face and computer interview modes. *American Journal of Public Health, 92*, 294–297.

NIAAA. (2005). Helping patients who drink too much: A clinician's guide. Updated 2005 Edition: Retrieved December 13 2007, from http://pubs.niaaa.nih.gov/publications/Practitioner/CliniciansGuide2005/guide.pdf.

Platt, L., Rhodes, T., Lowndes, C. M., Madden, P., Sarang, A., Mikhailova, L., Renton, A., Pevzner, Y., Sullivan, K., & Khutorskoy, M. (2005). Impact of gender and sex work on sexual and injecting risk behaviors and their association with HIV positivity among injecting drug users in an HIV epidemic in Togliatti City, Russian Federation. *Sexually Transmitted Diseases, 32*, 605–612.

Raj, A., Saitz, R., Cheng, D. M., Winter, M., & Samet, J. H. (2007). Associations between alcohol, heroin, and cocaine use and high risk sexual behaviors among detoxification patients. *American Journal of Drug and Alcohol Abuse, 33*, 169–178.

Rasch, R. F., Weisen, C. A., MacDonald, B., Wechsberg, W. M., Perritt, R., & Dennis, M. L. (2000). Patterns of HIV risk and alcohol use among African-American crack abusers. *Drug and Alcohol Dependence, 58*, 259–266.

Rhodes, T., Judd, A., Mikhailova, L., Sarang, A., Khutorskoy, M., Platt, L., Lowndes, C. M., & Renton, A. (2004). Injecting equipment sharing among injecting drug users in Togliatti City, Russian Federation: Maximizing the protective effects of syringe distribution. *Journal of the Acquired Immune Deficiency Syndrome, 35*, 293–300.

Shaboltas, A. V., Toussova, O. V., Hoffman, I. F., Heimer, R., Verevochkin, S. V., Ryder, R. W., Khoshnood, K., Purdue, T., Masse, B. R., & Koslov, A. P. (2006). HIV prevalence, sociodemographic, and behavioral correlates and recruitment methods among injection drug users in St. Petersburg, Russia. *Journal of the Acquired Immune Deficiency Syndrome, 41*(5), 657–663.

Shlay, J. C., McClung, M. W., Patnaik, J. L., & Douglas, J. M., Jr. (2004). Comparison of sexually transmitted disease prevalence by reported condom use: Errors among consistent condom users seen at an urban sexually transmitted disease clinic. *Sexually Transmitted Diseases, 31*, 526–532.

Somlai, A. M., Kelly, J. A., Benotsch, E., Gore-Felton, C., Ostrovski, D., McAuliffe, T., & Kozlov, A. P. (2002). Characteristics and predictors of HIV risk behaviors among injection-drug-using men and women in St. Petersburg, Russia. *AIDS Education and Prevention, 14*, 295–305.

Takacs, J., Amirkhanian, Y. A., Kelly, J. A., Kirsanova, A. V., Khoursine, R. A., & Mocsonaki L. (2006). "Condoms are reliable but I am not": A qualitative analysis of AIDS-related beliefs and attitudes of young heterosexual adults in Budapest, Hungary, and St. Petersburg, Russia. *Central European Journal of Public Health, 14*, 59–66.

UNAIDS. (2006a). 2006 Report on the Global AIDS Epidemic: Retrieved December 13 2007, from http://www.unaids.org/en/HIV_data/2006GlobalReport/default.asp.

UNAIDS. (2006b). Country report of the Russian Federation on the Implications of the Declaration of Commitment on HIV/AIDS Adopted at the 26th Special Session of the United Nations General Assembly, June 2001. Reporting period January–December 2005: Retrieved December 13 2007, from http://data.unaids.org/pub/Report/2006/2006_country_progress_report_russianfederation_en.pdf.

UNAIDS. (2007). AIDS epidemic update: Retrieved December 13 2007, from http://data.unaids.org/pub/EPISlides/2007/2007_epiupdate_en.pdf.

Vannappagari, V. (2004). Monitoring sexual behavior in the Russian Federation: the Russia longitudinal monitoring survey 1992–2003. Report submitted to the US Agency for International Development: Retrieved August 31 2007, from http://www.cpc.unc.edu/projects/rlms/papers/sex_03.pdf.

Vinson, D. C., Reidinger, C., & Wilcosky, T. (2003). Factors affecting the validity of a timeline follow-back interview. *Journal of Studies in Alcohol, 64*, 733–740.

Weinhardt, L. S., & Carey, M. P. (2000). Does alcohol lead to sexual risk behavior? Findings from event-level research. *Annual Review of Sex Research, 11*, 125–157.

Weinhardt, L. S., Carey, M. P., Maisto, S. A., Carey, K. B., Cohen, M. M., & Wickramasinghe, S. M. (1998). Reliability of the timeline follow-back sexual behavior interview. *Annals of Behavioral Medicine, 20*, 25–30.

World Health Organization (WHO). (2004). Global Status Report on Alcohol 2004: Retrieved April 13 2006, from http://www.who.int/substance_abuse/publications/alcohol/en/.

World Health Organization (WHO). (2005). Alcohol use and sexual risk behaviors: A cross-cultural study of eight countries: Retrieved April 13 2006, from http://www.who.int/substance_abuse/publications/alcohol_sexual_risk_crosscultural.pdf.

Published in final edited form as:
Subst Abus. 2009 ; 30(3): 213–222. doi:10.1080/08897070903040923.

Binge Drinking and Unsafe Sex: A Study of Narcology Hospital Patients from St. Petersburg, Russia

A Raj, Ph.D[1], **DM Cheng, Sc.D**[2,3], **EM Krupitsky, M.D, Ph.D, DSci**[4], **S. Levenson, M.P.H.**[5], **VY Egorova, M.D.**[4], **S Meli, M.P.H.**[3], **EE Zvartau, M.D, Ph.D, DSci**[4], and **JH Samet, M.D., M.A., M.P.H.**[1,3]

[1]Department of Social Behavioral Sciences, Boston University of Public Health, Boston, MA

[2]Department of Biostatistics, Boston University School of Public Health, Boston, MA

[3]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center, Boston, MA

[4]St. Petersburg State Pavlov Medical University, St. Petersburg, Russian Federation

[5]Data Coordinating Center, Boston University School of Public Health, Boston, MA

Abstract

The purpose of this study was to assess the association between binge alcohol use and unprotected sex in Russian substance users. Participants (N=181) were narcology hospital patients assessed on demographics, alcohol use, risky sex, and STD/HIV diagnoses. Adjusted GEE logistic regression analysis examined the association between binge drinking and same day unprotected sex across each of the past 30 days, per participant (N=5430 observations). Participants were age 18–55 years, 75% male, and 64% binge drinking. Sex trade was reported by 27%; history of STDs by 43%; and HIV by 15%. One-fourth of daily observations included sex; 88% of these involved unprotected sex. Binge drinking was not associated with same day unprotected sex (OR_{adj}=1.0, 95% CI=0.7–1.4, χ^2 (1, N=5219)=0.01, ns). Findings document substantial HIV/STD risk and prevalence among Russian narcology patients, but no link between binge drinking and unprotected sex in this population, possibly due to very low rates of condom use generally.

Keywords

unprotected sex; condom use; alcohol

INTRODUCTION

HIV infection was rare in Eastern Europe in the mid-1990's, but its prevalence has been increasing in the past decade. AIDS deaths in this region doubled between 2003 and 2005 (1). The Russian Federation (heretofore referred to as Russia) currently has the largest number of HIV cases in all of Europe (1). At the heart of the Russian epidemic is the large number of young injection drug users (IDUs), primarily in urban centers (1). Data from early 2004 indicated that 80% of all officially reported cases in the country were transmitted by injection drug use (1). However, sexual transmission of HIV in Russia is rapidly increasing; 25% of HIV infections were attributable to sexual contact in 2004, compared to 6% in 2001 (1). Further,

Contact Information: Anita Raj, Ph. D, 715 Albany Street Talbot Building, Boston University School of Public Health, Boston, MA 02118, Phone: (617) 838-6467, Fax: (617) 638-4483, Email: anitaraj@bu.edu.

clinical and epidemiologic data indicate that sexual transmission of HIV infection in this region may be the most rapidly increasing of all HIV infection transmission risk behaviors (2).

Sexual transmission of HIV infection within Russia is assumed to result from HIV-infected IDUs engaging in unprotected sex, bridging the epidemic to non-IDU populations. Studies with IDUs recruited from urban centers within Russia have found that the majority reports recent sexual activity, multiple partners and non-condom use (3,4). The epidemic has been propelled even further via the link between injection drug use and sex work, particularly for female IDUs. A substantial proportion of female IDUs (37%) report having engaged in sex work (5) and those engaging in sex work are more likely to report both risky injection drug use and a history of sexually transmitted diseases, as compared with male IDUs or IDUs reporting no history of sex work (6,7).

Some evidence from Russia raises the possibility that HIV may be spreading beyond IDUs and sex workers and reaching those with alcohol dependence (8). Increased risk for HIV among Russians with alcohol problems such as binge alcohol use would affect a substantial proportion of the Russian population, as Russia has one of the highest per capita use of alcohol in the world (5,9,10). The issue of a possible association between unsafe sex and alcohol use is important as the Russian Longitudinal Monitoring Survey, which assesses behavior and health with a representative sample of Russians aged 14–49 years, found that one-third of sexually active participants engaged in alcohol use at last sexual episode (11). To better address how alcohol may be linked to sexual risk within this population, research is needed to assess the association between binge drinking and unprotected sex within the Russian context. Recent research has examined sexual risk among IDUs in Russia (3,7) but not between binge alcohol use and risky sexual behaviors. Nonetheless, studies from other countries suggest such a link likely exists and is a concern for an expansion of the HIV epidemic to Russians engaging in binge alcohol use. Studies in the United States and Western Europe indicate that binge drinkers are more likely to engage in riskier sexual activities (e.g., multiple sex partners) and less likely to use condoms generally (12,13). Further, there is some, albeit inconsistent, evidence that excessive use of alcohol when engaging in sex reduces the likelihood of protected sex in that episode (12,13).

In summary, research demonstrates that HIV is at epidemic proportions in Russia and that those with substance abuse problems, specifically IDUs and perhaps binge drinkers, are at disproportionate risk for acquiring the virus and spreading it to others. The Joint United Nations Programme on AIDS recommends the use of sexual risk reduction interventions targeting substance abusing populations as an important means to inhibit the ongoing epidemic in this region (1). Development of such interventions will require more understanding of how substance use behaviors are linked to unprotected sex within a Russian treatment population than that provided by the current literature. This study seeks to build upon the growing body of work in the area of substance use and sexual risk in Russia by assessing the association between binge alcohol use and unprotected sex among Russian narcology hospital patients diagnosed with alcohol and/or drug dependence.

METHODS

Study Design & Subject Recruitment

Data for this research came from the Russian PREVENT (**P**artnership to **R**educe the **E**pidemic **V**ia **E**ngagement in **N**arcology **T**reatment) study, a randomized controlled trial (RCT) of an HIV behavioral intervention in narcology hospital in-patients in Russia. The PREVENT study included alcohol and/or drug-dependent men and women recruited from 2 narcology hospitals in the vicinity of St. Petersburg Russia: a) the Leningrad Regional Center for Addictions (LRCA) and b) the Medical Narcology Rehabilitation Center (MNRC). Narcology hospitals

are a standard setting for drug and alcohol dependent persons in Russia and Eastern Europe to receive treatment. Typically, the hospitalization is 3 to 4 weeks in length and involves patients undergoing detoxification and then receiving addiction treatment.

Participants for this study were recruited from October 2004 to April 2005. Trained physician research associates approached all patients after initial detoxification and assessed them for study eligibility. Eligible participants were 18 years or older, reported unprotected vaginal or anal sex in the past 6 months, and had a primary diagnosis of alcohol or drug dependence. Additional study entry criteria were the following: abstinence from alcohol and other abusive substances for 48 hours; willingness to undergo HIV testing as per standard narcology hospital protocol if not known to be HIV-infected; willingness and able to provide contact information for themselves as well as a relative or close friend through whom they could be contacted; residing within 150 kilometers of St. Petersburg; and possessing a home telephone. Individuals who were not fluent in Russian or demonstrated severe cognitive impairment as assessed by the research associate's clinical judgment at recruitment were excluded from the study. All eligible subjects provided written informed consent prior to study enrollment.

Procedure

Immediately subsequent to recruitment and eligibility assessment, all participants provided written informed consent and received their baseline survey, which assessed demographics, HIV risk behaviors, substance use behaviors and other key health indicators. At baseline, while subjects were in the narcology hospitals, HIV risk behavior questions were administered by both a face to face research associate interview as well as through an Audio Computer-Assisted Self Interviewing (ACASI) system. ACASI removes the interviewer and, therefore, allows additional privacy, minimizes literacy issues, encourages truth telling, and provides an identical recording of each question; using this system has been shown to enhance the quality of self-report behavioral assessments, to maintain confidentiality, and to provide an acceptable method for collecting self-reports of HIV risk behavior (14). All interviews were conducted in Russian, and participants were compensated US$5 for the baseline assessment. The current analyses include data collected at the baseline assessment.

Measures

Demographics including age, gender, marital status, education and employment were assessed via single survey items.

Main Independent Variable—Our main independent variable, daily binge alcohol use, was collected using a Timeline Followback (TLFB) approach (15,18). Participants noted the number of drinks they had in each of the past 30 days prior to hospitalization; daily binge alcohol use was defined as having 5 or more drinks per day for men, 4 or more for women. Additional data collected included substance use condition, based on the diagnosis received at the narcology hospital, and injection drug use and risky injection drug user ("sharing needles or works") in the past 6 months, assessed via single items from the Risk Assessment Battery (RAB) (19).

Outcomes—Our primary outcome variable, unprotected sex, was based on a TLFB assessment in which participants were asked the number of times they had vaginal or anal sex and if a condom was used in each of the past 30 days prior to hospitalization. Unprotected sex was modeled as a dichotomous outcome where subjects were categorized as either having an unprotected sex episode (any anal or vaginal sex without a condom on a given day) or having no unprotected sex (using a condom during all anal or vaginal sex on a given day or having no anal or vaginal sex) for each of the past 30 days. For descriptive purposes, we also assessed number of times unprotected vaginal or anal sex occurred with all primary and all casual

partners in the past 3 months via separate survey items. A primary sex partner was defined as "the person you have sex with most often and regularly and/or the person with whom you feel most attached," and casual partners were defined as "people you have sex with less frequently and with whom you do not consider yourself in a steady relationship."

Covariates—HIV behavioral risk factors (i.e., multiple sex partners, sex trade involvement, and recent risky injection drug use) were covariates in analyses and assessed via survey items from the Risk Assessment Battery (RAB). A single item asked participants the number of sex partners in the past 6 months; multiple sex partners was defined as 4 or more partners due to the large proportion of the sample reporting this behavior. Two additional RAB items with dichotomous responses were used to assess buying sex with money or drugs (buying sex) and selling sex for money or drugs (selling sex) in the past 6 months. Two dichotomous RAB items also assessed recent injection drug use and recent risky injection drug use. The first item asked whether the participant engaged in injection drug use in the past 6 months; the second asked, for those reporting yes on item one, whether they had shared needles or works when injecting drugs in the past 6 months.

STD/HIV diagnoses were included as covariates in analyses. Self-reported STD was assessed by asking whether participants had ever been diagnosed with syphilis, gonorrhea, chlamydia, genital warts, genital herpes, other STDs (defined to exclude HIV) or pelvic inflammatory disease (women only); no positive response yielded a "no STD diagnosis" dichotomous response. HIV serostatus was determined by HIV test results.

Data Analyses

Descriptive statistics were used to assess participants' characteristics for the sample (N=181). We also assessed bivariate associations between any binge drinking in the past 30 days with demographics and HIV risk factors using Chi-square and t-tests as appropriate.

For the primary hypothesis, we assessed the association between binge drinking and unprotected sex across each of the past 30 days per participant. Thus each of the 181 subjects could contribute a maximum of 30 observations to the analyses, resulting in N=5430 observations for the repeated measures analyses. We used generalized estimating equations (GEE) logistic regression models to examine the association between binge drinking and unprotected sex on the same day for each observation, adjusting for potential confounding factors: demographics (age, marital status, gender), HIV serostatus, and HIV risk factors (multiple partners, same sex partners, sex trade involvement, no primary partner, injection drug use and STD history). The GEE approach was used to adjust for the correlation due to analyzing repeated measures from the same subject (20).The empirical standard errors from the GEE approach were used for all analyses. The primary analyses utilized all available observations (N=5430). Secondary analyses were also conducted that excluded observations in which sex was not reported (N=1535). All analyses were conducted using two-sided tests and a significance level of 0.05. Note: Data presented in the results include and indicate whether they come from unique subjects (N=181) or if they include repeated observations from the same subject (N=5430).

RESULTS

Subject Characteristics

Study participants (N=181) were age 18–55 years (mean age 33.2 years), predominantly male (75%) and unmarried (67%). Although almost all had graduated from high school (94%), only half were employed. Nearly everyone (99%) identified as heterosexual; however 9% reported

having been with at least 1 same sex partner in the past 6 months, with women being more likely to report a same sex partner than men.

Substance Abuse Behaviors and Diagnosis, Unprotected Sex, and HIV Risk Factors

Almost three-quarters of our sample (72%) reported alcohol use in the past 30 days; 64% reported binge drinking in this same period. The 181 study subjects contributed a total of 5430 observations from the TLFB. Among the 5400 observations, 37% reported alcohol use (n=2020); 83% of these involved binge drinking (n=1669/2020). Recent injection drug use was reported by 40% of the sample. These findings are consistent with subjects' clinical diagnoses, which indicate that 60% of this sample is alcohol-dependent, 32% heroin dependent, and 8% both alcohol and heroin dependent. (Table I)

Three-fourths of participants reported having a primary sex partner, and 67% reported having a casual sex partner. Among those reporting recent involvement with a primary partner (n=136), 88% engaged in unprotected vaginal sex and 17% in unprotected anal sex with this type of partner. Among those reporting recent involvement with a casual partner (n=121), 76% engaged in unprotected vaginal sex and 8% in unprotected anal sex with this partner. (Table I) TLFB data demonstrated that, among the 5430 observations, 28% included at least one episode of either vaginal or anal sex (n=1535); 88% of these involved unprotected sex (n=1345/1535). Sex with 2 or more partners in the past 6 months was reported by 70% of the sample; 26% had been with 4 or more sex partners in the past 6 months. More than 1 in 4 participants (27%) reported sex trade involvement--19% had bought sex and 12% had sold sex. One third of participants (31%) had engaged in recent risky IDU; notably, this is the majority of those engaging in recent IDU (78%, 55/72). Consistent with these risks, high rates of HIV and STD were observed in this sample; almost half (43%) had a history of other STDs, and a substantial minority (15%) was HIV-infected. (Table I)

Bivariate Associations with Binge Drinking

Participants reporting binge alcohol use in the past 30 days were significantly more likely to be older, male, and employed, compared with non-binge drinkers, and they were also significantly less likely to be engaging in injection drug use, have a history of STDs, and be HIV-infected. (Table II)

Associations Between Binge Drinking and Same Day Unprotected Sex Across Each of the Past 30 Days, Per Participant

Repeated measures analyses of daily observations indicated no association between binge drinking and same day unprotected sex in either unadjusted (OR=1.0, 95% CI=0.8–1.3, χ^2 (1, N=5219)=0.01, ns) or adjusted analyses controlling for demographics and HIV risk factors (OR_{adj} =1.0, 95% CI=0.7–1.4, χ^2 (1, N=5219)=0.01, ns). (Table III) In analyses restricted to observations where sex was reported, the association between binge drinking and unprotected sex remained non-significant (OR_{adj} =1.0, 95% CI=0.5–1.7, χ^2 (1, N=1451)=0.02, ns). (Table IV)

Notably, the multivariable model of daily observation did indicate associations between other factors and unprotected sex. (Table III) The odds of an unprotected sex observation were significantly higher among those having a primary partner (OR_{adj}=2.9, 95% CI=1.7–4.7, χ^2 (1, N=5219)=15.65, <0.001) and those having 4 or more sex partners in the past 6 months (OR_{adj}=2.2, 95% CI=1.4–3.2, χ^2 (1, 5219)=9.89, <0.01); the odds of an unprotected sex observation were significantly lower among those who had purchased sex in the past 6 months (OR_{adj}=0.5, 95% CI=0.4–0.8, χ^2 (1, 5219)=7.84, <0.01). (Table III) In the model restricted to observations where sex was reported, the effects of having a primary partner, having 4 or more sex partners, and purchasing sex were attenuated and no longer statistically significant. (Table

IV) Neither of the multivariable observation models detected associations between HIV/STD infection, selling sex, or recent risky IDU and unprotected sex.

DISCUSSION

Despite reported associations between alcohol use and unsafe sex in the medical literature (21,23) in this Russian cohort of narcology hospital patients with unsafe sex in the past 6-months, binge drinking was not associated with an increased odds of same day unprotected sex. These findings are similar to other event specific studies of substance use and unprotected sex in US adolescents (24,26). The absence of an association in this study may reflect a reality that binge alcohol use does not impact unsafe sex in this population, however, another possibility must be considered: the impact of alcohol may be difficult to detect when 88% of daily observations in which sex occurs do not include condoms.

Currently, condoms in Russia are primarily used for pregnancy prevention and not, as yet, for prevention of sexually transmitted infections (11). However, as condom use becomes more normative in the Russian population, an impact of binge alcohol use may become evident. These data suggest that binge alcohol use, in a setting in which condom use is the exception, rather than the rule, is not a major issue for the promotion of the use of condoms. Nonetheless, as the HIV epidemic in Russia gains greater recognition in general, condom use will increase and the uptake of condoms may occur differentially between those with risky alcohol use and without such alcohol use behavior. Of note is the fact that for this sample, alcohol use predominantly meant binge alcohol use, with 83% of alcohol use incidents involving binge drinking levels.

In addition to binge drinking, no significant association was observed between the following variables and an observation of unprotected sex: HIV seropositive status, recent risky IDU, and history of STD diagnosis. However, we did observe higher odds of an unprotected sex observation among those reporting 4 or more sex partners and those who did not purchase sex in the past 6 months, as well as those with a primary partner. Notably, these significant associations with unprotected sex were lost when the model was limited to observations in which sex occurred, suggesting that these variables may be associated with having sex, rather than with an unprotected sex episode. Overall these findings demonstrate substantial HIV risk and pervasive unprotected sex among Russian narcology patients, with unprotected sex being no less likely among those with greater HIV risk.

These risk-specific findings are substantially different from those seen in the United States. Substance using and clinical samples from the United States demonstrate greater condom use among those with multiple sex partners, recent risky injection drug use and sex trade involvement (27,29). However, unprotected sex being more common among those with primary partners is consistent with US research (27,29) as well as with that seen in the general Russian population (11). Lack of significant findings in the model only including observations in which sex occurred may again be indicative of condom use simply not being normative in any Russia populations and more typically being used as a means of pregnancy prevention (11). These findings clearly speak to the need for better condom promotion in Russia generally and particularly among those in drug and alcohol treatment.

Risky sexual behaviors were remarkable for the cohort overall, especially compared with the general Russian population. Whereas 13% of sexually active Russians from a nationally representative sample reported 2 or more partners in the past year, 26% in this cohort reported 4 or more partners in the past year. Another dimension of risky sexual behavior is the high proportion involved in sex trade, an activity reported by 27% of this sample. While sex trade involvement has been discussed among Russian IDUs (5,7), it has not received as much focus

among binge drinkers in treatment; our study demonstrates that sex trade involvement is as likely for binge drinkers as IDUs. Future research to understand and address HIV risk among Russians in substance abuse treatment should consider multiple sex partners and sex trade involvement by alcohol dependent patients as well as IDUs in order to better address the epidemic in this clinical population.

While the current findings contribute to our growing understanding of the HIV epidemic in Russia, there are several limitations of the study. The study was observational at a single point in time which limits our ability to establish causality; however, the associations found do inform us regarding levels of HIV risk in sexually active Russian narcology hospital patients. Assessments were limited as they did not provide information on relative timing of same day binge alcohol use and unprotected sex; information on which occurred first and whether they occurred within the same relative period of the day was not collected. Additionally, reliance on self-report for behavioral risk variables potentially results in social desirability and recall biases on these, although, recall biases are likely minimal as timeframes for behavior assessment were short (i.e., past 1 to 6 months prior to hospitalization). We attempted to mitigate social desirability bias by use of the ACASI technology to assess risk behavior; such a bias typically results in an underestimate rather than overestimate of risk behavior and STD prevalence (14). Use of a single city in Russia and 2 narcology treatment sites for recruitment may also limit generalizablity of findings to other narcology hospital settings.

CONCLUSION and IMPLICATIONS

Sexual risk among alcohol and drug dependent patients is an important public health issue meriting research and clinical attention. While the hypothesized association between binge drinking and same day unprotected sex was not observed, findings did demonstrate substantial HIV risk for binge drinkers in treatment, as well as non-binge drinkers, a group that was predominantly IDUs. The exceedingly low condom use among a population at substantial risk for acquiring and transmitting HIV is alarming. The finding that episodes of unprotected sex are more common among some of those at greater risk for HIV infection, specifically those with a greater number of sex partners is of concern. These findings demonstrate that narcology treatment settings are an important venue in which HIV intervention could occur, and such intervention should not only promote condom use but guide patients to understand how their sexual behaviors increase risk for both acquiring and transmitting HIV. Efforts to address the HIV epidemic in Russia should address sex risk behaviors in the alcohol abusing population as well as among IDUs.

ACKNOWLDEGEMENTS

This work was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA), NIH: R21-AA014821 and K24-AA015674 (Dr. Samet).

Reference List

1. UNAIDS/WHO: AIDS Epidemic Update. 2005 [August 31, 2007]. from http://www.unaids.org/epi/2005/doc/report_pdf.asp
2. Aral SO, St Lawrence JS, Dyatlov R, Kozlov A. Commercial sex work, drug use, and sexually transmitted infections in St. Petersburg, Russia. Soc Sci Med 2005;60(10):2181–2190.
3. Rhodes T, Judd A, Mikhailova L, Sarang A, Khutorskoy M, Platt L, Lowndes CM, Renton A. Injecting equipment sharing among injecting drug users in Togliatti City, Russian Federation: maximizing the protective effects of syringe distribution. J Acquir Immune Defic Syndr 2004;35(3):293–300. [PubMed: 15076245]

4. Somlai AM, Kelly JA, Benotsch E, Gore-Felton C, Ostrovski D, McAuliffe T, Kozlov AP. Characteristics and predictors of HIV risk behaviors among injection-drug-using men and women in St. Petersburg, Russia AIDS Educ Prev 2002;14(4):295–305.
5. Benotsch EG, Somlai AM, Pinkerton SD, Kelly JA, Ostrovski D, Gore-Felton C, Kozlov AP. Drug use and sexual risk behaviours among female Russian IDUs who exchange sex for money or drugs. Int J STD AIDS 2004;15(5):343–347. [PubMed: 15117506]
6. Karapetyan AF, Sokolovsky YV, Araviyskaya ER, Zvartau EE, Ostrovsky DV, Hagan H. Syphilis among intravenous drug-using population: epidemiological situation in St Petersburg, Russia. Int J STD AIDS 2002;13(9):618–623. [PubMed: 12230926]
7. Platt L, Rhodes T, Lowndes CM, Madden P, Sarang A, Mikhailova L, Renton A, Pevzner Y, Sullivan K, Khutorskoy M. Impact of gender and sex work on sexual and injecting risk behaviors and their association with HIV positivity among injecting drug users in an HIV epidemic in Togliatti City, Russian Federation. Sex Transm Dis 2005;32(10):605–612. [PubMed: 16205301]
8. Krupitsky E, Zvartau E, Karandashova G, Horton NJ, Schoolwerth KR, Bryant K, Samet JH. The onset of HIV infection in the Leningrad region of Russia: a focus on drug and alcohol dependence. HIV Med 2004;5(1):30–33. [PubMed: 14731167]
9. Nemtsov AV. Estimates of total alcohol consumption in Russia, 1980–1994. Drug Alcohol Depend 2000;58(1–2):133–142. [PubMed: 10669064]
10. World Health Organization. Global Status Report on Alcohol 2004. 2004 [Retrieved on April 13, 2006]. from http://www.who.int/substance_abuse/publications/alcohol/en/
11. Vannappagari, V. Monitoring sexual behavior in the Russian Federation: the Russia longitudinal monitoring survey 1992–2003. Report submitted to the US Agency for International Development. 2004 [Retrieve on August 31, 2007]. from http://www.cpc.unc.edu/projects/rlms/papers/sex_03.pdf
12. Markos AR. Alcohol and sexual behaviour. Int J STD AIDS 2005;16(2):123–127. [PubMed: 15807939]
13. Weinhardt LS, Carey MP. Does alcohol lead to sexual risk behavior? Findings from event-level research. Annu Rev Sex Res 2000;11:125–157. [PubMed: 11351830]
14. Newman JC, Des J, Turner CF, Gribble J, Cooley P, Paone D. The differential effects of face-to-face and computer interview modes. Am J Public Health 2002;92(2):294–297. [PubMed: 11818309]
15. Dillon FR, Turner CW, Robbins MS, Szapocznik J. Concordance among biological, interview, and self-report measures of drug use among African American and Hispanic adolescents referred for drug abuse treatment. Psychol Addict Behav 2005;19(4):404–413. [PubMed: 16366812]
16. Midanik LT, Hines AM, Barrett DC, Paul JP, Crosby GM, Stall RD. Self-reports of alcohol use, drug use and sexual behavior: expanding the Timeline Follow-back technique. J Stud Alcohol 1998;59 (6):681–689. [PubMed: 9811089]
17. Vinson DC, Reidinger C, Wilcosky T. Factors affecting the validity of a Timeline Follow-Back interview. J Stud Alcohol 2003;64(5):733–740. [PubMed: 14572197]
18. Weinhardt LS, Carey MP, Maisto SA, Carey KB, Cohen MM, Wickramasinghe SM. Reliability of the timeline follow-back sexual behavior interview. Ann Behav Med 1998;20(1):25–30. [PubMed: 9755348]
19. Navaline HA, Snider EC, Petro CJ, Tobin D, Metzger D, Alterman AI, Woody GE. Preparations for AIDS vaccine trials. An automated version of the Risk Assessment Battery (RAB): enhancing the assessment of risk behaviors. AIDS Res Hum Retroviruses 1994;10:S281–S283. [PubMed: 7865319]
20. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.
21. Bagnall G, Plant M, Warwick W. Alcohol, drugs and AIDS-related risks: results from a prospective study. AIDS Care 1990;2(4):309–317. [PubMed: 2090201]
22. Halpern-Felsher BL, Millstein SG, Ellen JM. Relationship of alcohol use and risky sexual behavior: a review and analysis of findings. J Adolesc Health 1996;19(5):331–336. [PubMed: 8934293]
23. Leigh BC, Temple MT, Trocki KF. The relationship of alcohol use to sexual activity in a U.S. national sample. Soc Sci Med 1994;39(11):1527–1535. [PubMed: 7817217]
24. Bailey SL, Gao W, Clark DB. Diary study of substance use and unsafe sex among adolescents with substance use disorders. J Adolesc Health 2006;38(3):297–320. [PubMed: 16488830]

25. Fortenberry JD, Orr DP, Katz BP, Brizendine EJ, Blythe MJ. Sex under the influence. A diary self-report study of substance use and sexual behavior among adolescent women. Sex Transm Dis 1997;24 (6):313–319. [PubMed: 9243736]

26. Morrison DM, Gillmore MR, Hoppe MJ, Gaylord J, Leigh BC, Rainey D. Adolescent drinking and sex: findings from a daily diary study. Perspect Sex Reprod Health 2003;35(4):162–168. [PubMed: 12941648]

27. Bogart LM, Kral AH, Scott A, Anderson R, Flynn N, Gilbert ML, Bluthenthal RN. Sexual risk among injection drug users recruited from syringe exchange programs in California. Sex Transm Dis 2005;32 (1):27–34. [PubMed: 15614118]

28. Kwiatkowski CF, Stober DR, Booth RE, Zhang Y. Predictors of increased condom use following HIV intervention with heterosexually active drug users. Drug Alcohol Depend 1999;54(1):57–62. [PubMed: 10101617]

29. Shlay JC, McClung MW, Patnaik JL, Douglas JM Jr. Comparison of sexually transmitted disease prevalence by reported condom use: errors among consistent condom users seen at an urban sexually transmitted disease clinic. Sex Transm Dis 2004;31(9):526–532. [PubMed: 15480113]

Table I

Substance use behaviors and diagnosis, unprotected sex behaviors and HIV risk factors at baseline for narcology hospital patients enrolled in a sex risk reduction RCT in St. Petersburg, Russia (N=181)

	Total Sample % (n)
SUBSTANCE USE BEHAVIORS AND DIAGNOSIS	
Any Alcohol Use, Past 30 Days	72% (130)
Binge Drinking, Past 30 Days	64% (116)
IDU, Past 6 Months	40% (72)
Substance Abuse Diagnosis	
Alcohol-Dependent	60% (108)
Heroin-Dependent	32% (58)
BothAlcohol and Heroin-Dependent	8% (15)
UNPROTECTED SEX	
Any Unprotected Sex - Primary Sex Partner, past 3 months[a]	
Unprotected Vaginal Sex	88% (119)
Unprotected Anal Sex	17% (23)
Any Unprotected Sex - Casual Sex Partner, past 3 months[b]	
Unprotected Vaginal Sex	76% (92)
Unprotected Anal Sex	8% (10)
HIV RISK PROFILE	
Two or More Sex Partners	70% (126)
Four or More Sex Partners	26% (47)
Any Sex Trade	27% (49)
Buy Sex	19% (34)
Sell Sex	12% (21)
Recent IDU	40% (72)
Recent Risky IDU	31% (55)
STD History	43% (77)
HIV-infected	15% (27)

[a]Of our 181 participants, 136 (75.6%) reported a primary partner and were then asked questions about sex with a primary partner in the past 3 months.

[b] Of our 181 participants, 121(66.9%) reported a casual partner and were then asked questions about sex with a secondary partner in the past 3 months.

Table II

Baseline demographics and HIV risk factors for binge drinkers (n=116) and non-binge drinkers (n=65) enrolled in a sex risk reduction RCT in St. Petersburg, Russia (N=181) (stratified by binge drinking)

	Binge Drinkers[a] (n=116) %(n)	Non-Binge Drinkers (n=65) % (n)	Test statistic, degrees of freedom, p-value
DEMOGRAPHICS			
Age Range and Mean[b]	22–55; 36.0 (9.0)	18–55; 28.1 (7.2)	$t(181)= -6.02, 1, <.001$**
Gender			
Male	83% (96)	60% (39)	11.38, 1, <.001**
Female	17% (20)	40% (26)	
Full-time Employed	65% (75)	22% (14)	30.99, 1, <.001**
High School Graduate	94%(108)	94% (61)	0.00, 1, .97
Married	36% (42)	28% (18)	1.36, 1, .24
Primary Partner	71% (82)	83% (54)	3.12, 1, .08
Same Sex Partners	9% (10)	11% (7)	0.23, 1, .63
HIV RISK PROFILE			
Two or More Sex Partners	72% (83)	66% (43)	0.57, 1, .45
Four or More Sex Partners	28% (33)	22% (14)	1.03, 1, .31
Any Sex Trade	24% (28)	32% (21)	1.41, 1, .24
Buy Sex	17% (20)	22% (14)	0.50, 1, .48
Sell Sex	9% (10)	17% (11)	2.80, 1, .09
Injection Drug Use (IDU)	13% (15)	89% (57)	99.61, 1, <.001**
Risky IDU[c]	10% (11)	69% (44)	67.67, 1, <.001**
STD History	31% (36)	63% (41)	17.50, 1, <.001**
HIV-infected	8% (9)	28% (18)	13.04, 1, <.001**

[a] Any binge drinking in the past 30 days.

[b] Range and Mean (Standard Deviation) are provided for this continuous variable.

[c] Sample size used for this variable is N=72 injection drug users.

* p <.05.

** p < .01.

Table III

Adjusted odds ratios for unprotected sex from multivariable logistic regression analyses (N=5430).

Independent Variable	Adjusted Odds Ratio (95% Confidence Interval)	Score χ^2, degrees of freedom, p-value
Binge Drinking	1.01 (0.74–1.38)	0.01, 1, 0.94
Married	1.19 (0.72–1.97)	0.46, 1, 0.50
Female	1.18 (0.70–1.99)	0.37, 1, 0.54
Age[a]	0.83 (0.67–1.03)	2.84, 1, 0.09
Employed	0.76 (0.48–1.89)	1.40, 1, 0.24
Same Sex Partners	1.89 (0.90–3.99)	1.99, 1, 0.16
Primary Partner	2.85 (1.73–4.72)**	15.65, 1, <0.0001
Buy Sex	0.54 (0.36–.80)**	7.84, 1, 0.005
Sell Sex	0.70 (0.35–1.41)	1.01, 1, 0.31
Multiple Sex Partners (4+)	2.15 (1.42–3.24)**	9.89, 1, 0.002
Recent Risky IDU	1.46 (0.85–2.50)	1.92, 1, 0.16
STD Ever	0.73 (0.49–1.09)	2.34, 1, 0.13
HIV-infected	0.64 (0.31–1.35)	1.43, 1, 0.23

Note. N represents the total number of observations. Generalized estimating equations (GEE) were used to fit the logistic regression models

[a] Odds ratio corresponds to a 1 standard deviation (9.2 year) increase in age.

* p<.05.

** p<.01.

Table IV

Adjusted odds ratios for unprotected sex from multivariable logistic regression analyses including only observations in which sex occurred (N=1535).

Independent Variable	Adjusted Odds Ratio (95% Confidence Interval)	Score χ^2, degrees of freedom, p-value
Binge Drinking	0.96 (0.54–1.71)	0.02, 1, 0.90
Married	0.95 (0.36–2.52)	0.01, 1, 0.92
Female	1.02 (0.34–3.07)	0.00, 1, 0.97
Age[a]	1.38 (0.86–2.22)	1.78, 1, 0.18
Employed	1.68 (0.72–3.93)	1.46, 1, 0.23
Same Sex Partners	1.16 (0.33–4.11)	0.05, 1, 0.82
Primary Partner	2.01 (0.89–4.53)	2.63, 1, 0.01
Buy Sex	0.76 (0.32–1.80)	0.36, 1, 0.55
Sell Sex	0.74 (0.24–2.24)	0.29, 1, 0.59
Multiple Sex Partners (4+)	0.96 (0.49–1.88)	0.01, 1, 0.91
Recent Risky IDU	0.72 (0.24–2.12)	0.33, 1, 0.56
STD Ever	0.95 (0.43–2.11)	0.01, 1, 0.91
HIV-infected	0.83 (0.83–3.84)	0.06, 1, 0.81

Note. N represents the total number of observations. Generalized estimating equations (GEE) were used to fit the logistic regression models

[a] Odds ratio corresponds to a 1 standard deviation (9.2 year) increase in age.

* p<.05.

** p<.0.

Drug and Alcohol Dependence 101 (2009) 101–106

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

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

The associations of binge alcohol use with HIV/STI risk and diagnosis among heterosexual African American men

Anita Raj[a,*], Elizabeth Reed[b], M. Christina Santana[c], Alexander Y. Walley[d], Seth L. Welles[e], C. Robert Horsburgh[f], Stephen A. Flores[g], Jay G. Silverman[h]

[a] Social Behavioral Sciences Department, Boston University School of Public Health; 715 Albany Street, T2W, Boston, MA 02118, USA
[b] Center for Interdisciplinary Research on AIDS, 135 College Street, Suite 200, New Haven, CT 06510, USA
[c] Women's Center for Excellence (COE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center, 91 East Concord Street, Boston, MA 02118, USA
[d] Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center, 91 East Concord Street, Boston, MA 02118, USA
[e] Department of Epidemiology and Biostatistics, School of Public Health, Drexel University, 245 North 15th Street, Mailstop 660, Philadelphia, PA 19102, USA
[f] Department of Epidemiology, Boston University School of Public Health; 715 Albany Street, Boston, MA 02118, USA
[g] Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA
[h] Department for Society, Human Development, and Health, Harvard University School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Received 6 May 2008
Received in revised form
10 November 2008
Accepted 10 November 2008
Available online 30 December 2008

Keywords:
Binge alcohol use
HIV/STI
Sexual behavior

ABSTRACT

Background: Studies on the mechanisms of the association between illicit drug use and HIV/STI provide important insight into why there are disproportionate rates of HIV/STI among heterosexual African American men; far less work has been conducted to examine the associations between binge alcohol use and HIV/STI risks in this population.
Objective: To assess whether binge alcohol use is associated with risky sexual behaviors and recent HIV/STI diagnosis among heterosexual African American men reporting multiple sex partners in the past year.
Methods: Participants ($n = 672$) were heterosexually active African American men age 18–65 years recruited from urban health centers and clinics in Boston, MA, and who participated in a health survey. Logistic regression analyses were used to assess associations between past 30 day binge drinking and the following outcome variables: unprotected sex, six or more sex partners in the past year, sex trade involvement, and past 6 month HIV/STI diagnosis. Analyses were adjusted to control demographics, incarceration history, illicit drug use, and injection drug use.
Results: Significant associations were observed between binge alcohol use and unprotected vaginal sex with non-main female partners (AOR = 1.7, 95% CI = 1.2–2.3), unprotected anal sex with non-main female partners (AOR = 2.3, 95% CI = 1.4–4.0), sex trade involvement (AOR = 2.1, 95% CI = 1.3–3.5), and recent HIV/STI diagnosis (AOR = 1.9; 95% CI = 1.05–3.6).
Conclusion: Heterosexual African American men engaging in binge alcohol use may be at increased risk for HIV/STI; findings support the need for integrating alcohol risk reduction into HIV prevention programs targeting this population.

1. Introduction

Recent data from the United States (US) Centers for Disease Control and Prevention document that new HIV infections in the US are 40% greater than originally thought (56,300 new infections per year), and African Americans, though only 13% of the US population, account for 45% of these new infections (Hall et al., 2008; Centers for Disease Control and Prevention, 2008). Although African American MSMs remain the population at greatest risk for HIV/AIDS (Hall et al., 2008), 22% of HIV-infected African American males acquired the virus via heterosexual transmission (Centers for Disease Control and Prevention, 2008). Notably, 66% of US men who acquired HIV heterosexually are African American (Centers for Disease Control and Prevention, 2007). High rates of heterosexual HIV in this population are likely linked to their disproportionate representation in US STI cases (Centers for Disease Control and Prevention, 2006); STI-infected men (and women) are at heightened risk for acquiring HIV (Freeman et al., 2006; Fleming and Wasserheit, 1999; Cohen,

* Corresponding author. Tel.: +1 617 638 5160; fax: +1 617 638 4483.
E-mail address: anitaraj@bu.edu (A. Raj).

0376-8716/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.drugalcdep.2008.11.008

1998). Condom use has been identified as the most effective means of preventing HIV and some STIs among those who are sexually active, and as such, is recommended broadly to reduce HIV/STI risk across populations (Sangani et al., 2004). However, condom use is actually greater among African American compared with White males (Eaton et al., 2006), and racial/ethnic disparities in HIV/STI persist (Centers for Disease Control and Prevention, 2006, 2007, 2008). Other factors must be explored to understand what propels the HIV/STI epidemics in African American communities despite higher rates of condom use.

Both binge alcohol and illicit drug use have been identified as risk factors associated with HIV/STI diagnosis among with heterosexual African American males; notably, unprotected sex appears not to be associated with HIV/STI diagnosis in this population (Adimora et al., 2006a, 2006b; Raj et al., 2008). Such findings may best be explained by other risky sexual practices linked with substance use. In terms of illicit drug use, a number of studies indicate that such use, particularly crack/cocaine use, increases African American males' involvement in concurrent and multiple sex partnering as well as sex trade, risky sexual behaviors which are in turn associated with HIV/STI (Adimora et al., 2001, 2003, 2004, 2006a, 2006b; Adimora and Schoenbach, 2002, 2005; Braithwaite and Stephens, 2005; Essien et al., 2005; Jones, 1997; Lane et al., 2004a, 2004b; Whitehead, 1997). Those engaging in heavier illicit drug use appear additionally to be more likely to engage in unprotected sex (Braithwaite and Stephens, 2005; Raj et al., 2007; Rasch et al., 2000).

These risky sexual behaviors, and the illicit drug use related to them, are more common among African American men contending with social marginalization (i.e., poverty, homelessness) and a history of incarceration (Adimora et al., 2001, 2003, 2004, 2006a, 2006b; Adimora and Schoenbach, 2002, 2005; Cooke, 2004; Essien et al., 2005; Jones, 1997; Lane et al., 2004a, 2004b; Whitehead, 1997). Overall, these findings indicate that unprotected sex in isolation is not propelling the HIV epidemic among heterosexual African American men. However, among those men engaging in illicit drug use, unprotected sex does occur, and it occurs in the context of multiple partnering and sex trade involvement and within a climate of social risk. Hence, even if fewer unprotected sex episodes occur, exposure opportunities are greater for these more socially vulnerable men contending with illicit drug use.

The above findings provide important insight into the impact of illicit drug use on HIV/STI risk heterosexual African American men. However, such research on the role of binge alcohol use in this population is more limited. Alcohol use may be heightening risk for HIV/STI among heterosexual African American men by impeding their condom use at sexual incidents, however, it could also, like illicit drug use, be a marker or a risk factor for other risky sexual behaviors, such as multiple partnering and sex trade involvement (Collins et al., 2005; Corte and Sommers, 2005; Maisto et al., 2004). Research directly assessing associations between alcohol use and HIV/STI risks among exclusively or predominantly heterosexual samples of African American males has been limited to substance using or prison populations and reveals mixed findings. Alcohol was associated with multiple partnering and unprotected sex among African American crack users (Rasch et al., 2000), but not among detoxification patients (Raj et al., 2007) or incarcerated males (Braithwaite and Stephens, 2005). Research is needed to needed to explore these issues with a more generalizable samples of heterosexually at risk African American males, e.g., those not exclusively substance using or currently incarcerated. The purpose of this study is to assess associations between binge alcohol use and risky sex practices, as well as HIV/STI diagnosis, among heterosexually at risk African American men (i.e., men reporting multiple female sex partners), recruited from primary and urgent clinical care.

2. Methods

2.1. Study population

Study participants were from the Black and African American Men's Health Study (BAAMH), a cross-sectional study of black/African American adult men at sexual risk for HIV acquisition and/or transmission ($n = 703$). Participants were recruited from primary care and urgent care clinics within community health centers and a large teaching hospital; all recruitment sites were located in the Roxbury, Dorchester, and South End neighborhoods in Boston, Massachusetts; these neighborhoods have higher proportions of HIV/AIDS prevalence compared to that seen in the city as a whole (Boston Public Health Commission, 2007). The study involved a 30–45 min anonymous survey of participants on their sexual risk behaviors, demographics and related social and health concerns.

2.2. Recruitment and sample size

Trained research staff approached African American men attending five collaborating health centers and clinics during designated recruitment days and times from May 2005 to May 2006. Rotating days and times were used for recruitment across sites to reduce potential sample biases attached to work schedules. Eligible participants were those aged 18–65 years and reporting sex with two or more people in the past year; this latter criterion was used to obtain a more sexually at risk sample. Men scoring less than 21 on the Folstein Mini-Mental Exam (Folstein et al., 1975) were excluded from the study due to demonstrated cognitive impairment. Among the 2331 men approached, 85% ($n = 1988$) agreed to be screened for study eligibility. Of those screened ($n = 1988$), 47% (930) were eligible; 81% ($n = 754$) of eligible men agreed to study participation. Those who were ineligible or those who were eligible but refused participation were provided with local, low-cost or free social and health service referrals, including referral to HIV/STI counseling and testing, substance abuse treatment, batterers' intervention programs, mental health and trauma intervention, and job placement.

Of the 754 surveys collected, 51 (7%) were removed from further data analysis, as data from these participants indicated that they did not meet study criteria. Fifty participants did not respond to a question on number of sex partners; one participant indicated that he did not meet the age eligibility criterion. Current analyses were restricted to men reporting sex exclusively with women in the past year ($n = 617$); thus, an additional 86 men (12%) were excluded from analyses.

2.3. Study procedure

Written informed consent was obtained from eligible participants who were willing to complete the behavioral survey. After providing informed consent, participants completed an audio computer-assisted self-interview (ACASI) that was administered to obtain data on demographics, sexual risk histories, and related health and behavioral risks. ACASIs took approximately 20–25 min to complete and were administered in a private setting with a research staff member nearby to respond to participant questions, if needed. ACASI rather than interviewer-administered surveys were used due to evidence of more accurate responses to sensitive questions being obtained via ACASI (Simoes et al., 2006; Rogers et al., 2005; Abbey, 2005). Upon survey completion, all participants received $35 and low cost or free social and health service referrals. This study was approved by the Institutional Review Boards of Boston University Medical Campus and the Centers for Disease Control and Prevention. Additionally, a Federal Certificate of Confidentiality was obtained to provide further protections for study participants.

2.4. Measures

Measures for this study were obtained via self-report on ACASI items and included sociodemographics, sexual relationship characteristics, alcohol and illicit drug use, risky sexual behaviors and HIV/STI diagnosis.

2.4.1. Sociodemographics

Age, main relationship involvement, US born, education, employment status, homelessness, incarceration were assessed via single item measures. Although age was obtained as a continuous variable, this variable was dichotomized based on a median split due to skewed distribution. National origin was dichotomized to reflect population born in the US and US territories. Education was categorized to reflect whether the participant did not complete high school, received at least a high school education or equivalent (diploma or GED), or received at least some college. Employment was assessed by asking participants if they were unemployed, employed part-time or employed full time. Homelessness was assessed by asking where the participant was currently residing; those responding "on the streets" or "homeless in a housing shelter" were defined as homeless. Incarceration history was assessed via a single item on whether the participant had ever been to prison or jai and whether such an incident had occurred in the past year.

2.4.2. Alcohol use

Any Alcohol Use and Binge Alcohol Use were calculated using the past 30-day alcohol use questions from the Addiction Severity Index (ASI; McClellan et al., 1992).

Participants reporting alcohol use in the past 30 days were classified as having engaged in Any Alcohol Use; all reporting alcohol use were asked the number of days alcohol was used in the past 30 days. Additionally, binge alcohol use was assessed via an item on whether they had five or more drinks on an occasion, in the past 30 days.

2.4.3. Alcohol use and alcohol intoxication prior to sex and prior to unprotected sex

These items assessed, via yes/no questions created for this survey, whether the participant drank alcohol and whether they drank to intoxication within 2 h prior to vaginal or anal sex in the past 30 days; these items were each followed with an assessment of whether the participant used a condom the last time they engaged in that behavior.

2.4.4. Illicit drug use

Illicit drug use was calculated using the drug use frequency questions from the Addiction Severity Index (ASI; McClellan et al., 1992). Illicit drug use was assessed via a series of items on whether the individual used any of the following in the past 30 days: marijuana, heroin, hallucinogens, cocaine, inhalants, ecstasy, non-prescribed barbiturates, sedatives, opiates/analgesics and/or amphetamines. Those indicating use of any of these substances within this timeframe were defined as having engaged in illicit drug use. An additional item assessed whether injection drug use occurred in the past 6 months, to provide the variable *Any IDU*.

2.4.5. Unprotected Vaginal or Anal Sex with Main and Other (Non-Main) Female Partners

Any unprotected vaginal sex with a main female partner was constructed by asking the participant the number of times, in the past 90 days, they had penile–vaginal sex to ejaculation with a main female partner and subtracting that response from their response to the number of times, in the past 90 days, they used a condom when having this type of sex with that partner. This same procedure was also used to assess any unprotected vaginal sex with other (non-main) female partners in the past 90 days. The same questions (although specific to anal sex with main and non-main partners) and procedures described above were also used to assess any unprotected anal sex with a main female partner and any unprotected anal sex with other (non-main) female partners in the past 90 days. Due to highly skewed distributions, these variables were dichotomized as any versus no unprotected sex.

2.4.6. Number of sex partners and sex trade involvement

Number of female sex partners was assessed by asking participants the number of women with whom they had sex in the past year; due to skewed distribution, this variable was dichotomized via a median split (less than six female partners in the past year versus six or more female partners in the past year). Selling sex was assessed by asking participants if they had sold sex for drugs or money in the past 6 months, yes or no. Buying sex was assessed by asking participants whether or not they had given drugs or money to have sex with someone in the past 6 months, yes or no. Due to a high correlation between these variables and very low prevalence of selling sex, a summation score was created from these items and dichotomized to provide the *Sex Trade Involvement* variable.

Table 1
Demographic characteristics of the sample (African American men who have sex with women; $n = 617$) – Black and African American Men's Health Study, Boston, 2005–2006.

	% (n)
Age*	Mean age = 35.1 years (SD = 11.4)
US born	84.8% (523)
Education	
Less than high school completed	28.5% (176)
High school or GED completed	44.9% (277)
Some college or greater	26.6% (164)
Employment status	
Unemployed	60.9% (376)
Employed part time	19.1% (118)
Employed full time	19.9% (123)
Currently in relationship with main partner	73.3% (452)
Homelessness (on streets on in shelter)	22.7% (140)
Incarceration history	
Yes, in past year	26.1% (161)
Yes, but not in the past year	30.6% (189)
No, never	43.3% (267)

* As age is a continuous variable ranging from 18 to 65 years, mean and standard deviation rather than % (n) were used.

2.4.7. Recent HIV/STI diagnosis

Recent HIV/STI diagnosis was assessed by asking participants a series of items on whether or not they had ever or in the past 6 months been diagnosed with syphilis, gonorrhea, herpes or HPV (e.g., genital warts), or some other STI (unspecified) by a health provider. They were also asked if they had ever been told by a health care provider that they were HIV-positive and, if yes, when that occurred. STI responses were summated and dichotomized to yield STI diagnosis ever and in the past 6 months. HIV diagnosis was also created to yield ever and past 6-month data.

2.5. Data analysis

Descriptive data were obtained for all study measures, including frequencies on demographics, risky sexual behaviors and HIV/STI diagnoses, binge alcohol use and binge alcohol use-related unprotected sex. Simple logistic regression analyses were conducted to assess significant associations between binge alcohol use and the dependent variables (unprotected vaginal sex, unprotected anal sex, sex trade involvement, and HIV/STI diagnosis). Multivariate models were then created to determine whether binge alcohol use was significantly associated with the dependent variables, after controlling for potential confounders, including age, homelessness, employment, current steady relationship involvement, incarceration history, illicit drug use and IDU. Odds ratios and associated 95% confidence intervals were used to assess effect sizes and significance for variables in the crude and adjusted regression models.

3. Results

3.1. Sample characteristics

The majority of participants were unemployed (61%), and 23% were homeless (see Table 1). The majority (73%) reported current involvement in a steady relationship with a main partner, 88% of these had been in this relationship for 6 months or more, 40% had been in the relationship for 3 or more years.

3.2. Risky sexual behaviors and HIV/STI diagnosis history

In this sample of men reporting two or more female sex partners in the past year, 45% ($n = 279$) reported six or more female sex partners in the past year. (*Note*: Subsample sizes will be provided in this section, as the data are not presented in a table.)

Almost all participants (93%, $n = 581$) reported vaginal sex with a main female partner in the past 90 days; 53% of these ($n = 324$) reported never using condoms for any of these sexual episodes. More than one-third of participants (36%, $n = 220$) engaged in anal sex with a main partner in the past 90 days; 23% of these ($n/n = 50/220$) never used condoms in these contexts. The majority of the sample (79%, $n = 486$) reported vaginal sex with other female partners in the past 90 days; 21% of these never used condoms with these partners. One-fifth of men (22%, $n = 140$) reported anal sex with a non-main partner in the past 90 days; 24% reported never

Table 2
Prevalence of Alcohol use and alcohol use-related sexual behaviors reported by African American men who have sex with women ($n = 617$) – Black and African American Men's Health Study, Boston, 2005–2006.

	% (n)
Alcohol use, past 30 days	
Any alcohol use	57.7% (423)
Mean number of days alcohol used	Mean = 10.3 days (SD = 9.6)
Binge alcohol use	33.9% (209)
Alcohol use-related sexual behaviors	
Sex within 2 h after any alcohol use	33.7% (208)
Unprotected sex within 2 h after any alcohol use	49.0% (102/208)
Sex Within 2 H After Drinking to Intoxication	17.5% (108)
Unprotected sex within 2 h after drinking to intoxication	49.1% (53/108)

As number of drinking days in the past 30 days is a continuous variable ranging from 0 to 30 days, mean and standard deviation rather than % (n) were used.

Table 3
Crude and Adjusted regression analyses to assess associations between past 30 day binge alcohol use and HIV/STI behavioral risks and diagnosis among African American men who have sex with women ($n = 617$) – Black and African American Men's Health Study, Boston, 2005–2006.

	OR (95% CI)	AOR (95% CI)[a]
Unprotected vaginal sex – main partner, past 90 days	1.0 (0.7–1.3)	1.0 (0.7–1.5)
Unprotected anal sex – main partner, past 90 days	1.3 (0.8–2.1)	1.3 (0.8–2.1)
Unprotected vaginal sex – other female partners, past 90 days	1.7 (1.2–2.4)	1.7 (1.2–2.3)
Unprotected anal sex- other fm partners, past 90 days	2.4 (1.4–4.0)	2.3 (1.4–4.0)
Six or more female sex partners, past year	1.4 (1.02–2.0)	1.3 (0.9–1.9)
Sex trade involvement, past 6 months	2.2 (1.4–3.5)	2.1 (1.3–3.5)
HIV/STI diagnosis, past 6 months	2.2 (1.2–4.0)	1.9 (1.05–3.6)

Note: Significant findings characterized by 95% confidence intervals very close to 1.0 at the lower end of the interval are noted in the hundredths place to document significance.
[a] Adjusted models included the following variables: binge alcohol use, age, homelessness, employment status, incarceration history, relationship involvement, illicit drug use, IDU.

using condoms when engaging in anal sex with a non-main female partner in this timeframe.

One in seven men (15%, $n = 90$) reported sex trade involvement, with 13% ($n = 81$) having bought sex and 6% ($n = 34$) having sold sex. Approximately one-third of the sample (31%, $n = 190$) had a history of STI diagnosis, with 8% ($n = 47$) having received such a diagnosis in the past 6 months. HIV diagnosis was reported by 3% of the sample ($n = 18$), 1% ($n = 8$) were diagnosed with HIV in the past 6 months.

3.3. Alcohol use and alcohol use-related sexual behaviors; illicit drug use

Over half of participants (58%) reported past 30-day alcohol use, 34% reported binge alcohol use (5+ drinks in one sitting) in the past 30 days (see Table 2). One-third of the sample (34%) reported alcohol use prior to sex in the past 30 days, of these, 49% reported unprotected sex in this context. Eighteen percent of the sample reported binge alcohol use prior to sex in the past 30 days, 49% of these reported unprotected sex in this context.

Almost half the sample (44.7%) reported past 30-day illicit drug use. Approximately one-third of participants (31.8%) reported marijuana use, 16.0% reported crack/cocaine use; 5.8% reported heroin use, and 5.2% reported use of other opiates. Less than 2% reported past 30 days use of ecstasy, sedatives/benzos, barbiturates, amphetamines, hallucinogens, and inhalants, respectively.

3.4. Past 30 day binge alcohol use and relation to risky sexual behaviors and recent HIV/STI diagnosis

Binge alcohol use was significantly associated with unprotected vaginal sex with non-main female partners (AOR = 1.7, 95% CI = 1.2–2.3) and unprotected anal sex with non-main female partners (AOR = 2.3, 95% CI = 1.4–4.0); it was not significantly association with unprotected vaginal or anal sex with main female partners (see Table 3). Binge alcohol users were also significantly more likely to report sex trade involvement (AOR = 2.1, 95% CI = 1.3–3.5). Exploratory analyses were conducted to assess whether these effects differed for buying compared with selling sex, significantly correlated items collapsed to create the sex trade variable. Binge alcohol use was similarly associated with both variables (buying sex AOR = 1.8, 95% CI = 1.1–3.0; selling sex AOR = 2.3, 95% CI = 1.1–5.0). Finally, binge alcohol use was also significantly associated with recent HIV/STI diagnosis (AOR = 1.9; 95% CI = 1.05–3.6).

4. Discussion

Findings from the current study with heterosexually at risk African American men recruited from urban primary and urgent clinic care demonstrate that one-third of these men have engaged in binge alcohol use within the past 30 days; this is a higher level of binge alcohol use than that seen in population-based national data with African American men (34% vs. 19%) (Substance Abuse and Mental Health Services Administration, 2007). Study results additionally document that such behaviors are occurring, not uncommonly, prior to sex. Almost one in five participants reported alcohol intoxication prior to sex in the past 30 days; half of these reported no condom use at their last sex subsequent to alcohol intoxication. While data from the current study do not provide a comparison of condom use at last sex without alcohol use or even condom use at last sexual episode, these findings nonetheless reveal that heavy alcohol use prior to sex is an important issue to understand for heterosexually at risk African American men. This is reinforced by study findings demonstrating increased likelihood of riskier sex and recent HIV/STI diagnosis among those men reporting binge alcohol use.

Binge drinkers were more likely than non-binge drinkers to engage in unprotected vaginal and anal sex with non-main female partners and to be involved in sex trade, both buying and selling sex. Further, they were significantly more likely to have received a recent HIV/STI diagnosis. However, binge alcohol use was not significantly associated with unprotected sex in main relationships. Overall, these findings demonstrate that similar to illicit drug use, binge drinking in this population is linked to riskier sexual activities and HIV/STI. These results are consistent with previous research involving samples with a large proportion of African American male participants, including crack/cocaine users (Rasch et al., 2000), injection drug users (Stein et al., 2000, 2001), STI clinic patients (Kalichman and Cain, 2004), and HIV-infected patients (Kalichman et al., 2002; Stein et al., 2005). Notably, although many of these previous studies additionally documented an association between alcohol use and having a greater number of sex partners, such findings did not hold true for this sample; this is likely a consequence of the current samples' inclusion of only those reporting multiple sex partners in the past year. Further study is needed to determine if a link between alcohol and multiple sex partnering exists in a broader sample of African American men who have sex with women.

Overall, findings from this study provide important insight into studies, including previously published work with this study sample, which have found that unprotected sex is not significantly associated with recent HIV/STI diagnosis among heterosexual African American men (Adimora et al., 2006a; Raj et al., 2008). As described in the introduction and related to illicit drug use, study results indicate that binge alcohol use is linked to unprotected sex in the context of higher risk sexual situations, rather than unprotected sex generally (inclusive of sex with steady partners), for heterosexually at risk African American men. This unprotected sex in higher risk sexual situations (i.e., with casual sex partners, in a sex trade context) may help drive the disproportionate rates of HIV/STI in this population. Further research is needed to confirm that unprotected sex with sex trade partners is more likely in this population as current findings only provide information on sex trade involvement,

regardless of unprotected sex, and unprotected sex with casual sex partners.

4.1. Limitations

While current study offers important insight into HIV/STI risk among heterosexual African American men, these results must be considered in light of several study limitations. Findings have limited generalizability due to our use of a northeastern US community-based clinic sample of African American men reporting two or more sex partners in the past year and with very high rates of incarceration, unemployment and homelessness. Additionally, our assessments of numbers of partners did not provide sufficient information with regard to concurrent sexual partnering. Further, although our collaborating health centers and clinics are typical of those throughout the region in terms of location within a lower income area and serving predominantly racial/ethnic minority and lower income client populations, these collaborators may have stronger links to substance abuse treatment programs, prison re-entry programs, and transitional housing programs as compared to many other health centers. Hence, our sample may over-represent men misusing alcohol or illicit drugs, engaging in sex trade, having a history of incarceration or experiencing homelessness. Additionally, as our study included men seeking varied types of assistance and non-medical programs at our recruitment sites, findings cannot be generalized to those seeking traditional clinical care.

In addition to limitations related to generalizability, there are a number of limitations related to study design. This research was cross-sectional; hence, causality cannot be assumed. Further, timeframes used to assess substance use behaviors, risky sex practices, and HIV/STI diagnoses were not always consistent, ranging from 30 days to 1 year. Reliance on self-report also make these data subject to social desirability and recall biases. However, these biases would likely result in the under-reporting rather than over-reporting of sensitive issues such as illicit drug use and sex trade involvement. We used ACASI methods to at least partially mitigate these biases. Biological markers of HIV/STI and behaviors such as illicit drug use would have been helpful for validation of self-reported data. In addition to these concerns, analyses involved multiple comparisons increasing the risk for spurious findings. Longitudinal study of these issues among a larger and more representative sample of African American men and with more direct rather than self-report measurements is needed to improve examinations of these issues. Additionally, qualitative research to explore the mechanisms of observed associations is also needed to support development of HIV/STI prevention and intervention programs for this population (see Essien et al., 2005).

4.2. Conclusion and implications

In our primary and urgent care clinic sample of heterosexual African American men reporting multiple sex partners in the past year, those reporting recent binge alcohol use were significantly more likely than non-binge drinkers to report unprotected vaginal and anal sex with casual sex partners, sex trade involvement (buying and selling), and recent HIV/STI diagnosis. Such findings clarify recent research documenting a significant association between binge alcohol use and HIV/STI, but not unprotected sex and HIV/STI, among heterosexual African American men. Within this population, binge alcohol users appear to be more likely to engage in unprotected sex within higher exposure contexts (i.e., with casual or sex trade partners), and such behavior within such contexts, rather than unprotected sex generally, may be propelling the HIV/STI epidemic in this population. Longitudinal study is needed to determine whether this is the case; qualitative research is needed to understand why this may be the case. Effective HIV/STI behavioral interventions for heterosexually at risk African American men will likely need to be integrated with alcohol treatment or risk reduction to meet the needs of this vulnerable population.

Role of funding source

This project was funded under a grant from the Centers for Disease Control and Prevention (CDC) (grant number: CCU123364). Dr. Stephen A. Flores, CDC project officer on this study was involved with the development of this paper and is included as an author on the paper. However, the findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Contributors

Drs. Raj, Welles, Horsburgh and Flores were involved in the study design and protocol development for this project. Drs. Raj, Reed, Silverman and Walley conducted the literature review for the development of this paper. Drs. Raj, Reed, Silverman and Welles were involved in development of the analysis plan for the paper, and Drs. Raj and Reed conducted all data analyses. Ms. Santana assisted in the writing of this paper and oversaw all aspects of project implementation in the field. All authors were involved in interpretation of study findings for this paper. Drs. Raj and Silverman led conceptualization of the paper, and Dr. Raj led all writing for this paper and served as principal investigator on this CDC study. All authors contributed to and have approved the final manuscript.

Conflict of interest

There are no actual or potential conflicts of interest for any authors of this study.

Acknowledgements

We would like to thank the following organizations and agencies for housing and consulting on our study: Multicultural AIDS Coalition, Whittier Street Health Center, DOTWELL: Codman Square and Dorchester House Community Health Services, and Boston Medical Center's Urgent Care Clinic. We would also like to thank Dumas Lafontant and C. Robert Horsburgh, as well as our CDC Project Officers Stephen Flores and Lisa Fitzpatrick, for their support and guidance on this project.

References

Abbey, A., 2005. Lessons learned and unanswered questions about sexual assault perpetration. J. Interpers Violence 20 (1), 39–42.

Adimora, A.A., Schoenbach, V.J., 2005. Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. J. Infect. Dis. 191 (Suppl. 1), S115–S122.

Adimora, A.A., Schoenbach, V.J., 2002. Contextual factors and the black-white disparity in heterosexual HIV transmission. Epidemiology 13 (6), 707–712.

Adimora, A.A., Schoenbach, V.J., Martinson, F., Donaldson, K.H., Stancil, T.R., Fullilove, R.E., 2004. Concurrent sexual partnerships among African Americans in the rural south. Ann. Epidemiol. 14 (3), 155–160.

Adimora, A.A., Schoenbach, V.J., Martinson, F.E., Coyne-Beasley, T., Doherty, I., Stancil, T.R., Fullilove, R.E., 2006a. Heterosexually transmitted HIV infection among African Americans in North Carolina. J. Acquir. Immune Defic. Syndr. 41 (5), 616–623.

Adimora, A.A., Schoenbach, V.J., Doherty, I.A., 2006b. HIV and African Americans in the southern United States: Sexual networks and social context. Sex Transm. Dis. 33 (7 Suppl), S39–45.

Adimora, A.A., Schoenbach, V.J., Martinson, F.E., Donaldson, K.H., Fullilove, R.E., Aral, S.O., 2001. Social context of sexual relationships among rural African Americans. Sex Transm. Dis. 28 (2), 69–76.

Adimora, A.A., Schoenbach, V.J., Martinson, F.E., Donaldson, K.H., Stancil, T.R., Fullilove, R.E., 2003. Concurrent partnerships among rural African Americans with recently reported heterosexually transmitted HIV infection. J. Acquir. Immune Defic. Syndr. 34 (4), 423–429.

Boston Public Health Commission, 2007. Health of Boston 2007. Boston Public Health Commission Research Office. http://www.bphc.org/reports/pdfs/report_224.pdf. Accessed August 16, 2008.

Braithwaite, R., Stephens, T., 2005. Use of protective barriers and unprotected sex among adult male prison inmates prior to incarceration. Int. J. STD AIDS 16 (3), 224–226.

Centers for Disease Control and Prevention, 2008. CDC HIV/AIDS Fact Sheets. HIV/AIDS and African Americans. Revised August 2008. http://www.cdc.gov/hiv/topics/aa/resources/factsheets/pdf/aa.pdf. Accessed August 16, 2008.

Centers for Disease Control and Prevention, 2007. HIV/AIDS Surveillance Report, Volume 17, Revised Edition, June 2008. Table 9. Estimated numbers of persons living with HIV/AIDS at the end of 2005, by race/ethnicity, sex, and transmission category—33 states with confidential name-based HIV infection reporting; Table 19. Reported AIDS cases for male adults and adolescents, by transmission category and race/ethnicity, 2006 and cumulative—United States and dependent areas. http://www.cdc.gov/hiv/topics/surveillance/resources/reports/2006report/default.htm. Accessed August 16, 2008.

Centers for Disease Control and Prevention, 2006. STD Surveillance 2006. http://www.cdc.gov/std/stats/toc2006.htm. Accessed August 16, 2008.

Cohen, M.S., 1998. Sexually transmitted diseases enhance HIV transmission: no longer a hypothesis. Lancet 351 (Suppl. 2), 5–7.

Collins, R.L., Ellickson, P.L., Orlando, M., Klein, D.J., 2005. Isolating the nexus of substance use, violence and sexual risk for HIV infection among young adults in the United States. AIDS Behav. 9 (1), 73–87.

Cooke, C.L., 2004. Joblessness and homelessness as precursors of health problems in formerly incarcerated African American men. J. Nurs. Scholarsh. 36 (2), 155–160.

Corte, C.M., Sommers, M.S., 2005. Alcohol and risky behaviors. Annu. Rev. Nurs. Res. 23, 327–360.

Eaton, D.K., Kann, L., Kinchen, S., Ross, J., Hawkins, J., Harris, W.A., Lowry, R., McManus, T., Chyen, D., Shanklin, S., Lim, C., Grunbaum, J.A., Wechsler, H., 2006. Youth Risk Behavior Surveillance – United States, 2005. MMWR CDC Surveill. Summ. 55 (SS-05), 1–108.

Essien, E.J., Meshack, A.F., Peters, R.J., Ogungbade, G.O., Osemene, N.I., 2005. Strategies to prevent HIV transmission among heterosexual African-American men. BMC Public Health 5 (1), 3.

Fleming, D.T., Wasserheit, J.N., 1999. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex Transm. Infect. 75 (1), 3–17.

Freeman, E.E., Weiss, H.A., Glynn, J.R., Cross, P.L., Whitworth, J.A., Hayes, R.J., 2006. Herpes simplex virus 2 infection increases HIV acquisition in men and women: systematic review and meta-analysis of longitudinal studies. AIDS 20 (1), 73–83.

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. Psychiatry Res. 12 (3), 189–198.

Hall, H.I., Song, R., Rhodes, P., Prejean, J., An, Q., Lee, L.M., Karon, J., Brookmeyer, R., Kaplan, E.H., McKenna, M.T., Janssen, R.S., The HIV Incidence Surveillance Group, 2008. Estimation of HIV incidence in the United States. JAMA 300 (5), 520.

Jones, R., 1997. Violence and the politics of Black male identity in post-modern America. J. Afr. Am. Men 3 (2), 81–107.

Kalichman, S.C., Cain, D., 2004. A prospective study of sensation seeking and alcohol use as predictors of sexual risk behaviors among men and women receiving sexually transmitted infection clinic services. Psychol. Addict. Behav. 18 (4), 367–373.

Kalichman, S.C., Weinhardt, L., DiFonzo, K., Austin, J., Luke, W., 2002. Sensation seeking and alcohol use as markers of sexual transmission risk behavior in HIV-positive men. Ann. Behav. Med. 24 (3), 229–235.

Lane, S.D., Rubinstein, R.A., Keefe, R.H., Webster, N., Cibula, D.A., Rosenthal, A., Dowdell, J., 2004a. Structural violence and racial disparity in HIV transmission. J. Health Care Poor Underserved 15, 319–335.

Lane, S.D., Keefe, R.H., Rubinstein, R.A., Levandowski, B.A., Freedman, M., Rosenthal, A., Cibula, D.A., Czerwinski, M., 2004b. Marriage promotion and missing men: African American women in a demographic double bind. Med. Anthropol. Q. 18 (4), 405–428.

Maisto, S.A., Carey, M.P., Carey, K.B., Gordon, C.M., Schum, J.L., Lynch, K.G., 2004. The relationship between alcohol and individual differences variables on attitudes and behavioral skills relevant to sexual health among heterosexual young adult men. Arch. Sex Behav. 33 (6), 571–584.

McClellan, A.T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., Argeriou, M., 1992. The fifth edition of the addiction severity index. J. Subst. Abuse Treat. 9, 199–213.

Raj, A., Reed, E., Santana, M.C., Welles, S.L., Horsburgh, C.R., Flores, S.A., Silverman, J.G., 2008. History of incarceration and gang violence are associated with recent STD/HIV diagnosis in African American men. J. Acquir. Immune Defic. Syndr. 47 (1), 131–134.

Raj, A., Saitz, R., Cheng, D.M., Winter, M., Samet, J.H., 2007. Associations between alcohol, heroin, and cocaine use and high risk sexual behaviors among detoxification patients. Am. J. Drug Alcohol Abuse 33 (1), 169–178.

Rasch, R.F., Weisen, C.A., MacDonald, B., Wechsberg, W.M., Perritt, R., Dennis, M.L., 2000. Patterns of HIV risk and alcohol use among African-American crack abusers. Drug Alcohol Depend. 58 (3), 259–266.

Rogers, S.M., Willis, G., Al-Tayyib, A., Villarroel, M.A., Turner, C.F., Ganapathi, L., Zenilman, J., Jadack, R., 2005. Audio computer assisted interviewing to measure HIV risk behaviours in a clinic population. Sex Transm. Infect. 81 (6), 501–507.

Substance Abuse and Mental Health Services Administration, 2007. Results from the 2006 National Survey on Drug Use and Health: National Findings. Figure 3.3 Current, Binge, and Heavy Alcohol Use among Persons Aged 12 or Older, by Race/Ethnicity: 2006. Substance Abuse and Mental Health Services Administration. Office of Applied Studies. Department of Health and Human Services. http://www.oas.samhsa.gov/NSDUH/2k6NSDUH/2k6results.cfm#Ch3. Accessed August 16, 2008.

Sangani, P., Rutherford, G., Wilkinson, D., 2004. Population-based interventions for reducing sexually transmitted infections, including HIV infection. Cochrane Database Syst. Rev. 2, CD001220.

Simoes, A.A., Bastos, F.I., Moreira, R.I., Lynch, K.G., Metzger, D.S., 2006. A randomized trial of audio computer and in-person interview to assess HIV risk among drug and alcohol users in Rio De Janeiro, Brazil. J. Subst. Abuse Treat. 30 (3), 237–243.

Stein, M.D., Herman, D.S., Trisvan, E., Pirraglia, P., Engler, P., Anderson, B.J., 2005. Alcohol use and sexual risk behavior among human immunodeficiency virus-positive persons. Alcohol Clin. Exp. Res. 29 (5), 837–843.

Stein, M.D., Anderson, B., Charuvastra, A., Friedmann, P.D., 2001. Alcohol use and sexual risk taking among hazardously drinking drug injectors who attend needle exchange. Alcohol Clin. Exp. Res. 25 (10), 1487–1493.

Stein, M.D., Hanna, L., Natarajan, R., Clarke, J., Marisi, M., Sobota, M., Rich, J., 2000. Alcohol use patterns predict high-risk HIV behaviors among active injection drug users. J. Subst. Abuse Treat. 18 (4), 359–363.

Whitehead, T.L., 1997. Urban low-income African American men, HIV/AIDS, and gender identity. Med. Anthropol. Q. 11 (4), 411–447.

Review Article

Evaluating the Cost-Effectiveness of Cancer Patient Navigation Programs: Conceptual and Practical Issues

Scott Ramsey, MD, PhD[1]; Elizabeth Whitley, RN, PhD[2]; Victoria Warren Mears, RD, PhD[3]; June M. McKoy, MD, MPH JD[4,5]; Rachel M. Everhart, MS[2]; Robert J. Caswell, PhD[6]; Kevin Fiscella, MD, MPH[7]; Thelma C. Hurd, MD[8]; Tracy Battaglia, MD, MPH[9]; and Jeanne Mandelblatt, MD, MPH[10], for the Patient Navigation Research Program Group

Patient navigators—individuals who assist patients through the healthcare system to improve access to and understanding of their health and healthcare—are increasingly used for underserved individuals at risk for or with cancer. Navigation programs can improve access, but it is unclear whether they improve the efficiency and efficacy of cancer diagnostic and therapeutic services at a reasonable cost, such that they would be considered cost-effective. In the current study, the authors outline a conceptual model for evaluating the cost-effectiveness of cancer navigation programs. They describe how this model is being applied to the Patient Navigation Research Program, a multicenter study supported by the National Cancer Institute's Center to Reduce Cancer Health Disparities. The Patient Navigation Research Program is testing navigation interventions that aim to reduce time to delivery of quality cancer care (noncancer resolution or cancer diagnosis and treatment) after identification of a screening abnormality. Examples of challenges to evaluating cost-effectiveness of navigation programs include the heterogeneity of navigation programs, the sometimes distant relation between navigation programs and outcome of interest (eg, improving access to prompt diagnostic resolution and life-years gained), and accounting for factors in underserved populations that may influence both access to services and outcomes. In this article, the authors discuss several strategies for addressing these barriers. Evaluating the costs and impact of navigation will require

Corresponding author: Scott Ramsey, MD, PhD, Cancer Technology Assessment Group, Translational and Outcomes Research, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North M3-B232, PO Box 19024, Seattle, WA 98109; Fax: (206) 667-5977; sramsey@fhcrc.org

[1]Cancer Technology Assessment Group, Translational and Outcomes Research, Fred Hutchinson Cancer Research Center, Seattle, Washington; [2]Community Voices, Denver Health and Hospital Authority, Denver, Colorado; [3]Northwest Portland Area Indian Health Board, Northwest Tribal Epidemiology Center, Portland, Oregon; [4]Department of Medicine, Northwestern University, Feinberg School of Medicine, Chicago, Illinois; [5]Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, Chicago, Illinois; [6]College of Public Health, Ohio State University, Columbus, Ohio; [7]Department of Family Medicine, University of Rochester School of Medicine, Rochester, New York; [8]School of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas; [9]Women's Health Research Unit, Boston University, Boston, Massachusetts; [10]Cancer Control Program, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, District of Columbia

The following are Patient Navigation Research Program Principal Investigators: Charles L. Bennett, MD, PhD, MPP (Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, Ill); Elizabeth Calhoun, PhD (Health Policy and Administration, University of Illinois at Chicago, Chicago, Ill); Donald J. Dudley, MD (University of Texas Health Science Center, San Antonio, San Antonio, Tex); Kevin Fiscella, MD, MPH (University of Rochester Medical Center, Rochester, NY); Karen M. Freund, MD, MPH and Tracy Battaglia, MD, MPH (Co-Principal Investigators) (Boston University Medical Center, Women's Health Research Unit, Boston, Mass); Victoria Warren Mears, PhD (Northwest Portland Area Indian Health Board, Portland, Ore); Electra D. Paskett, PhD (Division of Epidemiology, College of Public Health, Comprehensive Cancer Center, Ohio State University, Columbus, Ohio); Steven R. Patierno, PhD (George Washington University Cancer Institute, Washington, DC); Peter C. Raich, MD (Denver Health and Hospital Authority, Denver, Colo); and Richard G. Roetzheim, MD, MSPH (H. Lee Moffitt Cancer Center and Research Institute, Tampa, Fla). The following are Center to Reduce Cancer Health Disparities Patient Navigation Research Program Directors: Roland Garcia, PhD; Mary Ann Van Duyn, PhD; and Emmanuel Taylor, DrPH. The following are NOVA Research Company Project Managers: Amanda Greene, PhD, MPH, RN and Paul Young, MPH, MBA.

Received: September 24, 2008; **Revised:** December 15, 2008; **Accepted:** January 13, 2009

Published online August 14, 2009 in Wiley InterScience (www.interscience.wiley.com)

DOI: 10.1002/cncr.24603, www.interscience.wiley.com

some novel methods, but will be critical in recommendations concerning dissemination of navigation programs. **Cancer 2009;115:5394–403. © 2009 American Cancer Society.**

KEY WORDS: cancer, navigator, cost-effectiveness, modeling.

Populations with limited access to or knowledge of the healthcare system often have difficulty using the system effectively for cancer services, and this may result in delays in cancer diagnosis,[1,2] added costs,[3] and less efficient and effective use of recommended therapies. Patient navigation programs provide support and guidance to persons with the goal of improving access to the cancer care system and overcoming barriers to timely, quality care.[4-14] In this article, we present a conceptual model for evaluating the cost-effectiveness of cancer patient navigation programs, discuss methodologic challenges, and suggest approaches for addressing these challenges.

Rationale for and History of Patient Navigation Programs

The origins of patient navigator programs are widely attributed to Harold Freeman, who, as president of the American Cancer Society (ACS), commissioned a study of barriers to cancer care among the poor in the United States. The report documented substantial disparities in both cancer care and outcomes between poor and nonpoor Americans, identifying, among other issues, significant barriers to care and a sense of fatalism regarding cancer that prevented many from seeking care in the first place.[15] As a result of this report, the ACS supported the first Patient Navigation program in 1990 at the Harlem Hospital Center. A pre-post comparison of women diagnosed with breast cancer at this facility demonstrated that 41% of breast cancer patients diagnosed between 1995 and 2000 were diagnosed with early disease, compared with 6% of patients diagnosed between 1964 and 1986.[16,17] Five-year survival rates increased from 39% to 70% over the same period.

Because of the success of this pioneer program, and in recognition that significant barriers to effective cancer screening, diagnosis, and care continue to exist among minority and underserved populations, patient navigation programs are becoming more common, particularly among health systems that serve these populations. The Centers for Medicare and Medicaid Services is funding demonstration projects to reduce barriers to care at all levels.[18] Despite their growing popularity and the publication of promising observational studies,[19-22] to our knowledge very few prospective, controlled trials have evaluated the efficacy of navigator programs. Controlled trials, most of which are small, have shown significant improvements in time to diagnosis, reductions in anxiety, and greater levels of satisfaction with the care process.[23-25] The impact of navigation programs on cancer-related morbidity and survival, and the cost-effectiveness of these programs, are not yet known.

The Patient Navigation Research Program

The National Cancer Institute and the ACS are sponsoring a 9-site Patient Navigation Research Program (Table 1).[10] The primary aim of the Patient Navigation Research Program is to evaluate navigation programs' impact on the time from an abnormal finding (from a screening test or clinical examination for case finding) to definitive diagnosis and treatment initiation. Secondary aims include evaluating the impact of navigation on patient satisfaction and the cost-effectiveness of navigation.

Patient Navigation Research Program sites serve diverse patient populations. Navigation programs focus on follow-up of abnormal breast, cervical, prostate, and colorectal cancer screening tests, among minority populations including African Americans, American Indians, Asians, Hispanics, and the rural underserved. Navigation models vary across sites, using different professionals and healthcare systems (Table 1) to follow patients through the completion of initial treatment.

Rationale for Evaluating the Cost-Effectiveness of Patient Navigation Programs

Patient navigator programs can be time and resource intensive. Similar to other interventions that may improve the health of poor and underserved populations, navigation programs must be viewed in the context of allocating

Table 1. Patient Navigation Research Program Study Populations, Setting, and Programs

PN Sites	Cancers	Populations	Navigator	Study Design	Setting	PN Intervention	Control
Boston University	Breast Cervix	B H U	6 O	Group randomized, controlled	Community health center	1200	1200
Denver Health and Hospital Authority	Breast Colorectal Prostate	B H U A/PI AI/AN	4.5 Lay	Randomized	Community health center, hospital	870	870
George Washington University, Washington, DC	Breast	B H U	1 NP 1 SW 7 O	Nonrandomized, controlled	Clinic	800	800
H. Lee Moffitt Cancer Center	Breast Colorectal	B H U	3 Lay	Group randomized, controlled	Clinic and hospital	600	600
Northwest Portland Area Indian Health Board	Breast Cervix Colorectal Prostate	AI/AN	3 RN 1 Lay	Nonrandomized, controlled	Clinic	650	650
University of Illinois at Chicago/Northwestern University, Chicago	Breast Cervix Colorectal Prostate	B H U	2 SW, 5 Lay	Randomized, controlled	Community health centers, clinics, hospital	2500	2500
University of Rochester, NY	Breast Colorectal	B U	3 LAY	Randomized, controlled (patient)	Hospital	400	400
University of Texas Health Science Center at San Antonio	Breast Cervix	B H U	4 PRO, 2 RN, 2 SW	Nonrandomized, controlled	Clinic	700	700
Ohio State University, Columbus	Breast Cervix Colorectal	B H U	3 LAY	Group randomized, controlled	Clinic	4258	4258
						11,978*	11,978*

PN indicates Patient Navigation; B, Black, H, Hispanic; U, underserved navigator; O, other; A/PI, Asian and Pacific Islander; AI/AN, American Indian/Alaska Native; Lay, lay or community worker; NP, nurse practitioner, nurse clinicians, physician assistant; SW, social worker; RN, registered nurse; PRO, Promotoras. For more information, visit http://crchd.cancer.gov/pnp/pnrp-index.html Accessed August 4, 2009.
*Totals.

resources such that health outcomes are maximized under limited budgets. It is particularly important to evaluate the cost-effectiveness of publicly funded navigator programs, because funding for these programs typically come from global health budgets that are fixed in the short run with many competing needs. Cost-effectiveness analysis can assist decision makers by demonstrating the health benefit for expenditure of navigator programs relative to other interventions, particularly those that are targeted to the same disease or condition of interest. The desirability of navigator programs can also be assessed in terms of commonly accepted thresholds (eg, $100,000 per quality-adjusted life-year [QALY] gained) in the health system or country.[26]

Conceptual Model for Cost-Effectiveness Analysis of Patient Navigation Interventions

For the Patient Navigation Research Program, we are using cost-effectiveness analysis to compare the added (incremental) costs of navigation interventions versus those of the status quo for the given target populations.[27] Cost-effectiveness analysis is a comparison of alternatives, typically a new intervention such as navigation versus usual care, which is patients and their family members seeking care without formal assistance. Costs and consequences flowing from each alternative (navigated vs usual care) are summarized over the time period that is relevant to the episode of care (Fig. 1). The incremental cost-effectiveness of navigation is derived using the following formula:

$$\text{Incremental Cost-Effectiveness Navigation} = (C_{Nav} - C_{UC})/(E_{Nav} - E_{UC}) \quad (1)$$

in which C_{Nav} and C_{UC} refer to the incremental difference in total costs of the navigation program compared with usual care, and E_{Nav} and E_{UC} refer to the difference in total effectiveness between navigation and usual care (Fig.

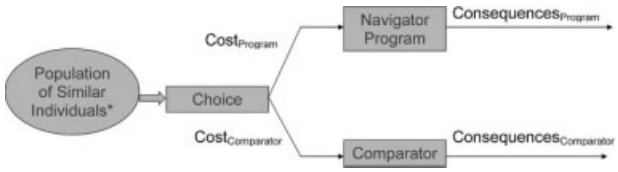


FIGURE 1. A conceptual model of patient navigator intervention versus usual care is shown. *Examples may include persons eligible for cancer screening procedures or those with cancer who are eligible for treatment.

1). Although the comparator is typically usual care; that is, care as it occurs in usual practice in the absence of navigators, one could also compare $\geq$2 navigation programs versus usual care, or 1 program with another. Both the navigator program and usual care have costs that flow from the point of entry (eg, abnormal finding on mammogram) to short-term and long-term downstream costs and consequences. Generally the time horizon is the individuals' remaining years of life. Because the Patient Navigation Research Program will only observe individuals over a maximum of the 5 years of the program, examining impact on survival and costs per QALYs saved will require estimation using mathematical models.

Navigation Cost-Effectiveness Analysis and Approaches for Addressing Challenges

Evaluating the cost-effectiveness of patient navigation programs poses several unique challenges (Table 2). In this section, we describe particular challenges for evaluating the cost-effectiveness of the Patient Navigation Research Program and how we plan to address those issues.

Defining the navigation intervention

The first issue in conducting the cost-effectiveness analysis of navigation is that the navigator intervention itself is not uniform for all patients, because part of the principle of navigation is to identify patient-specific issues and tailor the program to those needs. Moreover, navigation interventions (including the Patient Navigation Research Program) are quite heterogeneous, and are typically tailored to the needs and available resources of a particular region and the cancers of interest. Even within a single program site, the navigator will tailor the intervention to the needs of the particular patient-client, with wide variation in services provided between individuals. A related issue is that programs differ in expectations, qualifications, training, and supervision of navigators. In many settings, navigators are trained to assist patients with abnormal screening tests for several cancers (eg, cervical and colorectal, or breast and prostate). Although there are economies of scale in these situations, it is more difficult to segregate the time costs for each cancer and model each separately. One could capture the economies of scale by modeling all screening, but this requires extension of the time horizon in a model capturing the natural history of multiple cancers at once.

Table 2. Unique Challenges to Evaluating the Cost-Effectiveness of Navigation Programs

Relation between navigation and endpoints (costs, survival, QALY) is nonlinear
Content (and costs) of navigation interventions are variable due to site-specific program needs
Confounding between need for navigation and stage, mortality endpoints
Difficulty in allocating costs and effects over multiple cancers
Short-term intervention outcomes (eg, distress) do not map easily to QALYs
Difficulty collecting uniform data across sites and at relevant time points (eg, time costs)
Difficulty detecting the impact of modest reductions in diagnostic or treatment delays on mortality
Personal characteristics of navigators (difficult to measure) may influence program effectiveness

QALY indicates quality-adjusted life-year.

At present, we are not aware of models that are designed to incorporate the natural history of multiple cancers simultaneously. However, this is an important research priority, because the majority of providers recommend screening for multiple cancers to their patients, and navigators assist individuals in navigating through to diagnostic resolution for $>$1 cancer type.

Therefore, we address the issue of the heterogeneity of interventions by defining the navigation programs broadly, as specified by the study protocols.[28] This approach emphasizes the type of navigator (eg, nurse, layperson) and the general scope of services that that individual is able to provide. We will then have to model the cost-effectiveness of navigation for each individual cancer separately, allocating navigator time and other efforts in proportion for each cancer site.

Measuring effectiveness of navigation programs

The recommended measure of effectiveness of navigation programs for cost-effectiveness analyses is the

QALY,[29] which requires data on survival with and without the program and evaluation of health state preferences (utilities). However, outcome measures being directly tracked by the Patient Navigation Research Program research sites are intermediate outcomes: time to definitive diagnosis/resolution and time to initiation/completion of recommended cancer therapy for those with a cancer diagnosis.[28] Moreover, the period of observation under the 5-year Patient Navigation Research Program will be too short to observe any mortality endpoints.

Estimating QALYs will require simulation modeling. To address the need to extrapolate from the observation period to estimate the impact of navigation over a lifetime, we will use simulation models to extend the time frame of observation and look at stage distribution of patients diagnosed under navigation and usual care, using local cancer registries, hospitals, and patient charts. Age-specific, race-specific, and stage-specific survival from cancer registries (local or national) can then be used to project the life expectancy, or mortality experience of each group of patients.

Even using this approach, modeling the effects of mortality based on delays in diagnosis or treatment is challenging and requires modeling assumptions. For example, most models portray screening benefits in terms of decreases in tumor size (and number of lymph nodes involved) or stage shifts. In this situation, for navigation to demonstrate a benefit, the intervention would have to lead to an early stage diagnosis in a patient who would otherwise have been lost to follow-up and only presented clinically at more advanced stages. Less dramatic within-stage shifts (eg, early in the course of local disease vs later in local disease, but before transition to regional spread) are also likely to improve survival, but to our knowledge there are only limited primary data on which to model these effects. It is also possible that small within-stage shifts do not affect cancer-specific mortality. We will use sensitivity analysis to evaluate how different assumptions regarding stage shift or cure affect results. If navigation is not cost-effective under the most favorable assumptions concerning small effects, then one could conclude that the investment does not yield a return on investments in QALYs. However, if programs would be considered cost-effective under assumptions that are clinically reasonable, then programs with small effects could be considered to have the potential to be cost-effective.

The relation between the intervention (navigation) and the endpoints (survival, QALYs) may not be straightforward, because the intermediate outcome of navigation—adherence to timely diagnostic services (in which the majority does not have cancer) and to recommended therapy—will not necessarily be uniform and linear in its relation to endpoints. We address this issue with simulation modeling and sensitivity analysis, the latter evaluating how changes in the association between specific input parameters (eg, expenditures on navigation services and adherence to screening recommendations over time) influence long-term outcomes.

Even if navigation interventions do not improve survival, they still may improve an individual's quality of life. In cost-effectiveness analyses, these effects are recorded as health state utilities to be used in computing QALYs. Utilities are measures of health state preference, measured on a scale from 0 (death) to 1 (ideal health). QALYs are a summary measure of survival weighted by utilities over the period after the intervention.[29] Utility weights for navigator program participants and a comparator group can be measured using a generic multiattribute utility instrument such as the EQ-5D.[30] Multiattribute utility instruments are questionnaires filled out by respondents assessing their quality of life across several domains. The individual responses are weighted using data derived from large population surveys on the utility of the different quality of life states. Scores are summed and converted to a 0 to 1 scale, with 0 representing the worst health imaginable (or death) and 1 representing perfect health. This approach provides societal rather than individual patient ratings of the potential quality of life improvements that might occur with navigation, so that results are generalizable.

Because of budget constraints, not all Patient Navigation Research Program sites will administer multiattribute utility instrument surveys to their participants. Utility weights for the comparison (no navigator) group will be based on the literature and, when available, surveys of low-income populations with cancer but no navigation services.[31] We will compare patient populations where utilities are being collected and those where they are not. In cases in which health and socioeconomic status are similar, we use data from the populations in which utilities are collected as proxies for those where utilities were not collected. We also explore the use of regression models

based on navigator study populations with utility data to impute utilities for those without utility data.

It should be noted that problems that are highly prevalent in underserved populations that are being targeted by navigation (such as low literacy rates and frequently changing residences) pose challenges to measuring outcomes after navigation using existing utility surveys. For example, populations with very low literacy or special groups such as the homeless or persons with mental illnesses may have great difficulty completing written questionnaires. The Patient Navigation Research Program address this issue[28] by allowing telephone and face-to-face interviews with patients and, if necessary, patient representatives.

Another issue that is embedded in the navigation program that poses a challenge to cost-effectiveness analysts is that patients with significant barriers to access to health systems often have complex social and health issues, such as poor educational attainment or non–cancer-related comorbidity, that themselves may influence long-term outcomes, such as life expectancy and/or cancer-specific survival rates after treatment.[32,33] Education, health status, and comorbidity are measured in the parent Patient Navigation Research Program study. In our projections of effects from the trial horizon to a lifetime horizon, we construct multivariate models with covariates to account for these characteristics to allow us to vary projected outcomes based on the characteristics of the cohort of interest; we can also use national data on the distribution of these factors to conduct sensitivity analyses to estimate the impact of navigation in broader settings and populations.

Navigator programs also aim to improve patient satisfaction and self-efficacy, and reduce the short-term distress associated with evaluation of an abnormal screening result. However, self-efficacy and satisfaction with care are generally not incorporated in surveys that measure utilities. In such situations, one could calculate a cost per unit decrease in distress.[34] However, to our knowledge, there are no established benchmarks for comparison to determine whether particular reductions in stress are cost-efficient compared with other ways to accomplish the same goal.

Navigation programs aimed at cancer patients may also have goals such as informed use of procedures based on patient preference (eg, lumpectomy vs mastectomy) or completion rates of planned therapy. These measures of outcome, as well as distress and other outcomes (eg, stage at diagnosis, time to diagnostic resolution, and satisfaction), can be summarized using cost-consequence analysis.[35] Cost-consequence analyses summarize program costs and effects in tabular fashion (Table 3). For example, one can evaluate the costs per patient of timely diagnostic resolution for the navigator program versus usual care. Cost-consequence analysis can be useful to decision makers who use components of cost-effectiveness analysis rather than the cost per QALY ratio.[36]

Interpersonal styles and commitments of navigators may influence the outcomes of particular programs. Although this factor is very difficult to measure and account for across sites, we will evaluate variations in sensitivity analysis, using proxy measures such as volume-outcome correlations (eg, volume of patients seen and adherence to follow-up of abnormal mammograms) and sociodemographics of the navigators themselves (age, sex, education).

Cost Impact of Navigation Programs

Navigation program costs include allocated fixed (eg, office space, proportional allocation of supervisory personnel, new equipment or contracts initiated for the program) and variable (eg, navigator time and transportation costs, direct medical care) components (Table 3). We denote the sum of allocated fixed and variable costs as $C_{navigator(program)}$. There are also costs associated with training navigators, including replacements or additional navigators as needed ($C_{training[program]}$). We denote the total direct medical care cost of diagnostic services and treatments received for persons using navigation programs as $C_{medical(program)}$. Patients who receive care without using navigator services have a cost, denoted $C_{medical(usual\ care)}$.

Patients and their caregivers incur nonmedical costs when seeking care, such as transportation costs, time costs related to testing and treatment, and time lost from work. We denote related nonmedical patient costs for those receiving and not receiving navigator services as $C_{nonmed(program)}$ and $C_{nonmed(usual\ care)}$. Note that in the short run, medical and related nonmedical costs are likely to be higher for the navigation program because of improvements in patient access to care and adherence to protocols

Table 3. Cost Consequence Analysis Sample Table, With Specific Elements of Interest in Navigator Interventions

Costs
Training costs ($C_{training[program]}$)
Initial training
Training replacements and additional navigators
Navigation program ($C_{navigator[program]}$)
Fixed costs: navigator program
Costs associated with developing navigator-related materials (eg, pamphlets, telephone scripts)
Allocated fixed operation costs (office space leasing, telephone, furniture, etc)
Variable costs: navigator program
Time spent in navigation (travel, meeting with patients, documentation)
Travel-associated costs
Variable direct nonmedical costs: all patients ($C_{nonmed[program]}$ and $C_{nonmed[usual\ care]}$)
Patient time costs seeking treatment
Travel-associated costs
Variable direct medical costs: patients ($C_{medical[program]}$ and $C_{medical[usual\ care]}$)
Outcomes
Time from abnormal screening test or suspicious finding to diagnosis
Time from diagnosis to initial therapy
Time from initial therapy to resolution (end of initial therapy including therapeutic combinations such as surgery plus chemotherapy)
Percentage of patients receiving initial therapy (surgery, chemotherapy, radiotherapy)
Percentage completing therapy
Satisfaction with care
Quality of life during care
Quality of life after care
Survival (years of life)*
Quality-adjusted survival (QALY)*

QALY indicates quality-adjusted life-year.
* Modeled.

for care. Longer-term costs for the navigation program may be lower if a program results in diagnostic resolution at an earlier stage based on an abnormal screening test, because patients lost to follow-up are likely to present again with more advanced, more time-consuming (and costly) stages of disease. Navigation may also lower costs if patients use care more appropriately and efficiently or better adhere to planned therapy such that cancer recurrence rates fall. Thus, in the long run, the net cost of navigation programs can be more or less than those under usual care.

One of the potential cost offsets of a navigator program is decreasing the time required by the medical staff and office support staff in trying to support patients who need help through the complex medical system. Because of the heterogeneity of care settings involved, it is not possible to track these offsets directly. We will explore the impact of offsets, based on time navigators spend with patients, in sensitivity analyses.

Direct medical care related to navigation (eg, screening tests and care related to follow-up of abnormal tests) will be assessed based on the routine core data elements collected by the Patient Navigation Research Program and valued using representative reimbursement rates, such as regionally adjusted Medicare payments. Longer-term costs, such as lifetime costs related to cancer treatment, will be estimated based on the stage at diagnosis, using published sources.[37] Navigators' time costs are likely to be the most significant program cost. Time costs will vary substantially depending on training (eg, professionals vs laypersons), the complexity of the care system, and the needs of the target population. Time spent by volunteer navigators is not free and should be valued as the opportunity cost of those persons, given other options for spending their time. Time costs for professionals can be valued based on their wages. Valuing time costs for volunteers can be more difficult. For persons who are employed, time is typically valued based on their wages or the prevailing national wage rates for those of the individual"s age and sex. For those who do not work for pay (eg, homemakers or retired persons), there is no generally agreed on method, but most base costs on national wage surveys.[27] By using navigator logs, the Patient Navigation Research Program will collect self-reported information on the time spent by navigators in direct contact with patients and in activities required for coordination of care.

In the process of seeking care, patients incur costs that may be significant barriers to accessing care in the first place.[38] Patient costs can be evaluated using patient logs or, if this is infeasible, by estimating time and associated expenses when traveling to specific services. Although the Patient Navigation Research Program will not collect patient log data, navigator logs will include information on the provision of these patient services, including transportation and child care costs. Patient time costs will be valued using census region-specific wage rates for individuals that match the age and sex of the patient population.

It is important to separate research-related costs from intervention costs. For the Patient Navigation Research Program evaluations, research costs will be identified from audits of research budgets during site visits

with investigators (eg, navigator time filling out study-related paperwork and complying with institutional review board documentation). In practice, it can be difficult to separate research from intervention costs, thus necessitating the documentation and reporting of assumptions made when there is uncertainty.

In cases in which navigation influences the use of multiple cancer screening programs, we will disaggregate costs to particular services (eg, mammography) based on the patient and navigator diaries. If feasible, we will also estimate the cost-effectiveness of a bundle of services (eg, mammography + Papanicolaou smear + colorectal cancer screening) compared with usual care.

Perspective and Time Horizon

In cost-effectiveness analysis, perspective refers to the point of view taken for evaluating the impacts and costs of the study. The societal perspective is favored for cost-effectiveness analysis in which public health issues are under evaluation,[27] and is particularly important for navigation programs, because the resources for navigators may come from 1 source (eg, foundations, government programs, hospitals), whereas payment for medical care may come from another (eg, Medicaid). As discussed above, navigation programs have short-term and long-term impacts. Thus, the cost-effectiveness of navigation programs is best estimated over the entire period that the program is expected to influence costs and outcomes. The relevant time horizon for navigation programs that assist patients with evaluation of abnormal findings is the time from the initial point of detection of abnormal findings to their resolution. For navigation programs that change care such that longer-term endpoints are affected (eg, survival), this implies using a lifetime time horizon. Because the Patient Navigation Research Program will only observe participants over a 4-year to 5-year horizon, evaluating cost-effectiveness will require simulation modeling to estimate the lifetime impact of navigation on populations.

Uncertainty Analysis

One-way sensitivity and multiway uncertainty analyses can identify factors that most substantially influence the cost-effectiveness of the programs.[39] One-way sensitivity analysis is a process of varying individual parameters across a range, then recalculating the cost-effectiveness ratio. This gives a sense of the relative influence of individual factors (eg, the hourly wage of navigators) on the overall cost-effectiveness of the program. Multiway analysis is a process of varying all parameters simultaneously such that a distribution or confidence interval can be derived around the point estimate of cost-effectiveness.

Particular attention should be paid to the impact of various assumptions regarding costs, quality of life, and survival for the usual care (non-navigator) group. The comparison or usual care group in some Patient Navigation Research Program studies uses historical data from the period before navigation or convenience samples from comparable communities that are not involved in the Patient Navigation Research Program; to the best of our knowledge, few use randomized controlled trials (Table 1). Navigator program-specific factors that should be considered for sensitivity analyses include patient time, type of navigator used, ranges of time to navigate different subgroups of patients, and the basis for time costs (eg, local vs national, average or race-specific wages).

Conclusions

It is rare for an economic evaluation to be free of conceptual and/or practical challenges, and cost-effectiveness analysis of cancer patient navigation is no exception. In this report, we outline several special conceptual challenges to evaluating navigation interventions, as well as many practical issues of data collection, instrument choice, and cost measurement. We have outlined several issues related to assessing costs and effectiveness in navigation programs, as well as methods Patient Navigation Research Program investigators will take to identify them. Although it is possible to derive nationally representative estimates of cost-effectiveness for particular programs, many navigation programs are tailored to specific local situations, and thus also merit evaluation of economic value in a local context. However, we do not know if navigation will translate into improved cancer survival, and if it will improve the effectiveness of cancer care at a reasonable cost (ie, be cost-effective).[40-43] Thus, the process of defining processes, costs, and outcomes that is part and parcel of cost-effectiveness analysis can also provide valuable information for local decision makers allocating limited health resources to navigation programs.

Conflict of Interest Disclosures

Funding was provided by the National Cancer Institute, via the Center to Reduce Cancer Health Disparities, through Contract 263-FQ-612391; by National Institutes of Health Grants U01 CA116892, U01 CA117281, U01 CA116903, U01 CA116937, U01 CA116924, U01 CA116885, U01 CA116875, and U01 CA116925; and by American Cancer Society Grant SIRSG-05-253-01.

Dr. Mandelblatt is also supported in part by National Cancer Institute grants U01 CA88283 and KO5 CA96940.

References

1. Strzelczyk JJ, Dignan MB. Disparities in adherence to recommended followup on screening mammography: interaction of sociodemographic factors. *Ethn Dis.* 2002;12:77-86.
2. Gwyn K, Bondy ML, Cohen DS, et al. Racial differences in diagnosis, treatment, and clinical delays in a population-based study of patients with newly diagnosed breast carcinoma. *Cancer.* 2004;100:1595-1604.
3. Center to Reduce Cancer Health Disparities. Economic Costs of Cancer Health Disparities. Bethesda, Md: National Cancer Institute, National Institute of Health; 2007.
4. Lurie JD, Welch HG. Diagnostic testing following fecal occult blood screening in the elderly. *J Natl Cancer Inst.* 1999;91:1641-1646.
5. Baig N, Myers RE, Turner BJ, et al. Physician-reported reasons for limited follow-up of patients with a positive fecal occult blood test screening result. *Am J Gastroenterol.* 2003;98:2078-2081.
6. Bickell NA, Wang JJ, Oluwole S, et al. Missed opportunities: racial disparities in adjuvant breast cancer treatment. *J Clin Oncol.* 2006;24:1357-1362.
7. Griggs JJ, Sorbero ME, Stark AT, Heininger SE, Dick AW. Racial disparity in the dose and dose intensity of breast cancer adjuvant chemotherapy. *Breast Cancer Res Treat.* 2003;81:21-31.
8. Shavers VL, Harlan LC, Stevens JL. Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35. *Cancer.* 2003;97:134-147.
9. Freeman HP, Muth BJ, Kerner JF. Expanding access to cancer screening and clinical follow-up among the medically underserved. *Cancer Pract.* 1995;3:19-30.
10. Center to Reduce Cancer Health Disparities. Patient Navigation Research Program (PNRP) Overview. 2007. Available at: http://crchd.cancer.gov/pnp/pnrp-overview.html Accessed April 11, 2007.
11. Psooy BJ, Schreuer D, Borgaonkar J, Caines JS. Patient navigation: improving timeliness in the diagnosis of breast abnormalities. *Can Assoc Radiol J.* 2004;55:145-150.
12. Farber JM, Deschamps M, Cameron R. Investigation and Assessment of the "Navigator Role" in Meeting the Information, Decisional and Educational Needs of Women with Breast Cancer in Canada. Ottawa, Ontario, Canada: Canadian Breast Cancer Initiative, Health Canada, 2002.
13. Battaglia TA, Roloff K, Posner MA, Freund KM. Improving follow-up to abnormal breast cancer screening in an urban population. A patient navigation intervention. *Cancer.* 2007;109(2 suppl):359-367.
14. Wells KJ, Battaglia TA, Dudley DJ, et al. Patient navigation: state of the art or is it science? *Cancer.* 2008;113:1999-2010.
15. Freeman HP. Patient navigation: a community centered approach to reducing cancer mortality. *J Cancer Educ.* 2006;21(1 suppl):S11-S14.
16. Freeman HP, Wasfie TJ. Cancer of the breast in poor black women. *Cancer.* 1989;63:2562-2569.
17. Oluwole SF, Ali AO, Adu A, et al. Impact of a cancer screening program on breast cancer stage at diagnosis in a medically underserved urban community. *J Am Coll Surg.* 2003;196:180-188.
18. Centers for Medicare and Medicaid Services, the Schneider Institute for Health Policy, and the Heller School for Social Policy and Management. Evidence Report and Evidence Based Recommendations: Cancer Prevention and Treatment Demonstration for Ethnic and Racial Minorities. Waltham, Mass: Brandeis University; 2003.
19. Smedley BD, Stith AY, Nelson AR. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC: National Academy Press; 2002.
20. Weinrich SP, Boyd MD, Weinrich M, Greene F, Reynolds WA Jr, Metlin C. Increasing prostate cancer screening in African American men with peer-educator and client-navigator interventions. *J Cancer Educ.* 1998;13:213-219.
21. Freeman HP, Chu KC. Determinants of cancer disparities: barriers to cancer screening, diagnosis, and treatment. *Surg Oncol Clin N Am.* 2005;14:655-669.
22. Dohan D, Schrag D. Using navigators to improve care of underserved patients: current practices and approaches. *Cancer.* 2005;104:848-855.
23. Christie J, Itzkowitz S, Lihau-Nkanza I, Castillo A, Redd W, Jandorf L. A randomized controlled trial using patient navigation to increase colonoscopy screening among low-income minorities. *J Natl Med Assoc.* 2008;100:278-284.
24. Ferrante JM, Chen PH, Kim S. The effect of patient navigation on time to diagnosis, anxiety, and satisfaction in urban minority women with abnormal mammograms: a randomized controlled trial. *J Urban Health.* 2008;85:114-124.
25. Ell K, Vourlekis B, Lee PJ, Xie B. Patient navigation and case management following an abnormal mammogram: a randomized clinical trial. *Prev Med.* 2007;44:26-33.
26. Braithwaite RS, Meltzer DO, King JT Jr, Leslie D, Roberts MS. What does the value of modern medicine say about the $50,000 per quality-adjusted life-year decision rule? *Med Care.* 2008;46:349-356.
27. Gold MR, Siegal JE, Russel LB, Weinstein MC. Cost-Effectiveness in Health and Medicine. New York, NY: Oxford University Press; 1996.

28. Freund KM, Battaglia TA, Calhoun E, et al. National Cancer Institute Patient Navigation Research Program: methods, protocol, and measures. *Cancer.* 2008;113:3391-3399.

29. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA.* 1996;276:1253-1258.

30. The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy.* 1990;16:199-208.

31. Earle CC, Chapman RH, Baker CS, et al. Systematic overview of cost-utility assessments in oncology. *J Clin Oncol.* 2000;18:3302-3317.

32. Du XL, Fang S, Coker AL, et al. Racial disparity and socioeconomic status in association with survival in older men with local/regional stage prostate carcinoma: findings from a large community-based cohort. *Cancer.* 2006;106:1276-1285.

33. Ward E, Jemal A, Cokkinides V, et al. Cancer disparities by race/ethnicity and socioeconomic status. *CA Cancer J Clin.* 2004;54:78-93.

34. Mandelblatt JS, Cullen J, Lawrence WF, et al. Economic evaluation alongside a clinical trial of psycho-educational interventions to improve adjustment to survivorship among patients with breast cancer. *J Clin Oncol.* 2008;26:1684-1690.

35. Polsky D, Mandelblatt JS, Weeks JC, et al. Economic evaluation of breast cancer treatment: considering the value of patient choice. *J Clin Oncol.* 2003;21:1139-1146.

36. Mauskopf JA, Paul JE, Grant DM, Stergachis A. The role of cost-consequence analysis in healthcare decision-making. *Pharmacoeconomics.* 1998;13:277-288.

37. Yabroff KR, Lamont EB, Mariotto A, et al. Cost of care for elderly cancer patients in the United States. *J Natl Cancer Inst.* 2008;100:630-641.

38. Yabroff KR, Davis WW, Lamont EB, et al. Patient time costs associated with cancer care. *J Natl Cancer Inst.* 2007;99:14-23.

39. Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics.* 2000;17:479-500.

40. Mandelblatt J, Freeman H, Winczewski D, et al. The costs and effects of cervical and breast cancer screening in a public hospital emergency room. The Cancer Control Center of Harlem. *Am J Public Health.* 1997;87:1182-1189.

41. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR. Assessing the effectiveness of health interventions for cost-effectiveness analysis. Panel on Cost-Effectiveness in Health and Medicine. *J Gen Intern Med.* 1997;12:551-558.

42. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. *JAMA.* 1996;276:1172-1177.

43. Zhou X-HA, Ramsey SD. Assessing the equality of means of healthcare costs. *Expert Rev Pharmacoecon Outcomes Res.* 2003;3:561-567.

Viral RNA Testing in Hepatitis C Antibody–Positive Veterans

Catherine A. Rongey, MD, MSHS, Fasiha Kanwal, MD, MSHS, Tuyen Hoang, PhD, Allen L. Gifford, MD, Steven M. Asch, MD, MPH

Background: Chronic hepatitis C virus (HCV) infection affects approximately 1.3% of the U.S. population. As up to 30% of HCV-antibody (anti-HCV)–positive patients have negative HCV ribonucleic acid (RNA), indicating resolution of infection, VA (Veterans Affairs) guidelines recommend RNA testing on all anti-HCV–positive patients. As HCV RNA testing is a prequel to treatment, perceived eligibility for treatment may influence the decision to order an RNA test. This study was designed to determine the patient and healthcare facility factors associated with patient receipt of HCV RNA testing.

Methods: Two logistic regression analyses were conducted in anti-HCV–positive patients, including the entire sample and then on a subsample excluding sites with routine HCV RNA testing policies, using data stored in the VA Southern California Network data warehouse. Significant patient- and site-level predictors of patient receipt of HCV RNA testing were determined.

Results: Of the 13,257 antibody-positive patients, 76% received HCV RNA testing. Excluding sites with routine HCV RNA testing, patients aged >65 years (RR=0.79) and illicit drug users (RR=0.94) were significantly less likely to receive HCV RNA testing. Patients with abnormal transaminases (RR=1.14), presence of non-HCV hepatitis (RR=1.08), or decompensated liver disease (RR=1.22) were significantly more likely to receive HCV RNA testing.

Conclusions: Without policies for routine RNA testing, patients with hepatitic C who either are aged >65 years or are illicit drug users are less likely to be tested. Also, patient receipt of RNA testing becomes dependent on clinical cues of hepatic decompensation or inflammation. The results support the implementation of routine RNA testing for anti-HCV–positive patients.

(Am J Prev Med 2009;36(3):235–238)

Introduction

Chronic hepatitis C virus (HCV) infection is prevalent and expensive, affecting more than 1.3% of the U.S. and 5.4% of the Veteran Affairs (VA) population.[1,2] In clinical practice, the first step in HCV diagnosis is to test patients for HCV antibodies (anti-HCV). However, as up to 30% of anti-HCV–positive patients have negative HCV ribonucleic acid (RNA), indicating resolution of HCV infection, VA guidelines recommend ribonucleic acid (RNA) testing in all anti-HCV–positive patients.[3] In addition to establishing viremia, HCV RNA status modifies clinical follow-up measures, such as hepatic function assessment and hepatocellular cancer screening, and health counseling, such as risk behavior reduction. Because HCV RNA testing is thought of as a prequel to treatment, perceived eligibility for treatment may influence the decision to order an RNA test.

Published studies within the VA population have examined the proportion and characteristics of HCV-infected veterans eligible for treatment.[4,5] This is the first study to examine the patient and center characteristics associated with the decision to initiate RNA testing. Administrative and clinical data were used from anti-HCV–positive patients within VA facilities in Southern California. Two had a policy to test routinely for HCV RNA after a positive antibody test. In sites without routine testing policies, it was anticipated that patients with illicit drug-use behaviors, psychiatric disorders, medical contraindications, or a normal hepatic profile

From the Robert Wood Johnson Clinical Scholars Program (Rongey), Center for Outcomes Research and Education (Kanwal), Department of Internal Medicine, David Geffen School of Medicine (Asch), University of California at Los Angeles; Veterans Affairs Greater Los Angeles Healthcare System (Rongey, Hoang, Asch); Division of Gastroenterology, University of Southern California (Rongey), Los Angeles, California; John Cochran VA Medical Center (Kanwal), Division of Gastroenterology, Saint Louis University (Kanwal), Saint Louis, Missouri; Department of Internal Medicine, Boston University (Gifford), Boston, Massachusetts

Address correspondence and reprint requests to: Catherine A. Rongey, MD, MSHS, Department of Medicine, GI/Liver Division, USC School of Medicine, 1200 N. State Street, LAC Room 12-137, Los Angeles CA 90033. E-mail: rongey@usc.edu.

0749-3797/09/$–see front matter
doi:10.1016/j.amepre.2008.10.013

would be less likely to receive testing. This study examines biases that may be eliminated by routine testing policies.

Methods

Data Source

Data were collected from the VA Southern California Network (VISN-22), which includes clinical and administrative medical record information from five VISN-22 centers (Los Angeles, Las Vegas, San Diego, Loma Linda, and Long Beach). The database includes patient demographics, outpatient and inpatient services utilization, vital signs, pharmacy utilization, and laboratory data.

Study Subjects

Patients defined as positive for HCV were those with one or more positive HCV-antibody tests between 10/02/2000 and 3/30/2006. A patient's treating facility was defined as the most frequently visited site during the 5-year study period, with at least two visits/year to the treating facility.

Outcome Variable

Outcome was receipt of an HCV RNA test as either a qualitative or quantitative polymerase chain reaction.

Predictors of HCV RNA Testing

Patient factors included sociodemographic characteristics (age, race/ethnicity, household income, marital status); healthcare utilization (number of clinic visits); HCV and liver characteristics (diagnosis of cirrhosis, alanine aminotransferase [ALT] level, other viral co-infectivity, portal hypertension); health behaviors (illicit drug and alcohol use); and comorbidities (psychiatric, cardiopulmonary, renal, endocrine, cancer, cytopenia). Table 1 includes a list of patient factors by category.

Absolute and relative treatment contraindications are based on guidelines from the American Association for the Study of Liver Diseases (AASLD) and the VA.[3,6] Treatment contraindications are defined as the presence of one inpatient or two outpatient diagnostic codes for renal failure, cardiac disease, cancer, organ transplant, autoimmune hepatitis, autoimmune conditions, cytopenias, chronic obstructive pulmonary disease, or HIV.

Demographic data from VISN do not include race/ethnicity; race/ethnicity was listed as "missing" for 40% in the primary analysis. A sensitivity analysis in the multivariate models was performed. The impact of RNA routine testing policies was evaluated at each of the five sites and adjusted for facility.

Table 1. Demographic and clinical characteristics of the study population, % (N/*n*) unless otherwise indicated

Variable	Total sample (N=13,257)	Sites without reflex testing policies (*n*=8686)	Sites with reflex testing policies (*n*=4571)
Receipt of RNA testing (%)	75.7 (10,038)	64.0 (5562)	97.9 (4476)
Age (M [SD])	57.7 (9.0)	57.5 (9.1)	58.0 (8.9)
Primary care visits per year (M [SD])	17.1 (19.8)	15.6 (18.8)	19.9 (21.4)
Mental health visits per year (M [SD])	4.5 (18.6)	3.3 (8.1)	6.7 (29.6)
Years in VA healthcare system (M [SD])	3.6 (1.8)	3.5 (1.8)	3.9 (1.7)
Male	97.0 (12,685)	97.0 (8422)	97.2 (4443)
Race/ethnicity			
White	29.6 (3918)	31.0 (2686)	27.0 (1232)
Hispanic	8.0 (1060)	7.5 (648)	9.0 (412)
Black	14.0 (1862)	10.5 (914)	20.7 (948)
Other	8.4 (1114)	8.6 (751)	7.9 (363)
Missing	40 (5303)	42.4 (3687)	35.4 (1616)
Diagnosis			
Medical contraindications	27.5 (3649)	27.8 (2411)	27.1 (1238)
Non-HCV hepatitis	6.4 (809)	5.2 (452)	8.7 (397)
Psychiatric comorbidity	46.5 (6302)	44.5 (3870)	53.2 (2432)
Illicit drug use	40.7 (5394)	38.4 (3333)	45.1 (2061)
Decompensated cirrhosis	3.7 (485)	3.9 (337)	3.2 (148)
Abnormal transaminases	76.4 (10,123)	78.0 (6773)	73.3 (3350)

HCV, hepatitis C virus; VA, Veterans Affairs

Statistical Analysis

Two logistic regression analyses were conducted including the entire sample and a subsample excluding sites with routine RNA testing. The unit of analysis was a patient with a positive HCV-antibody test. The dependent variable was HCV RNA testing, and the predictors were those mentioned above. Intrasite correlation was adjusted using the generalized estimating equation method. Risk estimation was expressed as relative risk.

Results

Of the 13,257 anti-HCV–positive patients in the total sample, 10,038 (76%) received RNA testing. Table 1 provides the descriptive analysis. The total sample mean age was 58 years; 97% were male, 30% were white, and 40% were African-American patients. Of the 8686 observations in the subsample excluding routine testing sites, 5562 (64%) received RNA testing. The two routine testing sites had HCV RNA testing of 92.5% and 99%, whereas sites without routine testing policies ranged between 47% and 76%. In adjusted analysis, routine testing sites are 1.5 times more likely to check HCV RNA than sites without policies.

Table 2 provides multivariate logistic regression analysis results. In the total sample, controlling for the presence of routine RNA testing, patients with decompensated cirrhosis (RR=1.14; 95% CI=1.1, 1.2), other viral hepatitis (RR=1.04; 95% CI=1.0, 1.07), or elevated transaminases (RR=1.08; 95% CI=1.05, 1.1) were significantly more likely to receive RNA testing. Conversely, patients aged >65 years (RR=0.89; 95% CI=0.8, 0.99) or who were illicit drug users (RR=0.98; 95% CI=0.96, 0.99) were

Table 2. Predictors of HCV RNA testing, RR (CI)[a]

Variable	Total sample	Sites without reflex testing policies	Sites with reflex testing policies
Male	1.04 (0.98, 1.1)	1.03 (0.9, 1.1)	1.02 (0.99, 1.1)
Age (years)			
40–50	1.04 (0.9, 1.1)	1.05 (0.9, 1.2)	1.03 (0.95, 1.1)
51–65	1.02 (0.92, 1.1)	1.04 (0.9, 1.2)	1.02 (0.95, 1.1)
>65	0.89 (0.8, 0.99)**	0.79 (0.69, 0.92)***	1.01 (0.99, 1.02)
Diagnosis			
Psychiatric comorbidity	1.03 (1, 1.05)***	1.03 (0.9, 1.07)*	1.01 (0.99, 1.03)
Illicit drug use	0.98 (0.96, 0.99)**	0.94 (0.91, 0.97)***	1.00 (0.99, 1.0)
Non-HCV viral hepatitis	1.04 (1, 1.07)***	1.08 (1.02, 1.14)***	1.00 (0.99, 1.0)
Decompensated cirrhosis	1.14 (1.1, 1.2)***	1.22 (1.1, 1.3)***	1.02 (1.0, 1.03)
Abnormal transaminases	1.08 (1.05, 1.1)***	1.14 (1.03, 1.20)***	1.01 (1.0, 1.02)
Race/ethnicity			
African American	1.02 (0.99, 1.05)*	1.04 (0.9, 1.1)*	1.01 (0.99, 1.0)
Hispanic	1.02 (0.99, 1.05)	1.04 (0.98, 1.1)	1.01 (1.0, 1.0)
Other	1.0 (0.97, 1.0)	1.0 (0.95, 1.07)	1.02 (0.99, 1.0)
Missing	1.03 (1.01, 1.06)***	1.1 (1.06, 1.15)***	0.99 (0.98, 1.0)
Routine testing policy	1.5 (1.48, 1.52)***		

[a]Controlling for gender, medical contraindications, number of primary care and mental health visits, and years in the VA healthcare system. For age and race/ethnicity, reference categories are aged <40 years and non-Hispanic Caucasian.
*$p \leq 0.10$; **$p \leq 0.05$; ***$p \leq 0.01$
HCV, hepatitis C virus; VA, Veterans Affairs

significantly less likely to receive HCV testing. The subsample excluding routine testing sites yielded the same significant variables but with stronger effect. Patients aged >65 years (RR=0.79; 95% CI=0.69, 0.92) or who were illicit drug users (RR=0.94; 95% CI=0.91, 0.97) were significantly less likely to receive RNA testing. Patients with non-HCV hepatitis (RR=1.07; 95% CI=1.02, 1.14), decompensated cirrhosis (RR=1.2; 95% CI=1.1, 1.3), or elevated transaminases (RR=1.1; 95% CI=1.03, 1.20) were more likely to receive RNA testing.

Discussion

The results show significant underutilization of RNA testing. Providers use treatment eligibility as a prompt for obtaining HCV RNA tests, whereas routine testing policies eliminate biases in RNA testing. In sites without routine testing, several groups of patients were less likely to receive RNA testing. Infection with HIV at age >50 years is associated with more rapid progression to cirrhosis,[7] but such patients are less likely to receive RNA testing. Active drug users, in whom reported HCV prevalence is 65%–95%,[8,9] were less likely to receive RNA testing. Individualized assessment of treatment eligibility in active users is recommended by AASLD.[6] There is increasing evidence that illicit drug users can comply with HCV treatment[10,11]; regardless, review of RNA test results provides an important counseling opportunity to remind patients that HCV is transmitted more efficiently than HIV.[12]

Providers relied on clinical cues such as elevated transaminases and hepatic decompensation to initiate RNA testing. Typically, HCV is an indolent disease with poor correlation between degree of transaminase elevation and liver injury.[13,14] The disease is more treatment-responsive prior to the development of fibrosis.[15] Treatment becomes precarious after hepatic decompensation[16,17] and would merely serve as a bridge to liver transplantation. Without clinical guidance, patients may not seek care until they have noticed clinical signs of decompensation.[18]

Regardless of a patient's appropriateness for therapy, RNA confirmation of infectivity presents an important patient educational and counseling opportunity.[19] Risk-reduction counseling is important in the predominantly male VA population, as there is some evidence that progression to cirrhosis is faster and that risk of hepatocellular carcinoma is higher in men.[20,21] Smoking,[22] alcohol consumption,[23,24] hepatitis B,[25] and HIV[26] co-infection increase the likelihood of progression to cirrhosis. The benefit of counseling to reduce the likelihood of progression to cirrhosis outweighs the additional RNA testing cost, as the cost of end-stage liver disease is estimated to be $10.7 billion by 2019.[7]

The results support the implementation of routine RNA testing in anti-HCV–positive patients. Historically, VA quality initiatives have benefited other healthcare centers.[27] Although this study focused on a VA population, it is anticipated that most medical centers do not conduct routine RNA testing in anti-HCV–positive patients.

Laboratory routine testing algorithms improve the efficiency of laboratory evaluation in other condtions.[28] In the absence of such policies, many patients will miss the benefits of knowing how their RNA status should affect their care, whether that care be pharmacologic measures, clinical surveillance, or behavior modification counseling.

Dr. Rongey is supported by the Robert Wood Johnson Clinical Scholars and Foundation. This research was also supported by the VA HIV–Hepatitis Quality Enhancement Research Initiative.

No financial disclosures were reported by the authors of this paper.

References

1. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. Ann Intern Med 2006;144:705–14.
2. Dominitz JA, Boyko EJ, Koepsell TD, et al. Elevated prevalence of hepatitis C infection in users of United States veterans medical centers. Hepatology 2005;41:88–96.
3. Yee HS, Currie SL, Darling JM, Wright TL. Management and treatment of hepatitis C viral infection: recommendations from the Department of Veterans Affairs Hepatitis C Resource Center program and the National Hepatitis C Program office. Am J Gastroenterol 2006;101:2360–78.
4. Bini EJ, Brau N, Currie S, et al. Prospective multicenter study of eligibility for antiviral therapy among 4,084 U.S. veterans with chronic hepatitis C virus infection. Am J Gastroenterol 2005;100:1772–9.
5. Muir AJ, Provenzale D. A descriptive evaluation of eligibility for therapy among veterans with chronic hepatitis C virus infection. J Clin Gastroenterol 2002;34:268–71.
6. Strader DB, Wright T, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C. Hepatology 2004;39:1147–71.
7. Wong JB, McQuillan GM, McHutchison JG, Poynard T. Estimating future hepatitis C morbidity, mortality, and costs in the United States. Am J Public Health 2000;90:1562–9.
8. Hagan H, Thiede H, Des Jarlais DC. Hepatitis C virus infection among injection drug users: survival analysis of time to seroconversion. Epidemiology 2004;15:543–9.
9. Alter MJ, Moyer LA. The importance of preventing hepatitis C virus infection among injection drug users in the United States. J Acquir Immune Defic Syndr Hum Retrovirol 1998;18S:S6–S10.
10. Neri S, Bruno CM, Abate G, et al. Controlled clinical trial to assess the response of recent heroin abusers with chronic hepatitis C virus infection to treatment with interferon alpha-n2b. Clin Ther 2002;24:1627–35.
11. Sylvestre D. Hepatitis C treatment in drug users: perception versus evidence. Eur J Gastroenterol Hepatol 2006;18:129–30.
12. Hagan H, Thiede H, Weiss NS, Hopkins SG, Duchin JS, Alexander ER. Sharing of drug preparation equipment as a risk factor for hepatitis C. Am J Public Health 2001;91:42–6.
13. Okanoue T, Makiyama A, Nakayama M, et al. A follow-up study to determine the value of liver biopsy and need for antiviral therapy for hepatitis C virus carriers with persistently normal serum aminotransferase. J Hepatol 2005;43:599–605.
14. Pradhan MM, Horswell R, Jones G, Ramsey JL, Cassidy W. Evaluation of The Federal Bureau of Prisons protocol for selection of which hepatitis C-infected inmates are considered for treatment. Dig Dis Sci 2005; 50:714–8.
15. De Luca A, Di Giambenedetto S, Cingolani A, et al. Liver fibrosis stage predicts early treatment outcomes with peginterferon plus ribavirin in HIV/hepatitis C virus co-infected patients. AIDS 2004;18:1602–4.
16. Hoofnagle JH, Di Bisceglie AM, Waggoner JG, Park Y. Interferon alfa for patients with clinically apparent cirrhosis due to chronic hepatitis B. Gastroenterology 1993;104:1116–21.
17. Crippin JS, McCashland T, Terrault N, Sheiner P, Charlton MR. A pilot study of the tolerability and efficacy of antiviral therapy in hepatitis C virus-infected patients awaiting liver transplantation. Liver Transpl 2002; 8:350–5.
18. Khaw FM, Stobbart L, Murtagh MJ. 'I just keep thinking I haven't got it because I'm not yellow': a qualitative study of the factors that influence the uptake of hepatitis C testing by prisoners. BMC Public Health 2007;7:98.
19. Arora S, O'Brien C, Zeuzem S, et al. Treatment of chronic hepatitis C patients with persistently normal alanine aminotransferase levels with the combination of peginterferon alpha-2a (40 kDa) plus ribavirin: impact on health-related quality of life. J Gastroenterol Hepatol 2006;21:406–12.
20. Wright M, Goldin R, Fabre A, et al. Measurement and determinants of the natural history of liver fibrosis in hepatitis C virus infection: a cross sectional and longitudinal study. Gut 2003;52:574–9.
21. Degos F, Christidis C, Ganne-Carrie N, et al. Hepatitis C virus related cirrhosis: time to occurrence of hepatocellular carcinoma and death. Gut 2000;47:131–6.
22. Kermit K, Devine DA, Tatman SM. High risk model of threat perception in chronic pain patients: implications for primary care and chronic pain programs. J Nerv Ment Dis 2000;188:577–82.
23. Harris DR, Gonin R, Alter HJ, et al. The relationship of acute transfusion-associated hepatitis to the development of cirrhosis in the presence of alcohol abuse. Ann Intern Med 2001;134:120–4.
24. Strader DB, Seeff LB. Hepatitis C: a brief clinical overview. Ilar J 2001;42:107–16.
25. Gaeta GB, Stornaiuolo G, Precone DF, et al. Epidemiological and clinical burden of chronic hepatitis B virus/hepatitis C virus infection. A multicenter Italian study. J Hepatol 2003;39:1036–41.
26. Graham CS, Baden LR, Yu E, et al. Influence of human immunodeficiency virus infection on the course of hepatitis C virus infection: a meta-analysis. Clin Infect Dis 2001;33:562–9.
27. Stetler CB, Mittman BS, Francis J. Overview of the VA Quality Enhancement Research Initiative (QUERI) and QUERI theme articles: QUERI Series. Implement Sci 2008;3:8.
28. Wu AH, Ghani F, Prigent F, Petry C, Armstrong G, Graff L. Reflex testing II: evaluation of an algorithm for use of cardiac markers in the assessment of emergency department patients with chest pain. Clin Chim Acta 1999;288(1–2):97–109.

Performance Improvement

Measuring Quality of Oral Anticoagulation Care: Extending Quality Measurement to a New Field

Adam J. Rose, M.D., M.Sc.; Dan R. Berlowitz, M.D.; M.P.H.; Susan M. Frayne, M.D., M.P.H.; Elaine M. Hylek, M.D., M.P.H.

Quality measurement efforts have grown in sophistication and impact during the past three decades. However, some areas of clinical practice remain relatively unaffected by quality measurement. In some cases, this is because it would be difficult to measure quality for uncommon conditions, when clinicians or even whole medical centers may manage only a few patients per year.[1] In other cases, quality measurement may not seem warranted for conditions that seem relatively benign and self-limited, although it could be argued that any condition with sufficient expenditures deserves a program of rigorous quality measurement. There are still conditions, however, that are both common and serious for which quality measurement has not yet become a reality—it should be a high priority to institute programs of quality measurement for these conditions.

Oral anticoagulation with warfarin is an increasingly common medical intervention in the United States: The number of outpatient prescriptions for warfarin increased from 21.1 million in 1998 to 30.6 million in 2004.[2] The increasing use of warfarin is due to two factors. First, warfarin is highly efficacious for such relatively common conditions as atrial fibrillation,[3–9] venous thromboembolism,[10–14] and valvular heart disease.[15] Second, as the population ages, indications for anticoagulation such as atrial fibrillation are increasing in prevalence.[16]

Despite its efficacy, warfarin is notoriously difficult to manage: Its therapeutic window is narrow,[17,18] it has significant interactions with diet and other medications,[19,20] and it its action is affected by comorbid conditions and other inherent patient characteristics.[19] The difficulty of managing warfarin contributes to great potential for patient harm,[21] both from excessive anticoagulation[22–24] and insufficient anticoagulation (which can allow the occurrence of thromboembolic events despite warfarin therapy).[18,23,24] Therefore, it is a major patient safety goal to improve the quality of oral anticoagulation care. Many have hoped that novel anticoagulants, which would be easier to use than warfarin, would themselves improve quality and outcomes in oral anticoagulation care.[25–27] This may even-

Article-at-a-Glance

Background: Oral anticoagulation with warfarin is an increasingly common medical intervention. Despite its efficacy, warfarin is difficult to manage, contributing to potential for patient harm. Efforts to measure the quality of oral anticoagulation care have focused disproportionately on the identification of ideal candidates for warfarin therapy, with comparatively little effort in measuring the quality of oral anticoagulation care once therapy has begun. To address this gap in the literature, a MEDLINE search was conducted for all papers relevant to possible quality measures in oral anticoagulation care, including measures of structure, process, and outcomes of care.

Limitations, Concerns, and Challenges of Quality Measurement in Oral Anticoagulation: Because they do not have intrinsic significance, measures of structure and process should be strongly related to outcomes that matter to merit our interest. Consensus guidelines may provide useful guidance to practicing clinicians but may not represent valid process measures. Outcome measures must be studied with databases that provide sufficient statistical power to reliably demonstrate real differences between providers or sites of care.

Conclusion: Oral anticoagulation care, a common and serious condition, is in need of a program of quality measurement. This article suggests a research agenda to begin such a program. Previous research has established the evidence for anticoagulant therapy across a broad spectrum of indications and has helped to achieve consensus on the optimal target intensity for various indications. The next task will be to use this body of evidence to develop valid measures of the structure, process, and outcomes of oral anticoagulation care. Quality indicators provide a framework for quality improvement, two goals of which are to maximize the effectiveness of therapy and to minimize harm.

tually happen, but we do not know when such agents will be clinically available, and when they are introduced, their use may initially be limited to certain groups of patients in which they were initially studied. In addition, doubts about the safety[28] and cost-effectiveness[29] of such novel agents may slow their adoption into routine clinical practice. Therefore, it remains worthwhile to improve the quality of oral anticoagulation care as it currently exists, that is, with warfarin. Previous efforts in measuring the quality of oral anticoagulation care have focused disproportionately on the identification of ideal candidates for warfarin therapy, and comparatively little effort has gone into measuring the quality of oral anticoagulation care once therapy has begun.

To address this gap in the literature, we searched MEDLINE for the intersection of "warfarin OR anticoagulants" and "quality of care." We performed an exhaustive search for other papers using the "find related articles" function of PubMed and the bibliographies of all articles retrieved. Through further searches, we specifically located all papers relevant to possible quality measures in oral anticoagulation care, including measures of structure, process, and outcomes of care.[30] We completed our first search in March 2008 and repeated our search again in December 2008 prior to the final acceptance of this article.

In this narrative review, we discuss candidate quality measures for anticoagulation care and what is already known about them. We sequentially discuss quality measures in the three domains of quality first proposed by Donabedian: structure, process, and outcomes.[30] We then detail a research agenda to advance the understanding of how to measure the quality of care in oral anticoagulation. This narrative review can also serve as a general example of how to conceptualize a program of quality measurement for any field that does not yet have one.

Developing an Evidence Base in Oral Anticoagulation Care

A fundamental step in quality of care research in anticoagulation was establishing that anticoagulation does in fact prevent thromboembolism. During the last 50 years, the efficacy of warfarin in the prevention and treatment of thromboembolic disease has been firmly established, especially for atrial fibrillation,[9] venous thromboembolism,[12] and valvular heart disease.[15] The exact parameters of these indications are still under investigation in some cases. For example, research is ongoing to define the optimal duration of therapy for venous thromboembolism[31–33] and to define patients for whom the benefits of anticoagulation for atrial fibrillation are likely to exceed the risks.[34–38]

Along with the establishment of the major indications for long-term anticoagulation, a parallel research agenda aimed to determine the optimal degree of anticoagulation for each indication. A necessary first step was the development of a standardized test to measure the degree of anticoagulation. The prothrombin time was used to monitor anticoagulation as early as the 1950s[39] but was plagued by a lack of standardization between laboratories.[40] This issue was resolved through the development and widespread adoption of the International Normalized Ratio (INR), which allows results from different laboratories to be comparable.[40] Having agreed on a way to measure anticoagulation, investigators were now able to achieve consensus regarding optimal INR target ranges for various indications.[9,12,15]

Achieving consensus about indications for long-term anticoagulation, a method for measuring the degree of anticoagulation, and INR targets for different indications have been major research accomplishments. The existence of such high-quality evidence produces new obligations to use it to guide treatment decisions. Quality indicators can help us to quantify the degree to which we are succeeding in this endeavor.

What Is a Quality Indicator, and How Might It Be Used?

Ideally, a quality indicator will either be an outcome that is relatively common or an aspect of structure or process that has been linked to outcomes.[41] Quality indicators can be used to profile and compare performance between different providers or sites of care.[42] In addition to profiling process or outcome measures, it is useful to study sites that are performing best or worst on a measure ("outliers") to determine which factors may be contributing to their performance. The results of such study can be used to define and disseminate "best practices" that may improve care through widespread implementation. Finally, there has recently been great interest in tying reimbursement levels to improvements in process and outcomes of care ("pay for performance")[43–47] to provide an additional incentive for providers and sites of care to pursue quality improvement.

In addition to profiling performance on a provider or site level, quality indicators can also be used to examine disparities in health care.[48,49] Disparities in the structure, process, and outcomes of care based on race, socioeconomic status, mental illness, and other patient characteristics have been documented in almost every conceivable area of medical practice.[49–51] The reduction and elimination of such disparities is an important goal, and one that will not necessarily be fulfilled through general quality improvement alone.[49–52] Valid systems of quality

measurement are necessary to monitor, and hopefully reduce and eliminate, such disparities over time.[53]

Many candidate measures in oral anticoagulation care have already been used extensively in research, but have not yet been used as quality indicators. The review will now discuss possible quality indicators for oral anticoagulation care. Each section will begin with what is known about some candidate measures and will finish by proposing a research agenda to advance knowledge in that area.

Structural Measures of Quality

Structural measures of quality identify health care delivery systems that represent best practices, ideally ones that have been linked to improvements in the process and outcomes of care.[30,41] For anticoagulation care, structural quality measures that have been examined include different systems of management and the presence or absence of computerized decision support.

Oral anticoagulation can be managed in one of three settings: physician's offices ("usual care"), dedicated anticoagulation clinics (ACCs), and patient self-testing and self-management (PST/PSM). If one of these settings were clearly superior to the others, then the proportion of patients receiving care in that setting could be used as a quality indicator. Although observational studies have suggested that ACCs may improve INR control and decrease complications compared with usual care,[54,55] the few randomized trials that have been undertaken have not been able to show a clear difference between settings, possibly due to a lack of statistical power.[56,57] Several studies have suggested that PST/PSM can produce improved INR control compared with routine care[58,59] or ACC,[60–62] whereas another study found it to be at least as good as ACC.[63] In one study, there were fewer major complications of therapy in the PST/PSM group compared with ACC[60]; in no study was the reverse effect seen.

Although we may be moving toward a consensus that ACC and PST/PSM represent an improvement over usual care, we are in need of structural measures that might be used to compare ACCs with one another. For example, nursing ratios are known to affect outcomes in many domains of care,[64] but it is not yet known whether the ratio of staff in an ACC to the patient load has an impact on outcomes of care. Similarly, studies of care for other conditions have shown that high-volume centers produce superior outcomes.[65,66] Do higher-volume ACCs similarly achieve better outcomes than lower-volume ACCs? Finally, there may be differences between ACCs in staffing, leadership, or organizational structure. It is not known whether such factors have measurable consequences in terms of outcomes or whether they have utility as structural quality indicators.

Another body of literature has examined a different structure of care issue: anticoagulation management software. Studies have consistently shown that managing warfarin doses via standardized computer-based algorithms results in improved INR control and stability compared with management without computer assistance.[67–71] The research findings in this area are strong enough that the use of such software can already be considered a quality indicator, even if no further work is done.

Process Measures of Quality

Process measures of quality assess whether optimal care is provided to the correct patient in a timely fashion.[30,41] Ideally, improvements in process measures can be linked to improved outcomes of care.[30,41] Many studies of process measures in oral anticoagulation have focused on atrial fibrillation, the most common indication for long-term anticoagulation. Two relatively recent studies, both of which represent great advances in the general science of process measurement, nevertheless illustrate the limitations of previous process measures in oral anticoagulation.

In the Community Quality Index (CQI) Study[72] (Table 1, page 149), anticoagulation-related indicators assess the initiation of warfarin therapy for atrial fibrillation and the timing of such initiation. The only indicator that relates to the quality of management of warfarin, beyond the decision to use it at all, is Number 10, which states that the first INR test should occur within a week of the first dose. In the more recent Assessing Care of Vulnerable Elders (ACOVE) Study[73–75] (Table 2, page 149), some additional dimensions of process are measured. Compared with the earlier CQI Study, the ACOVE authors added a requirement to set a target INR range of 2–3 or to document the reason for another range (Stroke and Atrial Fibrillation, Indicator 5). In addition to specifying the timing of the first INR test after initiation of therapy (now within four days), a maximum interval of six weeks between subsequent INR tests is now specified (Medication Use, Indicator 7). Finally, an indicator measures the receipt of comprehensive education regarding warfarin therapy, or else referral to an ACC, where such education would presumably occur (Medication Use, Indicator 6).

In addition to CQI and ACOVE, a recently released set of performance measures deserves mention here: the American College of Cardiology/American Heart Association performance measures for atrial fibrillation.[76] The committee set forth three performance measures: assessing all patients for stroke risk fac-

Table 1. Quality Indicators Related to Oral Anticoagulation in the Community Quality Index Study*

Atrial Fibrillation, Indicator 3: Patients with atrial fibrillation of greater than 48 hours duration or of unknown duration who do not have contraindications to warfarin should receive warfarin if they are under 65† with one or more other risk factors for stroke.

Atrial Fibrillation, Indicator 4: Patients with atrial fibrillation of greater than 48 hours duration or of unknown duration who do not have contraindications to warfarin should receive warfarin if they are 65† years of age or older.

Atrial Fibrillation, Indicator 5: Patients with chronic atrial fibrillation who have contraindications to warfarin or have declined warfarin therapy should receive aspirin if they are under 65† with one or more other risk factors for stroke.

Atrial Fibrillation, Indicator 6: Patients with atrial fibrillation who do not have contraindications to warfarin should be started on warfarin within 2 weeks of presenting with new onset ischemic or embolic stroke.

Atrial Fibrillation, Indicator 7: Patients with atrial fibrillation who do not have contraindications to warfarin should be started on warfarin within 1 week of presenting with new onset transient ischemic attack.

Atrial Fibrillation, Indicator 8: Patients with atrial fibrillation of greater than 48 hours duration or of unknown duration who are undergoing elective electrical or chemical cardioversion should receive anticoagulation for at least 3 weeks prior to cardioversion unless they have had a transesophageal echocardiogram within 24 hours of cardioversion that indicates no clot.

Atrial Fibrillation, Indicator 9: All patients with atrial fibrillation of greater than 48 hours or unknown duration should receive anticoagulation for at least 4 weeks after cardioversion unless there are contraindications to anticoagulation.

Atrial Fibrillation, Indicator 10: Patients with atrial fibrillation started on warfarin should have an INR checked within 1 week of the first dose.

* Source: McGlynn E.A., et al.: *Appendix: The Quality of Health Care Delivered to Adults in the United States.* http://www.rand.org/pubs/working_papers/WR174-1/ (last accessed Jan. 26, 2008). INR, international normalized ratio.

† This age cutoff of 65 reflects knowledge and practice at the time these measures were developed. Current guidelines emphasize age 75 as a cutoff for receiving warfarin for atrial fibrillation, in the absence of other stroke risk factors. (Reference 78: Fuster V., et al.: ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation. *Circulation* 114:e257–e354, Aug. 15, 2006. Erratum in: *Circulation* 116(6):e138, Aug. 7, 2007.)

Table 2. Quality Indicators Related to Oral Anticoagulation in the Most Recent Version of the Assessing Care of Vulnerable Elders (ACOVE) Quality Indicators*

Stroke and Atrial Fibrillation, Indicator 3: IF a vulnerable elder (VE) has chronic atrial fibrillation and is at medium to high risk for stroke, THEN anticoagulation should be offered, BECAUSE anticoagulants reduce the risk of stroke, as well as vascular events (stroke, myocardial infarction [MI], and vascular death).

Stroke and Atrial Fibrillation, Indicator 4: IF a VE has chronic atrial fibrillation, is at medium to high risk for stroke, and has a contraindication to anticoagulation, THEN antiplatelet therapy should be prescribed, BECAUSE antiplatelet therapy reduces the risk of stroke, as well as vascular events (stroke, MI, and vascular death), although not as much as anticoagulants.

Stroke and Atrial Fibrillation, Indicator 5: IF a VE is prescribed anticoagulants for atrial fibrillation, THEN there should be documentation that the goal for the INR is 2.0 to 3.0 or reason for another goal, BECAUSE INR values kept within this range have the best trade-off between stroke prevention and risk of hemorrhage.

Stroke and Atrial Fibrillation, Indicator 6: IF a VE has had a TIA or ischemic stroke, THEN outpatient antiplatelet or anticoagulant therapy should be prescribed within 3 months after stroke or TIA or entering a new practice, BECAUSE antithrombotic treatment reduces the risk of recurrent stroke, as well as vascular events (stroke, MI, and vascular death).

Medication Use, Indicator 6: IF a VE receives a new prescription for warfarin, THEN he or she should receive education about diet and drug interactions and the risk of bleeding complications, or should be referred to an anticoagulation clinic, BECAUSE certain drugs and dietary substances interact with warfarin and can increase the risk of bleeding complications.

Medication Use, Indicator 7: IF a VE is prescribed warfarin, THEN an INR should be determined within 4 days after initiation of therapy and at least every 6 weeks thereafter, BECAUSE VEs are at particularly high risk for drug toxicity, and regular monitoring can help maintain patients within the therapeutic index.

* Sources: References 73–75: Cheng E.M., Fung C.H.: Quality indicators for the care of stroke and atrial fibrillation in vulnerable elders. *J Am Geriatr Soc* 55(suppl. 2):S431–S437, Oct. 2007; Shrank W.H., Polinski J.M., Avorn J.: Quality indicators for medication use in vulnerable elders. *J Am Geriatr Soc* 55(suppl. 2):S373–S382, Oct. 2007; Wenger N.S., Shekelle P.G.: Assessing care of vulnerable elders: ACOVE project overview. *Ann Intern Med* 135:642–646, Oct. 16, 2001. INR, international normalized ratio; TIA, transient ischemic attack.

tors, offering warfarin to all patients with any high-risk factor or more than one moderate-risk factor for stroke, and monitoring INR at least monthly. These performance measures generally echo the approaches used in the CQI and ACOVE studies.

The use of INR testing interval as a quality indicator (ACOVE Medication Use, Indicator 7, and ACC/AHA Guidelines) illustrates some of the limitations of previous efforts at process measurement in oral anticoagulation. Current consensus guidelines recommend that all patients have an INR test at least every 28 days or every six weeks,[19,77–79] but these recommendations are based on expert opinion rather than evidence that a minimum testing interval improves outcomes of

care. Because of the expense and inconvenience of laboratory checks, it is important to measure the INR sufficiently frequently to optimize therapy, but not more often than that, given that patients and their caregivers bear the burden of frequent clinic visits. Kent colleagues have derived a computerized algorithm to titrate follow-up intervals on the basis of INR variability,[80] with the premise that the risk associated with a testing interval of six weeks in a patient with stable control would be different than that of a patient with erratic control. They were able to demonstrate, in a randomized trial, that use of the algorithm increased follow-up intervals without compromising INR control, a desirable result.[81] Although their system is not widely used, their findings do not necessarily suggest the need for an absolute maximum interval between INR tests but rather suggest that testing intervals can be tailored to the individual patient.

Another potential process measure that was not included in prior efforts is the timeliness of follow-up for an aberrant INR value. In clinical practice, the exigency related to an out-of-range value is largely predicated on a particular patient's baseline risk of hemorrhage or thromboembolism and the degree of INR derangement. Evidence regarding the optimal follow-up interval for aberrant INR values is lacking, and the topic is unlikely to be subjected to a clinical trial due to ethical concerns. However, the importance of adequate follow up for especially extreme INR values (> 5.0, for example) could be empirically demonstrated through a link to intermediate or definitive outcomes, providing evidence in an area unlikely to see a clinical trial.

There is increasing interest in the use of genetic markers to predict a patient's steady state dose of warfarin before inception of therapy or early in the course of therapy.[82] Such a strategy might allow the earlier achievement of a therapeutic INR while avoiding supratherapeutic INR values and their attendant risk of hemorrhage.[82] Trials comparing this strategy to usual care at the inception of warfarin therapy are ongoing[83]; if this strategy is shown to improve care in a cost-effective manner, it could be an important process measure in the future.

Another aspect of process deserving further study is the optimal way to adjust doses of warfarin. There have been important advances in this area; for example, a validated dose calculator is now freely available on the Internet.[84] In addition, our group has recently shown that reserving dose changes for patients whose INR is at least 0.3 outside the target range has the potential to improve percent time in the therapeutic INR range (TTR) in clinical practice.[85] Optimal warfarin dose management remains an important but understudied topic; further study may yield better evidence and useful quality indicators.

Other process measures could be studied but have not been, possibly because of a lack of sufficiently detailed data. For example, there are guidelines for which patients should receive vitamin K in response to a very high INR[19]; adherence to such guidelines could be studied, and the effect on outcomes quantified.

Intermediate Outcomes

It is often impractical to follow large enough groups of patients for long enough to study definitive outcomes, which may occur rarely. As long as intermediate outcomes of care have been convincingly linked to definitive outcomes, they may serve as a useful surrogate, allowing for more-feasible study designs. The most commonly used intermediate outcome in anticoagulation is TTR, which uses linear interpolation to assign an INR value to each day between INR measurements.[86]

Several studies have provided strong evidence of a link between TTR and definitive outcomes.[23,24,87–89] The most direct evidence for TTR as a predictor of adverse events comes from a study that divided patients into three groups by TTR: < 60%, 60%–75%, and > 75%. The group with the worst control had more adverse outcomes than the other groups, including death, stroke or systemic embolus, major bleeding, and myocardial infarction.[23] Another study estimated that 26% of hemorrhagic and 11% of thromboembolic events among warfarin patients are attributable to time spent above and below the target range, respectively.[24] The importance of keeping patients within the target range was further illustrated by another study, which estimated that unless a TTR of at least 58% can be achieved, patients with atrial fibrillation would receive no more benefit from warfarin than they would from aspirin, the second-line therapy.[87]

Intermediate outcomes other than TTR have also been used to assess INR control. Fihn et al. have proposed an INR variability measure and have demonstrated that it also predicts adverse events.[90,91] This measure is calculated by finding the mean INR value for each patient and then the standard deviation around that mean; the standard deviation is a measure of INR variability.[80] Although it is known that both TTR and INR variability predict adverse events, their predictive ability has never been compared in the same study. It is even possible that both together will predict adverse events better than either one alone.

An even simpler measure would be to calculate the proportion of INR values within the target range, without linear interpolation between values. This is especially attractive because

sites of care could measure the quality of anticoagulation care without the need for more complicated statistical analysis. However, it would be important to document that the proportion of INR values in range has the same ability to predict definitive outcomes as TTR.[23] Several studies have compared this simpler measure to interpolated TTR in terms of the numerical results produced,[92,93] but it would be more important to compare intermediate outcome measures with regard to a gold standard of definitive outcomes.

In addition, it is likely that TTR can be improved as a quality indicator by identifying and remediating measurement artifacts. For example, some ACCs may intentionally document low INR values prior to procedures, a practice that would reduce TTR but would not imply inferior quality of care. Deleting INR values proximal to an intentional interruption of therapy may improve the performance of TTR as a quality indicator.[94] In addition, previous studies using TTR to compare sites have been ambiguous about the method used to accommodate differential contributions of patient-time. Some studies have evaluated entire clinics as if they are one patient observed for thousands of person-years[57,95]; others have calculated TTR for each patient and then averaged them, without weighting for time in the database,[56,60,96,97] and others have calculated mean site TTR both ways[63] or have not specified the method used.[88,97] A formal comparison of these methods could establish a standard, which would be important for future attempts to compare sites of care on TTR.

Whether TTR or some other measure is used, intermediate outcomes seem to be ideal quality indicators for oral anticoagulation care. They measure what is arguably the single most important function of an anticoagulation provider—to keep the patient within the target range as much as possible. In addition, they have been linked to definitive outcomes, which have obvious importance.[23,24,87–89] However, if intermediate outcomes are to be used to profile sites of care, they must be risk adjusted, and no study has yet attempted to formulate a risk-adjustment scheme for TTR or any other intermediate outcome in oral anticoagulation. A clinically credible and statistically valid risk-adjustment mechanism would be an important step toward comparing intermediate outcomes among sites of care.

Definitive Outcomes: Major Bleeding and Thromboembolic Events

The two most commonly studied definitive outcomes of oral anticoagulation care are thromboembolism and hemorrhage; disease-specific mortality is a third definitive outcome, but one that occurs too infrequently for practical study. Thromboembolic events potentially preventable through anticoagulation include stroke, systemic embolus, and venous thromboembolism. In most studies, bleeding is divided into minor hemorrhage, which is generally conceptualized as bleeding that does not require medical management, and major hemorrhage, which includes bleeding requiring transfusion or bleeding into a critical anatomic site. The definitions for thrombosis, and even more so for hemorrhage,[98] have varied among studies, making comparisons difficult. This issue has diminished over time but has not disappeared entirely.

The most important difficulty in studying these definitive outcomes is that they occur infrequently. This would severely limit their use as quality indicators, even for the largest sites of care.[1] The traditional solution to such limitations is to link definitive outcomes to intermediate outcomes, and then to use the more convenient intermediate outcomes in subsequent studies. As discussed above, previous research has sufficiently established a link between at least one intermediate outcome (TTR) and definitive outcomes to enable the use of TTR as a quality indicator.[23,24,87–89] Future research should aim to more precisely define the relationship between intermediate and definitive outcomes of anticoagulation care, including effect sizes. Figure 1 (page 152) summarizes our research agenda to advance quality measurement in oral anticoagulation care.

Limitations, Concerns, and Challenges of Quality Measurement in Oral Anticoagulation

It will be important to continuously gauge our level of confidence in our quality indicators and to have realistic expectations of those that may be more imperfect than others. For example, because they do not have intrinsic significance, measures of structure and process should be strongly related to outcomes that matter to merit our interest.[30,41] Process measures, in particular, should be sophisticated enough to account for the complexity of real patients with multiple comorbidities. Overly simplistic or poorly conceived process measures may do much harm by oversimplifying what are often complex clinical decisions, especially in patients who have multiple comorbid conditions.[99,100] For example, a simplistic process measure might require that any patient with atrial fibrillation receive warfarin, whereas a better process measure might make an exception for a patient with a history of gastrointestinal hemorrhage. An even better process measure might weigh the relative importance of a non–life-threatening hemorrhage in the distant past against a markedly elevated stroke risk in this particular patient. Recent advances have greatly increased the validity of quality measure-

Eleven Research Goals to Advance Quality Measurement in Oral Anticoagulation

STRUCTURE OF CARE

1. Document relationship between site volume and outcomes in anticoagulation clinics (ACCs).
2. Document the effect of staff-to-patient ratio on outcomes in an ACC.
3. Describe aspects of ACC organizational management or leadership that contribute to outcomes.

PROCESS OF CARE

4. Document impact of high-quality patient education regarding warfarin on outcomes.
5. Develop an evidence base regarding optimal strategies for adjusting warfarin doses. Document impact on outcomes.
6. Study time to a next visit (or next INR value) when the INR is very high or very low. Document impact on outcomes.
7. Study adherence to guidelines regarding the provision of vitamin K for very high INR values. Document impact on outcomes.

INTERMEDIATE OUTCOMES OF CARE

8. Evaluate the relative ability of TTR, INR variability, and other measures of INR stability to predict definitive outcomes. This includes simpler measures than TTR (such as proportion of INR values in range) and more complex measures than TTR.
9. Develop and validate risk adjustment models for intermediate outcomes of anticoagulation care, such as TTR, to allow fair comparisons between sites of care.
10. Profile sites of care using risk-adjusted TTR (or another intermediate outcome). Identify high and low outlier sites for further study, and use them to identify strategies to improve TTR at all sites of care.

DEFINITIVE OUTCOMES OF CARE

11. More precisely define the relationship between TTR and definitive outcomes of care (stroke, venous thromboembolism, major hemorrhage).

Figure 1. *Eleven research goals to advance quality measurement in oral anticoagulation, organized by domain of quality measurement, are shown. INR, international normalized ratio; TTR, percent time in the therapeutic range.*

ment in the area of process,[72,75,101–104] but there is still potential to improve the sophistication of process measurement.

It is important to note that although consensus guidelines may provide useful guidance to practicing clinicians, they cannot automatically be assumed to represent valid process measures.[99,105] Consensus guidelines may be heavily based on the results of randomized trials, which would have excluded most patients.[106] In addition, many consensus guidelines are written by experts in a single field, with relatively little attention to the whole patient and his or her many comorbid conditions and competing priorities.[100,107] Although elements of consensus guidelines may be excellent candidates to serve as quality indicators, the burden of proof is on investigators to demonstrate that they have utility as quality indicators before they are ready for such use.

Outcome measures, to be useful as quality indicators, must be studied with databases that provide sufficient statistical power to reliably demonstrate real differences between providers or sites of care rather than merely showing differences attributable to chance alone.[1] As discussed, this would preclude the use of definitive outcomes as quality indicators for all but the largest sites of care; instead, attention should be focused on intermediate outcomes as possible quality indicators. To be credible, outcome measures must be adjusted for case mix to ensure that differences in performance are due to the quality of care rather than inherent patient characteristics.[42] Ideally, such risk adjustment would control for not only patient demographics and comorbidities but also for variables that may be harder to obtain, such as socioeconomic status.

Finally, it is important to move beyond a relatively narrow focus on outcomes as comprising definitive clinical events and intermediate measures of INR control. Of the six dimensions of health care performance identified by the Institute of Medicine—safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity[108]—only safety and effectiveness are addressed by the above research agenda. Although some studies in oral anticoagulation have focused on patient satisfaction or the cost of care,[29,56,62,97,109] more work is needed in this area; greater consideration of these outcomes will help us to make our care more responsive to patient needs and the needs of society. Despite some previous research describing disparities in oral anticoagulation,[110–114] additional research will be needed to define and reduce such disparities over time.

Conclusion

Oral anticoagulation care, a common and serious condition, is in need of a program of quality measurement. Previous research

has established the evidence for anticoagulant therapy across a broad spectrum of indications and has helped to achieve consensus on the optimal target intensity for various indications. The next task will be to use this body of evidence to develop valid measures of the structure, process, and outcomes of oral anticoagulation care. Valid quality indicators will give us a framework for quality improvement, whose two goals will be to maximize the effectiveness of therapy and to minimize harm. Similar methods could be used to develop a program of quality measurement in other areas of clinical practice.

Dr. Rose is supported by a Career Development Award from the Department of Veterans Affairs Health Services Research and Development Service. Dr. Hylek has received research funding from Bristol-Myers Squibb, the manufacturers of Coumadin® brand warfarin. The opinions expressed in this article do not necessarily represent the views or policies of the Department of Veterans Affairs.

Adam J. Rose, M.D., M.Sc. is Investigator, Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Bedford, Massachusetts; and Assistant Professor, Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston. **Dan R. Berlowitz, M.D., M.P.H.,** is Director, Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center; and Professor, Department of Medicine, Boston University School of Medicine, and Department of Health Policy and Management, Boston University School of Public Health. **Susan M. Frayne, M.D., M.P.H.,** is Investigator, Center for Health Care Evaluation, VA Palo Alto Healthcare System, Palo Alto, California; and Associate Professor, Division of General Internal Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, California. **Elaine M. Hylek, M.D., M.P.H.,** is Associate Professor, Department of Medicine, Boston University School of Medicine. Please address correspondence to Adam J. Rose, adamrose@bu.edu.

References

1. Hofer T.P., et al.: The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA* 281:2098–2105, Jun. 9, 1999.

2. Wysowski D.K., Nourjah P., Swartz L.: Bleeding complications with warfarin use: A prevalent adverse effect resulting in regulatory action. *Arch Intern Med* 167:1414–1419, Jul. 9, 2007.

3. The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation. The Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators. *N Engl J Med* 323:1505–1511, Nov. 29, 1990.

4. Stroke Prevention in Atrial Fibrillation Study. Final results. *Circulation* 84:527–539, Aug. 1991.

5. Secondary prevention in non-rheumatic atrial fibrillation after transient ischaemic attack or minor stroke. EAFT (European Atrial Fibrillation Trial) Study Group. *Lancet* 342:1255–1262, Nov. 20, 1993.

6. Connolly S.J., et al.: Canadian Atrial Fibrillation Anticoagulation (CAFA) Study. *J Am Coll Cardiol* 18:349–355, Aug. 1991.

7. Ezekowitz M.D., 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* 327:1406–1412, Nov. 12, 1992.

8. Petersen P., et al.: Placebo-controlled, randomised trial of warfarin and aspirin for prevention of thromboembolic complications in chronic atrial fibrillation. The Copenhagen AFASAK study. *Lancet* 1:175–179, Jan. 28, 1989.

9. Singer D.E., et al.: Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133(suppl. 6):546S–592S, Jun. 2008.

10. Hull R., et al.: Warfarin sodium versus low-dose heparin in the long-term treatment of venous thrombosis. *N Engl J Med* 301:855–858, Oct. 18, 1979.

11. Hull R., et al.: Different intensities of oral anticoagulant therapy in the treatment of proximal-vein thrombosis. *N Engl J Med* 307:1676–1681, Dec. 30, 1982.

12. Kearon C., et al.: Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133(suppl. 6):454S–545S, Jun. 2008.

13. Lagerstedt C.I., et al.: Need for long-term anticoagulant treatment in symptomatic calf-vein thrombosis. *Lancet* 2:515–518, Sep. 7, 1985.

14. Lohr J.M., et al.: Lower extremity calf thrombosis: To treat or not to treat? *J Vasc Surg* 14:618–623, Nov. 1991.

15. Salem D.N., et al.: Valvular and structural heart disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133(suppl. 6):593S–629S, Jun. 2008.

16. Miyasaka Y., 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* 114:119–125, Jul. 11, 2006.

17. Hylek E.M., Singer D.E.: Risk factors for intracranial hemorrhage in outpatients taking warfarin. *Ann Intern Med* 120:897–902, Jun. 1, 1994.

18. Hylek E.M., et al.: An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. *N Engl J Med* 335:540–546, Aug. 22, 1996.

19. Ansell J., et al.: Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133(suppl. 6):160S–198S, Jun, 2008.

20. Wells P.S., et al.: Interactions of warfarin with drugs and food. *Ann Intern Med* 121:676–683, Nov. 1, 1994.

21. Budnitz D.S., et al.: Medication use leading to emergency department visits for adverse drug events in older adults. *Ann Intern Med* 147:755–765, Dec. 4, 2007.

22. Schulman S., et al.: Hemorrhagic complications of anticoagulant and thrombolytic treatment: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133(suppl. 6):257S–298S, Jun, 2008.

23. White H.D., et al.: Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: Results from SPORTIF III and V. *Arch Intern Med* 167:239–245, Feb. 12, 2007.

24. van Walraven C., et al.: Burden of potentially avoidable anticoagulant-associated hemorrhagic and thromboembolic events in the elderly. *Chest* 131:1508–1515, May 2007.

25. Olsson S.B.: 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* 362:1691–1698, Nov. 22, 2003.

26. Boos C.J., More R.S.: Anticoagulation for non-valvular atrial fibrillation: Towards a new beginning with ximelagatran. *Curr Control Trials Cardiovasc Med* 5:3, Apr. 22, 2004.

27. Agnelli G., et al.: Treatment of proximal deep-vein thrombosis with the oral direct factor Xa inhibitor rivaroxaban (BAY 59-7939): The ODIXa-DVT (Oral Direct Factor Xa Inhibitor BAY 59-7939 in Patients With Acute Symptomatic Deep-Vein Thrombosis) study. *Circulation* 116:180–187, Jul. 10, 2007.

28. Harenberg J., Jorg I., Weiss C.: Observations of alanine aminotransferase

and aspartate aminotransferase in THRIVE studies treated orally with ximelagatran. *Int J Toxicol* 25:165–169, May–Jun. 2006.

29. O'Brien C.L., Gage B.F.: Costs and effectiveness of ximelagatran for stroke prophylaxis in chronic atrial fibrillation. *JAMA* 293:699–706, Feb. 9, 2005.

30. Donabedian A. *Explorations in Quality Assessment and Monitoring.* Ann Arbor, MI: Health Administration Press, 1980.

31. Agnelli G., et al.: Three months versus one year of oral anticoagulant therapy for idiopathic deep venous thrombosis. Warfarin Optimal Duration Italian Trial Investigators. *N Engl J Med* 345:165–169, Jul. 19, 2001.

32. Kearon C., et al.: A comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. *N Engl J Med* 340:901–907, Mar. 25, 1999.

33. Ridker P.M., et al.: Long-term, low-intensity warfarin therapy for the prevention of recurrent venous thromboembolism. *N Engl J Med* 348:1425–1434, Apr. 10, 2003.

34. Beyth R.J., Quinn L.M., Landefeld C.S.: Prospective evaluation of an index for predicting the risk of major bleeding in outpatients treated with warfarin. *Am J Med* 105:91–99, Aug. 1998.

35. Gage B.F., et al.: Validation of clinical classification schemes for predicting stroke: Results from the National Registry of Atrial Fibrillation. *JAMA* 285:2864–2870, Jun. 13, 2001.

36. Gage B.F., et al.: Clinical classification schemes for predicting hemorrhage: Results from the National Registry of Atrial Fibrillation (NRAF). *Am Heart J* 151:713–719, Mar. 2006.

37. Hylek E.M., et al.: Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. *Circulation* 115:2689–2696, May 29, 2007.

38. McAlister F.A., et al.: A randomized trial to assess the impact of an antithrombotic decision aid in patients with nonvalvular atrial fibrillation: The DAAFI trial protocol [ISRCTN14429643]. *BMC Cardiovasc Disord* 4:5, May 5, 2004.

39. Owren P.A., Aas K.: The control of dicumarol therapy and the quantitative determination of prothrombin and proconvertin. *Scand J Clin Lab Invest* 3:201–208, 1951.

40. Hirsh J., Bates S.M.: Clinical trials that have influenced the treatment of venous thromboembolism: A historical perspective. *Ann Intern Med* 134:409–417, Mar 6, 2001.

41. Brook R.H., McGlynn E.A., Cleary P.D.: Quality of health care. Part 2: Measuring quality of care. *N Engl J Med* 335:966–970, Sep. 26, 1996.

42. Iezzoni L.I. (ed.): *Risk Adjustment for Measuring Health Care Outcomes*, 3rd ed. Chicago: Health Administration Press, 2003.

43. Porter M.E., Teisberg E.O.: *Redefining Healthcare: Value Based Competition on Results*. Boston: Harvard Business School Press, 2006.

44. Rosenthal M.B., et al.: Pay for performance in commercial HMOs. *N Engl J Med* 355:1895–1902, Nov. 2, 2006.

45. Epstein A.M.: Pay for performance at the tipping point. *N Engl J Med* 356:515–517, Feb. 1, 2007.

46. Lee T.H.: Pay for performance, version 2.0? *N Engl J Med* 357:531–533, Aug. 9, 2007.

47. Lindenauer P.K., et al.: Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 356:486–496, Feb. 1, 2007.

48. Rathore S.S., Krumholz H.M.: Differences, disparities, and biases: Clarifying racial variations in health care use. *Ann Intern Med* 141:635–638, Oct. 19, 2004.

49. Institute of Medicine: *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.* Washington, DC: National Academy Press, 2002.

50. Kressin N.R.: Separate but not equal: The consequences of segregated health care. *Circulation* 112:2582–2584, Oct. 25, 2005.

51. Fiscella K., et al.: Inequality in quality: Addressing socioeconomic, racial, and ethnic disparities in health care. *JAMA* 283:2579–2584, May 17, 2000.

52. Werner R.M., Asch D.A., Polsky D.: Racial profiling: The unintended consequences of coronary artery bypass graft report cards. *Circulation* 111:1257–1263, Mar. 15, 2005.

53. Lavizzo-Mourey R., Jung M.: Fighting unequal treatment: The Robert Wood Johnson Foundation and a quality-improvement approach to disparities. *Circulation* 111:1208–1209, Mar. 15, 2005.

54. Cortelazzo S., et al.: Thrombotic and hemorrhagic complications in patients with mechanical heart valve prosthesis attending an anticoagulation clinic. *Thromb Haemost* 69:316–320, Apr. 1, 1993.

55. Witt D.M., et al.: Effect of a centralized clinical pharmacy anticoagulation service on the outcomes of anticoagulation therapy. *Chest* 127:1515–1522, May 2005.

56. Chiquette E., Amato M.G., Bussey H.I.: Comparison of an anticoagulation clinic with usual medical care: Anticoagulation control, patient outcomes, and health care costs. *Arch Intern Med* 158:1641–1647, Aug. 10–24, 1998.

57. Matchar D.B., et al.: Improving the quality of anticoagulation of patients with atrial fibrillation in managed care organizations: Results of the managing anticoagulation services trial. *Am J Med* 113:42–51, Jul. 2002.

58. Siebenhofer A., et al.: Self-management of oral anticoagulation in the elderly: Rationale, design, baselines and oral anticoagulation control after one year of follow-up. A randomized controlled trial. *Thromb Haemost* 97:408–416, Mar. 2007.

59. Watzke H.H., et al.: A prospective controlled trial comparing weekly self-testing and self-dosing with the standard management of patients on stable oral anticoagulation. *Thromb Haemost* 83:661–665, May 2000.

60. Menendez-Jandula B., et al.: Comparing self-management of oral anticoagulant therapy with clinic management: A randomized trial. *Ann Intern Med* 142:1–10, Jan. 4, 2005.

61. Soliman Hamad M.A., et al.: Self management program improves anticoagulation control and quality of life: A prospective randomized study. *Eur J Cardiothorac Surg* Nov. 26, 2008 (epub).

62. Jacobson A.K., et al.: A prospective randomized controlled trial of the impact of home INR testing on clinical outcomes: The Home INR Study (THINRS), VA Cooperative Study #481 (abstract). Presented at the 2008 Annual Meeting of the American Heart Association, New Orleans, Nov. 8–12, 2008.

63. Gadisseur A.P., et al.: Comparison of the quality of oral anticoagulant therapy through patient self-management and management by specialized anticoagulation clinics in the Netherlands: A randomized clinical trial. *Arch Intern Med* 163:2639–2646, Nov. 24, 2003.

64. Aiken L.H., et al.: Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA* 288:1987–1993, Oct. 23–30, 2002.

65. Birkmeyer J.D., et al.: Hospital volume and surgical mortality in the United States. *N Engl J Med* 346:1128–1137, Apr. 11, 2002.

66. Peterson E.D., et al.: Procedural volume as a marker of quality for CABG surgery. *JAMA* 291:195–201, Jan. 14, 2004.

67. Manotti C., et al.: Effect of computer-aided management on the quality of treatment in anticoagulated patients: A prospective, randomized, multicenter trial of APROAT (Automated PRogram for Oral Anticoagulant Treatment). *Haematologica* 86:1060–1070, Oct. 2001.

68. Poller L., et al.: Multicentre randomised study of computerised anticoagulant dosage. European Concerted Action on Anticoagulation. *Lancet* 352:1505–1509, Nov. 7, 1998.

69. Poller L., Wright D., Rowlands M.: Prospective comparative study of computer programs used for management of warfarin. *J Clin Pathol* 46:299–303, Apr. 1993.

70. Ryan P.J., Gilbert M., Rose P.E.: Computer control of anticoagulant dose for therapeutic management. *BMJ* 299:1207–1209, Nov 11, 1989.

71. Wilson R., James A.H.: Computer assisted management of warfarin treatment. *Br Med J (Clin Res Ed)* 289:422–424, Aug. 18, 1984.

72. McGlynn E.A., et al.: *Appendix: The Quality of Health Care Delivered to Adults in the United States.* http://www.rand.org/pubs/working_papers/WR174-1/ (last accessed Jan. 26, 2008).

73. Cheng E.M., Fung C.H.: Quality indicators for the care of stroke and atrial fibrillation in vulnerable elders. J *Am Geriatr Soc* 55(suppl. 2):S431–S437, Oct. 2007.

74. Shrank W.H., Polinski J.M., Avorn J.: Quality indicators for medication use in vulnerable elders. *J Am Geriatr Soc* 55(suppl. 2):S373–S382, Oct. 2007.

75. Wenger N.S., Shekelle P.G.: Assessing care of vulnerable elders: ACOVE project overview. *Ann Intern Med* 135:642–646, Oct. 16, 2001.

76. Estes N.A., 3rd, et al.: ACC/AHA/Physician Consortium 2008 Clinical Performance Measures for Adults with Nonvalvular Atrial Fibrillation or Atrial Flutter. *J Am Coll Cardiol* 51:865–884, Feb. 26, 2008.

77. The use of oral anticoagulants (warfarin) in older people. AGS Clinical Practices Committee. American Geriatric Society. *J Am Geriatr Soc* 48:224–227, Feb. 2000.

78. Fuster V., et al.: ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation. *Circulation* 114:e257–e354, Aug. 15, 2006.

79. Fairweather R.B., et al.: College of American Pathologists Conference XXXI on laboratory monitoring of anticoagulant therapy: Laboratory monitoring of oral anticoagulant therapy. *Arch Pathol Lab Med* 122:768–781, Sep. 1998.

80. Kent D.L., et al.: A model for planning optimal follow-up for outpatients on warfarin anticoagulation. Warfarin Optimal Outpatient Follow-up Study Group. *Med Decis Making* 12:132–141, Apr.–Jun. 1992.

81. Fihn S.D., et al.: A computerized intervention to improve timing of outpatient follow-up: a multicenter randomized trial in patients treated with warfarin. National Consortium of Anticoagulation Clinics. *J Gen Intern Med* 9:131–139, Mar. 1994.

82. Gage B.F., Lesko L.J.: Pharmacogenetics of warfarin: Regulatory, scientific, and clinical issues. *J Thromb Thrombolysis* 25:45–51, Feb. 2008.

83. Schwarz U.I., et al.: Genetic determinants of response to warfarin during initial anticoagulation. *N Engl J Med* 358:999–1008, Mar 6. 2008.

84. Gage B.F.: *Warfarin Dosing*. http://www.warfarindosing.org (last accessed Jan. 26, 2009).

85. Rose A.J., et al.: Warfarin dose management affects INR control. *J Thromb Haemost* 7:94–101, Jan. 2009.

86. Rosendaal F.R., et al.: A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost* 69:236–239, Mar. 1, 1993.

87. Connolly S.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* 118:2029–2037, Nov. 11, 2008.

88. Jones M., et al.: Evaluation of the pattern of treatment, level of anticoagulation control, and outcome of treatment with warfarin in patients with nonvalvar atrial fibrillation: A record linkage study in a large British population. *Heart* 91:472–477, Apr. 2005.

89. Veeger N.J., et al.: Individual time within target range in patients treated with Vitamin K antagonists: Main determinant of quality of anticoagulation and predictor of clinical outcome. A retrospective study of 2300 consecutive patients with venous thromboembolism. *Br J Haematol* 128:513–519, Feb. 2005.

90. Fihn S.D., et al.: The risk for and severity of bleeding complications in elderly patients treated with warfarin. The National Consortium of Anticoagulation Clinics. *Ann Intern Med* 124:970–979, Jun. 1, 1996.

91. Fihn S.D., et al.: Risk factors for complications of chronic anticoagulation. A multicenter study. Warfarin Optimized Outpatient Follow-up Study Group. *Ann Intern Med* 118:511–520, Apr. 1, 1993.

92. Hutten B.A., et al.: Comparison of three methods to assess therapeutic quality control of treatment with vitamin K antagonists. *Thromb Haemost* 82:1260–1263, Oct. 1999.

93. Schmitt L., Speckman J., Ansell J.: Quality assessment of anticoagulation dose management: comparative evaluation of measures of time-in-therapeutic range. *J Thromb Thrombolysis* 15:213–216, Jun. 2003.

94. Rose A.J., et al.: Warfarin for atrial fibrillation in community-based practise. *J Thromb Haemost* 6:1647–1654, 2008.

95. Samsa G.P., et al.: Quality of anticoagulation management among patients with atrial fibrillation: Results of a review of medical records from 2 communities. *Arch Intern Med* 160:967–973, Apr. 10, 2000.

96. Fitzmaurice D.A., et al.: Oral anticoagulation management in primary care with the use of computerized decision support and near-patient testing: A randomized, controlled trial. *Arch Intern Med* 160:2343–2348, Aug. 14–28, 2000.

97. Menzin J., 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* 39:446–451, Mar. 2005.

98. Seto A.C., Kenyon K., Wittkowsky A.K.: Discrepancies in identification of major bleeding events in patients taking warfarin. *Pharmacotherapy* 28:1098–1103, Sep. 2008.

99. Pogach L.M., Rajan M., Aron D.C.: Comparison of weighted performance measurement and dichotomous thresholds for glycemic control in the Veterans Health Administration. *Diabetes Care* 29:241–246, Feb. 2006.

100. Werner R.M., et al.: Measuring quality of care in patients with multiple clinical conditions: Summary of a conference conducted by the Society of General Internal Medicine. *J Gen Intern Med* 22:1206–1211, Aug. 2007.

101. Higashi T., et al.: Quality of care is associated with survival in vulnerable older patients. *Ann Intern Med* 143:274–281, Aug. 16, 2005.

102. McGlynn E.A., et al.: The quality of health care delivered to adults in the United States. *N Engl J Med* 348:2635–2645, Jun. 26, 2003.

103. Shekelle P.G., et al.: Acove quality indicators. *Ann Intern Med* 135:653–667, Oct. 16, 2001.

104. Shekelle P.G., et al.: Assessing care of vulnerable elders: Methods for developing quality indicators. *Ann Intern Med* 135:647–652, Oct. 16, 2001.

105. Walter L.C., et al.: Pitfalls of converting practice guidelines into quality measures: Lessons learned from a VA performance measure. *JAMA* 291:2466–2470, May 26, 2004.

106. Van Spall H.G., et al.: Eligibility criteria of randomized controlled trials published in high-impact general medical journals: A systematic sampling review. *JAMA* 297:1233–1240, Mar. 21, 2007.

107. Tinetti M.E., Bogardus S.T. Jr., Agostini J.V.: Potential pitfalls of disease-specific guidelines for patients with multiple conditions. *N Engl J Med* 351:2870–2874, Dec. 30, 2004.

108. Institute of Medicine: *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press, 2001.

109. Gage B.F., Cardinalli A.B., Owens D.K.: Cost-effectiveness of preference-based antithrombotic therapy for patients with nonvalvular atrial fibrillation. *Stroke* 29:1083–1091, Jun. 1998.

110. Schauer D.P., et al.: Racial disparities in the filling of warfarin prescriptions for nonvalvular atrial fibrillation. *Am J Med Sci* 333:67–73, Feb. 2007.

111. Schauer D.P., et al.: Psychosocial risk factors for adverse outcomes in patients with nonvalvular atrial fibrillation receiving warfarin. *J Gen Intern Med* 20:1114–1119, Dec. 2005.

112. Birman-Deych E., et al.: Use and effectiveness of warfarin in Medicare beneficiaries with atrial fibrillation. *Stroke* 37:1070–1074, Apr. 2006.

113. Bhandari V.K., et al.: Quality of anticoagulation control: Do race and language matter? *J Health Care Poor Underserved* 19:41–55, Feb. 2008.

114. Aujesky D., et al.: African American race was associated with an increased risk of complications following venous thromboembolism. *J Clin Epidemiol* 60:410–416, Apr. 2007.

Comparing Methods of Measuring Treatment Intensification in Hypertension Care

Adam J. Rose, MD, MSc; Dan R. Berlowitz, MD, MPH; Meredith Manze, MPH; Michelle B. Orner, MPH; Nancy R. Kressin, PhD

Background—Greater treatment intensification (TI) improves hypertension control. However, we do not know the ideal way to measure TI for research and quality improvement efforts. We compared the ability of different TI measures to predict blood pressure (BP) control.

Methods and Results—We enrolled 819 hypertensive outpatients from an urban academic hospital. Each patient was assigned 3 scores to characterize TI. The any/none score divides patients into those who had any therapy increases during the study versus none. The norm-based method models the chance of a medication increase at each visit, then scores each patient based on whether they received more or fewer medication increases than predicted. The standard-based method is similar to the norm-based method but expects a medication increase whenever the blood pressure is uncontrolled. We compared the ability of these scores to predict the final systolic blood pressure (SBP). The any/none score showed a paradoxical result: any therapy increase was associated with SBP 4.6 mm Hg higher than no increase ($P<0.001$). The norm-based method score did not predict SBP in a linear fashion ($P=0.18$); further investigation revealed a U-shaped relationship between the norm-based method score and SBP. However, the standard-based method score was a strong linear predictor of SBP (2.1 mm Hg lower for each additional therapy increase per 10 visits, $P<0.001$). Similarly, the standard-based method predicted dichotomized blood pressure control, as measured by SBP <140 mm Hg (odds ratio, 1.30; $P<0.001$).

Conclusions—Our results suggest that standard-based method is the preferred measure of treatment intensity for hypertension care. **(*Circ Cardiovasc Qual Outcomes*. 2009;2:385-391.)**

Key Words: hypertension ■ chronic disease ■ research ■ quality of health care ■ ambulatory care

Improving cardiovascular outcomes will require valid approaches to measuring quality of care. Measuring the quality of hypertension management is an especially important goal because improved blood pressure (BP) control has great potential to improve cardiovascular outcomes. One possible measure of the quality of hypertension care is the intensity of clinical management when BP is uncontrolled. As early as 1979, the Hypertension Detection and Follow-up Program demonstrated that as compared with usual care, an algorithmic stepped-care approach to treating hypertension improves BP control and reduces morbidity and mortality.[1] More recently, investigators have also shown that in observational settings, patients who receive more intensive management for hypertension have better BP control.[2,3] Because of this demonstrated importance, there is increasing interest in measuring treatment intensification (TI) in the management of hypertension. A valid measure of TI could be used to profile providers and could be an important element of research and quality improvement efforts in hypertension care. However, there is no consensus regarding the best way to measure TI, and at least 3 methods have been used in previous studies.[2–4]

The first method examines whether a patient has had any medication increases during a period of time (versus none).[4,5] This approach has 2 flaws. First, it cannot distinguish gradations of TI, only any intensification versus none. Second, it fails to account for confounding by indication, the phenomenon wherein patients with the most severe disease receive more intensive medical therapy.[6] As might be expected, therefore, previous studies using this method have found the paradoxical result that greater TI seems to worsen BP control.[4] However, this method has remained in use through 2008,[4,5] so it is important to evaluate its validity.

The other 2 approaches can measure gradations of TI and do not produce paradoxical results, suggesting that they are not confounded by indication. One of these approaches relies

Received November 25, 2008; accepted April 21, 2009.

From the Center for Health Quality, Outcomes, and Economic Research (A.J.R., D.R.B., M.B.O., N.R.K.), Bedford VA Medical Center; the Department of Medicine, Section of General Internal Medicine (A.J.R., D.R.B., M.M., N.R.K.), Boston University School of Medicine; the Department of Health Policy and Management (D.R.B., M.M.), Boston University School of Public Health; and the VA Boston Healthcare System (N.R.K.), Boston, Mass.

The online-only Data Supplement can be found at http://circoutcomes.ahajournals.org/cgi/content/full/10.1161/CIRCOUTCOMES.108.838649/DC1.

Correspondence to Adam J. Rose, MD, MSc, Center for Health Quality, Outcomes, and Economic Research, 200 Springs Road, Building 70, Bedford, MA 01730. E-mail adamrose@bu.edu

Circ Cardiovasc Qual Outcomes **is available at http://circoutcomes.ahajournals.org** **DOI: 10.1161/CIRCOUTCOMES.108.838649**

WHAT IS KNOWN

- More intensive management of hypertension improves blood pressure over time in both experimental and observational studies.
- Several different systems of measuring treatment intensification have been used in the literature, but these systems have not previously been compared regarding their ability to predict blood pressure over time (predictive criterion validity).

WHAT THE STUDY ADDS

- We found that a standard-based method, which essentially "expects" a medication change whenever the blood pressure is elevated, performs better than other methods of measuring treatment intensification.
- Future research and quality improvement efforts should preferably use the standard-based method of measuring treatment intensification rather than the other methods we studied.
- Improved measures of treatment intensification and other quality-related constructs can increase the relevance of research efforts and magnify the effect on clinical care.

on a norm-based method (NBM) for defining care as more or less intensive, whereas the other relies on a standard-based method (SBM). The NBM, described by Berlowitz et al,[2] first derives a model to predict the probability of a dose increase at each visit according to various visit characteristics, then compares observed versus predicted dose changes to characterize each patient's care as more or less intensive than expected. The SBM, described by Okonofua et al,[3] simply compares the number of dose changes observed with the number of occasions on which the BP was 140/90 mm Hg or higher. In this system, a dose change is essentially "expected" whenever the BP is uncontrolled. Some have noted that SBM has certain inherent advantages over NBM because it is easier to calculate and interpret.[7] However, NBM incorporates a more nuanced view of clinical decision making because it allows for the possibility that factors other than the BP may influence the decision to intensify therapy, as well as the possibility that gradations of BP may exert differential influence on this decision. If NBM were the most valid measure of TI, as measured by BP control, it might be preferred, despite difficulties of calculation and interpretation.

However, different methods of measuring TI have not been directly compared regarding their ability to predict BP control over time. Because TI is a measure of process of care, linking it to BP control outcomes can demonstrate its validity and utility.[8,9] We therefore used data from a study of hypertensive patients at an academic urban safety net medical center to address 2 questions: (1) To what extent do these different measures of TI identify the same patients as having received more or less intensive management, and (2) Which, if any, of these 3 measures of TI best predicts BP control over time? Whatever our results, we expected them to inform future efforts to measure TI in the management of hypertension.

Methods

Enrollment

This report is a secondary analysis of data from a randomized trial designed to test whether a clinician-directed curriculum about patient-centered counseling could improve doctor–patient communication, adherence to therapy, and BP control (clinicaltrials.gov identifier: NCT00201149). Patients were enrolled from 7 outpatient primary care clinics at Boston Medical Center, an inner-city safety net hospital affiliated with the Boston University School of Medicine. The study was approved by the Institutional Review Board of Boston University Medical Center. We identified all patients of white or black race, age 21 and older, with outpatient diagnoses of hypertension on at least 3 separate occasions between August 2004 and June 2006.

Using this "universe" of 10 125 hypertensive patients from 7 clinics, the study staff tracked these patients' clinic visits over a 19-month period, and, as they presented for care, approached 3526 of them to request participation in the study. All willing respondents were then asked a series of questions and administered a cognitive screen to determine eligibility. A total of 1082 patients were excluded. Reasons included seeing a medical student at their visit (n=257), use of a daily medication dispenser (because it might invalidate collection of adherence data, n=247), cognitive impairment according to our cognitive screen (n=199), ethnicity other than white or black (n=149), unable to speak English (n=71), not prescribed antihypertensive medication (n=61), participation in another hypertension study (n=30), hearing impairment (n=16), and other (n=52), leaving 2444 eligible patients. Of those, 654 patients overtly refused to participate and 920 patients responded that they did not have time to participate that day. Total enrollment was therefore 870 patients.

Dependent Variable: Final Systolic Blood Pressure

The primary outcome was each patient's final systolic blood pressure (SBP) value, drawn from the clinical record of Boston Medical Center. We chose SBP rather than diastolic blood pressure (DBP) as our primary outcome because many more patients have poorly controlled SBP.[10] However, we also examined several secondary outcomes of hypertension care, including DBP and dichotomized measures of SBP, DBP, and overall BP control.

Categorizing Medication Increases

Automated data from Boston Medical Center's electronic medical record were examined. Our database included all prescriptions written, as well as all clinical BP values recorded within the study period. The unit of analysis was a visit to the primary care clinic, as identified by a date on which a BP value was recorded. When there were multiple BP values recorded on one date, we chose the one with the lowest SBP; if two values were tied, we selected the one with the lower DBP.

We recorded the patient's initial regimen of antihypertensive medications, ie, the regimen before study inception. One of the authors (A.J.R.) manually reviewed all prescriptions for each patient to see when the BP regimen was increased. An increase in medication was defined as either a new medication being added to the regimen or an increase in the dose of an existing medication. The period between each 2 BP values was assigned a 1 if the regimen was increased during that period, or a 0 if it was not. Multiple increases during a single period were counted as a 1. Dose changes occurring after the final visit were not recorded. A subset of 42 patients, representing 495 (5%) of all clinic visits, were randomly selected for blind reabstraction by another author (D.R.B.). Agreement between the 2 reviewers was excellent (κ=0.93; 95% CI, 0.87 to 0.98).

Covariates

We collected patient demographic data, including race (black or white), gender, and age. Using both ICD-9 codes and problem lists from the electronic medical record, we noted whether the patients had the following comorbid conditions, all of which could impact the BP, the use of antihypertensive medications, or the perceived urgency of controlling hypertension: benign prostatic hypertrophy,

cerebrovascular disease, congestive heart failure, chronic kidney disease, coronary artery disease, diabetes mellitus, hyperlipidemia, obesity (body mass index >30), peripheral vascular disease, and tobacco use.

Independent Variable: Any/None Score

The any/none score was "1" if the patient had at least 1 therapy increase during the study; otherwise, it was "0." The any/none score does not account for the number of visits or the degree of BP elevation.

Independent Variable: Norm-Based Method Score

To create the NBM score,[2] we began by deriving and validating a model to predict medication increases at each visit. The unit of analysis was each individual clinic visit; the outcome was whether or not the medications were increased at the visit. Our hypotheses regarding likely predictors were derived from our clinical judgment as well as our experience with the strongest predictors in previous similar models.[2,11,12] We considered the following possible predictors: SBP at the current and the previous visit, DBP at the current and the previous visit, number of days since the previous visit, whether the medications were increased at the previous visit, and the entire list of variables described above under "Covariates."

We initially screened variables using recursive partitioning (CART modeling),[13] using the R statistical package, version 2.6 (R Foundation, 2007). This method assigns each clinic visit into 1 of several categories according to several important predictors; each category is characterized by a particular frequency of medication increase. The important variables and cutoff values are empirically determined by the modeling procedure.

Having used CART to screen variables, we proceeded to derive and validate our predictive model using logistic regression. The dataset was split 60/40, with the larger subset used for derivation and the smaller for validation. We tried all candidate variables in our models, focusing particularly on those identified as important by CART modeling. In selecting cutoff values for continuous variables, we were guided by the output from CART model results and results of bivariate analyses. There were 5 predictors in the final model: (1) current SBP, (2) current DBP, (3) days since last visit, (4) DBP at previous visit, and (5) whether the medication was adjusted at the last visit (see online-only Data Supplement A for model details). The c-statistic was 0.74 in the derivation set and 0.72 in the validation set; the Hosmer–Lemeshow test indicated good model fit ($P=0.59$ in the derivation set and 0.44 in the validation set).

We then calculated the total number of expected medication changes for each patient in the dataset by summing probabilities over all of their visits. For example, if a patient had 3 visits, with predicted probabilities of a medication change of 0.20, 0.30, and 0.50, then exactly one medication change would be expected over this 3-visit period. We assigned each patient an NBM score, using the following formula:

(observed medication changes − NBM-predicted medication changes)/number of clinic visits

NBM scores are between −1 and 1, with 0 as the midpoint of the score. A score of 0 indicates a precise match between observed and expected medication increases, with positive numbers indicating more medication increases than expected and negative numbers indicating fewer increases than expected. As an example, over a 10-visit period, a patient might have a total of 5 predicted medication increases using NBM. If this patient actually had 3 visits with medication increases, the NBM score would be −0.2, indicating 2 fewer medication increases than expected per 10 visits. If the patient had 6 visits with therapy increases, the NBM score would be 0.1, indicating 1 more medication increase than expected per 10 visits.

We also created an alternative NBM score for each patient, based solely on the results of our CART model (online-only Data Supplement B), as in the original article by Berlowitz et al.[2] Results obtained using this score were not meaningfully different from our main NBM score and are not shown.

Independent Variable: Standard-Based Method Score

For the SBM analysis,[3] the expected number of medication increases was the number of occasions on which the recorded BP was 140/90 mm Hg or higher. Using this number and the number of occasions on which the medication was intensified each patient was assigned a score between −1 and 1. To make comparisons with NBM more straightforward, we reversed the polarity of the SBM score from what was originally described by Okonofua et al.[3] Therefore, we computed the SBM score using the following formula:

(observed medication changes − SBM-predicted medication changes)/number of clinic visits

For example, a patient with 5 elevated BP values over 10 visits would have a predicted value of 5 therapy increases. If this patient actually had 3 visits with medication increases, the score would be 3/10 to 5/10=−0.2, or 2 fewer therapy increases than expected per 10 visits. If the patient had 6 visits with therapy increases, the score would be 6/10 to 5/10=0.1, or 1 more therapy increase than expected per 10 visits.

We recognize that for patients with diabetes or chronic kidney disease, current guidelines set a lower BP target (ie, 130/80 mm Hg).[14] We therefore created an alternative SBM score only for patients with a low BP target. For this alternative SBM score, a medication increase was expected on each occasion when the recorded BP is 130/80 mm Hg or higher as opposed to 140/90 mm Hg for the main TI score. We divided the sample into patients with the higher and the lower BP thresholds and repeated our analyses for each group using the appropriate TI score. Results of this sensitivity analysis were similar to our main analysis and are not shown.

Statistical Analyses

Each patient was assigned 3 scores to measure TI in their hypertension care: any/none, NBM, and SBM. We examined the degree to which these 3 measures of TI were intercorrelated. For comparisons involving the any/none score, we used *t* tests to compare means of the other 2 scores when the any/none score was "any" versus "none." We compared the NBM and SBM scores using Spearman correlation (because of the non-Gaussian distribution of the SBM score) as well as dividing them into quartiles and constructing a 4×4 table.

We then examined the predictive validity of these 3 scores for the main dependent variable, the final SBP (continuous), as well as several secondary measures of BP control, including final DBP (continuous) and whether the final SBP was <140 mm Hg (categorical). For the any/none score, we compared the "any" group with the "none" group using *t* tests or χ^2, as appropriate. For the NBM and SBM scores, we used linear or logistic regression to model the relationship between the score and the BP outcomes, as appropriate. We repeated these analyses, controlling for patient-level covariates. We also divided the NBM and SBM scores into quartiles and performed ANOVA tests regarding the ability of the quartiles to predict the final SBP. For all analyses except the CART modeling, we used SAS, version 9.1 (SAS Institute). The authors had full access to the data and take responsibility for their integrity. All authors have read and agree to the manuscript as written.

Results

Patient Characteristics

Of 870 patients enrolled in the study, 51 were excluded from this analysis because they had 2 or fewer BP values. Therefore, 819 patients with hypertension, managed at Boston Medical Center, constituted our study population (Table 1). The mean follow-up time was 24 months; on average, patients visited the clinic once every 2 months. The mean age was 59.6 years, 34% of patients were male, and most (58%) were of black race. Considering their relatively young age, the

Table 1. Patient Characteristics (n=819)

Characteristic	No. (%) or Mean (SD)
Age, y	59.6 (11.4)
Male gender	278 (34)
Race	
White	343 (42)
Black	476 (58)
Current smoker	61 (7)
Obese	483 (59)
Comorbid conditions	
Benign prostatic hypertrophy	30 (4)
Cerebrovascular disease	46 (6)
Chronic kidney disease	55 (7)
Congestive heart failure	29 (4)
Coronary artery disease	105 (13)
Diabetes	272 (33)
Hyperlipidemia	439 (54)
Peripheral vascular disease	43 (5)
Frequency of clinic visits	
Person-time, mo	24.3 (7.9)
No. of clinic visits	12.0 (10.4)
Clinic visits/month	0.50 (0.40)
Medication classes received at baseline	
ACE inhibitors or ARBs	529 (65)
β-blockers	370 (45)
Calcium channel blockers	297 (36)
Diuretics, thiazide or loop	532 (65)
All other classes combined	95 (12)
Baseline No. of medications	
None	6 (1)
1	206 (25)
2	299 (37)
3	201 (25)
4 or more	106 (13)
Baseline BP control	
Baseline BP, mm Hg	134/80 (17/11)
Patients with baseline BP <140/90 mm Hg	447 (55)
Final BP control	
Final BP, mm Hg	133/79 (17/11)
Patients with final BP <140/90 mm Hg	489 (60)

population had a high burden of comorbidity: 54% had hyperlipidemia, 33% had diabetes, 13% had coronary artery disease, and 59% were obese. Most patients (74%) were receiving 2 or more antihypertensive medications at study inception. The population was characterized by relatively well controlled hypertension at baseline: the mean BP was 134/80 mm Hg, and 55% of patients were below 140/90 mm Hg.

Medication Increases and Measures of Treatment Intensity

After excluding the initial and final clinic visits for each patient (which were not analyzed regarding therapy increases), therapy was increased at 835 of the 9828 clinic visits (8.5%); 406 patients (50%) had at least 1 therapy increase during the study. Among patients with at least 1 therapy increase, the mean number of increases was 2.1, and the median was 2.0. We calculated NBM and SBM scores for each patient in the database. NBM scores were narrowly distributed (median, −0.04; interquartile range, [IQR] −0.06, 0.05; 5th and 95th percentiles, −0.13, 0.25). SBM scores were more widely distributed (median, −0.25; IQR, −0.50, −0.05; 5th and 95th percentiles, −0.80, 0.05).

Table 2. Comparison of Quartile Classifications of Patients Using the NBM Score and the SBM Score

SBM	Lowest TI (NBM)	Lower TI (NBM)	Higher TI (NBM)	Highest TI (NBM)	Totals
Lowest TI	132	18	36	22	208
Lower TI	61	39	55	61	216
Higher TI	11	77	46	49	183
Highest TI	4	75	60	73	212
Totals	208	209	197	205	819

Quartiles are not equal in size because of patients with identical TI scores.

Before examining the scores as predictors of BP control, we compared their classification of patients. The mean NBM score was 0.09 when the any/none score was "any," versus −0.07 when it was "none" ($P<0.001$). In contrast, the SBM score did not differ meaningfully between the 2 groups (−0.27 versus −0.30, $P=0.15$). The Spearman correlation between the NBM and SBM scores was 0.44, a fairly low correlation for 2 scores that are intended to measure the same construct. We also divided the NBM and SBM scores into quartiles and compared their classification of patients (Table 2). The κ statistic for agreement between these 2 scores was 0.14 (95% CI, 0.09 to 0.18). Extreme differences in quartile classification were not uncommon; for example, there were 209 patients (26%) whose quartile classifications differed by more than 1 category.

Any Therapy Increases as a Predictor of Blood Pressure Control

Any therapy increase (versus none) was examined as a predictor of the final BP. Patients with at least 1 therapy increase had a mean final SBP of 135.2 mm Hg as compared with a mean final SBP of 130.6 mm Hg among patients who had no therapy increases ($P<0.001$). As expected, because this measure does not control for confounding by indication, it produces a paradoxical result (therapy increases are associated with a higher final BP).

Norm-Based Method Score as a Predictor of BP Control

NBM score was a poor predictor of BP control (Table 3). In a linear regression, the NBM score was not a significant predictor of the final SBP (model $R^2=0.001$, $P=0.28$). Adding patient-level covariates improved the model fit somewhat. We investigated further by dividing the NBM score into quartiles (Table 4). A U-shaped relationship, rather than a linear relationship, was observed between NBM score quartiles and the final SBP. The NBM score also performed

Table 3. Performance of the NBM Score as a Predictor of the Final SBP, With and Without Covariates

Variable	NBM Score Alone	NBM Score With Covariates
Intercept	132.9*	124.8*
NBM Score (per 0.1)	−0.5	−0.6
Age		
Oldest (70+)		6.5†
Older (60–69)		5.0†
Younger (50–59)		2.9
Youngest (Under 50)		—
Benign prostatic hypertrophy		−2.9
Black race (vs white)		3.4†
Cerebrovascular disease		8.2†
Chronic kidney disease		−2.0
Congestive heart failure		1.0
Coronary artery disease		0.6
Diabetes mellitus		2.6
Female gender (vs male)		−0.1
Hyperlipidemia		0.3
Obesity		1.7
Peripheral vascular disease		1.2
Tobacco user		0.3
Model R^2	0.001	0.05

A positive change of 0.1 in the NBM score indicates 1 more therapy change per 10 visits. Intercepts and β coefficients predict the final systolic blood pressure and are expressed in units of mm Hg. A positive β coefficient denotes an increase in the final systolic blood pressure, whereas a negative β coefficient denotes a decrease.

*$P<0.001$; †$P<0.05$.

Table 5. Performance of the SBM Score as a Predictor of the Final SBP, With and Without Covariates

Variable	SBM Score Alone	SBM Score With Covariates
Intercept	127.1*	119.3*
SBM Score (per 0.1)	−2.1*	−2.0*
Age		
Oldest		5.9†
Older		5.3†
Younger		3.2
Youngest		—
Benign prostatic hypertrophy		−2.0
Black race (vs white)		2.1
Cerebrovascular disease		6.5†
Chronic kidney disease		−1.0
Congestive heart failure		1.5
Coronary artery disease		0.1
Diabetes mellitus		2.2
Female gender (vs male)		0.2
Hyperlipidemia		1.9
Obesity		0.9
Peripheral vascular disease		1.8
Tobacco user		0.4
Model R^2	0.12	0.16

A positive change of 0.1 in the SBM score indicates 1 more therapy change per 10 visits. Intercepts and β coefficients predict the final systolic blood pressure and are expressed in units of mm Hg. A positive β coefficient denotes an increase in the final systolic blood pressure, whereas a negative β coefficient denotes a decrease.

*$P<0.001$; †$P<0.05$.

poorly as a predictor of the final DBP ($R^2=0.002$, $P=0.19$) and as a predictor of whether the final SBP would be below 140 mm Hg (OR, 1.06 per change of 0.1; c-statistic, 0.56, $P=0.28$).

Standard-Based Method Score as a Predictor of Blood Pressure Control

In contrast to the NBM score, the SBM score was an excellent predictor of the final BP (Table 5). In a linear regression, the β coefficient was −2.1, indicating that for each 0.1 of the SBM score (1 more therapy increase per 10 visits), the final SBP was 2.1 mm Hg lower ($R^2=0.12$, $P<0.001$). Adding covariates to the model improved its fit by a margin similar to that with the NBM model, but SBM persisted as a powerful predictor of the final SBP. In additional stratified analyses, SBM performed similarly in males and females, in white and black patients, and among subgroups of patients with particularly severe comorbid conditions such as chronic kidney disease, congestive heart failure, and peripheral vascular disease.

Table 4. Quartiles of NBM Score and SBM Score as Predictors of the Final SBP

Treatment Intensity	NBM Scoring	SBM Scoring
Most intensive management	133.7	125.3
More intensive management	131.3	129.2
Less intensive management	127.9	135.6
Least intensive management	138.5	141.0
P value, ANOVA test	<0.001	<0.001

Mean systolic blood pressure for each group is given in mm Hg.

We investigated further by dividing the SBM score into quartiles (Table 4). A strong linear relationship was observed between SBM score quartiles and final SBP (P for linear trend <0.001). The SBM score was also a predictor of the final DBP (β coefficient, −0.8; $P<0.001$) and of whether the final SBP would be below 140 mm Hg (OR, 1.30 per change of 0.1; c-statistic, 0.70, $P<0.001$).

Discussion

Optimizing approaches to measuring the quality of care delivered to patients with chronic diseases is an important research goal. This is particularly true for measuring treatment intensity in the care of hypertension because we have decades of evidence showing that more intensive treatment improves BP control.[1–3] We therefore compared the predictive criterion validity of 3 approaches of measuring TI in hypertension care. We found that the any/none measure produces paradoxical results because it does not account for confounding by indication. To our surprise, we found that the

NBM score was not predictive of BP control. Further investigation demonstrated that the NBM score appeared to have a U-shaped relationship with BP outcomes, complicating its use as a predictor and calling into question its validity as a measure of TI, which is meant to be monotonic.

In contrast, the SBM score was a powerful predictor of the final BP, a relationship that remained undiminished after controlling for covariates. It is important to remember that the β coefficient we found for the effect of SBM on final SBP, −2.1 mm Hg, was for each additional therapy increase per 10 visits, a relatively small difference in management. Larger differences in management would obviously improve BP control much more. For example, the difference in final SBP between the highest and lowest quartiles of TI (125 mm Hg versus 141 mm Hg) suggests an effect of considerable magnitude and clinical significance.

We had expected to find that the any/none measure performs poorly as a measure of TI because previous studies have shown that a failure to account for confounding by indication produces paradoxical (or attenuated) results.[4–6] We had also expected to find that NBM is superior to SBM as a predictor of BP control because it incorporates a more nuanced representation of clinical decision making. The apparent lack of predictive criterion validity for NBM in our study contrasts with the findings of earlier studies, particularly the original article by Berlowitz et al.[2] This difference may be attributable to improved BP control: mean initial BP was 134/80 mm Hg in our study versus 146/83 in the earlier study.[2] NBM may have worked better in an era of mediocre BP control, whereas SBM may be more suited to pursuing what are ultimately smaller improvements in BP.

Our study has several limitations. First, TI is not universally accepted as an ideal theory to understand poor control of asymptomatic chronic conditions, especially when it is presented as "clinical inertia,"[15] the obverse of TI. Some studies have suggested that on deeper inspection, what seems to be clinical inertia could also be attributed to "competing demands,"[16,17] "clinical uncertainty,"[18] or "appropriate inaction."[19] Other studies have explored the relationship between TI and adherence,[4,5,20–23] or the patient and visit-level predictors of TI.[10,17,24–26] This study did not include specific measures of adherence, competing demands, patient complexity, clinical uncertainty, or appropriate inaction, although we did account for the burden of comorbid disease, which relates to several of these concepts (competing demands and patient complexity). However, because we compared multiple measures of TI using the same database, we can be assured that unmeasured covariates would have been equally true for all comparisons. In addition, although refinements to the TI concept are always welcome, our study reinforces the notion that TI, as embodied in the SBM score, is an important determinant of BP control.

Second, our study compared different methods of measuring TI in hypertension care. It should not be assumed, however, that SBM would also be the ideal system for measuring TI in the care of diabetes or hyperlipidemia; future research should address those questions. Finally, our data were drawn from an academic urban hospital, which may limit generalizability. The clinicians at Boston Medical Center may have managed hypertension differently than nonacademic clinicians. Similarly, the BP control in this cohort was quite good; it is possible that the SBM score may work particularly well in such a setting. In addition, many of the patients in our study were immigrants, ethnic minorities, and of low socioeconomic status. However, given the relatively good BP control achieved among this population, the challenges these patients face in their everyday lives do not seem to threaten the generalizability of our findings.

We have known for 30 years that more intensive management leads to better hypertension outcomes, in both clinical trials and observational settings.[1–3] What we have lacked is consensus about the best method to measure TI in the care of hypertension. Our study found that any/none and NBM were not valid measures of TI, whereas a SBM was an excellent predictor of BP control. Unless these results are challenged by other studies, SBM should be the preferred method of characterizing TI in future studies of hypertension care. SBM can now serve as the basis of research and quality improvement efforts to improve the process and outcomes of hypertension care.

Acknowledgments

We thank Mark Glickman, PhD, for his help with the CART models and Al Ozonoff, PhD, for other statistical advice.

Sources of Funding

This research was supported by a grant from the National Institutes of Health (HL072814, N.R. Kressin, Principal Investigator). Dr Rose is supported by a career development award from the Department of Veterans Affairs, Health Services Research and Development Service. Dr Kressin is supported by a Research Career Scientist award from the Department of Veterans Affairs, Health Services Research and Development Service (RCS 02-066-1). The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

Disclosures

None.

References

1. Hypertension Detection and Follow-up Program Cooperative Group. Five-year findings of the hypertension detection and follow-up program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension. *JAMA*. 1979;242:2562–2571.
2. Berlowitz DR, Ash AS, Hickey EC, Friedman RH, Glickman M, Kader B, Moskowitz MA. Inadequate management of blood pressure in a hypertensive population. *N Engl J Med*. 1998;339:1957–1963.
3. 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;47:345–351.
4. Ho PM, Magid DJ, Shetterly SM, Olson KL, Peterson PN, Masoudi FA, Rumsfeld JS. Importance of therapy intensification and medication nonadherence for blood pressure control in patients with coronary disease. *Arch Intern Med*. 2008;168:271–276.
5. Rose AJ, Berlowitz DR, Orner MB, Kressin NR. Understanding uncontrolled hypertension: is it the patient or the provider? *J Clin Hypertens*. 2007;9:937–943.
6. Rubin HR, Pronovost P, Diette GB. The advantages and disadvantages of process-based measures of health care quality. *Int J Qual Health Care*. 2001;13:469–474.
7. O'Connor PJ. Commentary: improving diabetes care by combating clinical inertia. *Health Serv Res*. 2005;40:1854–1861.
8. Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2. Measuring quality of care. *N Engl J Med*. 1996;335:966–970.

9. Donabedian A. *Explorations in Quality Assessment and Monitoring*. Ann Arbor, Mich: Health Administration Press; 1980.
10. Rose AJ, Shimada SL, Rothendler JA, Reisman JI, Glassman PA, Berlowitz DR, Kressin NR. The accuracy of clinician perceptions of "usual" blood pressure control. *J Gen Intern Med*. 2008;23:180–183.
11. Berlowitz DR, Ash AS, Glickman M, Friedman RH, Pogach LM, Nelson AL, Wong AT. Developing a quality measure for clinical inertia in diabetes care. *Health Serv Res*. 2005;40:1836–1853.
12. Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Hypertension control: how well are we doing? *Arch Intern Med*. 2003;163: 2705–2711.
13. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Monterey, Calif: Wadsworth and Brooks/Cole; 1984.
14. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr, Jones DW, Materson BJ, Oparil S, Wright JT Jr, Roccella EJ. 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.
15. Phillips LS, Branch WT, Cook CB, Doyle JP, El-Kebbi IM, Gallina DL, Miller CD, Ziemer DC, Barnes CS. Clinical inertia. *Ann Intern Med*. 2001;135:825–834.
16. Parchman ML, Pugh JA, Romero RL, Bowers KW. Competing demands or clinical inertia: the case of elevated glycosylated hemoglobin. *Ann Fam Med*. 2007;5:196–201.
17. Turner BJ, Hollenbeak CS, Weiner M, Ten Have T, Tang SS. Effect of unrelated comorbid conditions on hypertension management. *Ann Intern Med*. 2008;148:578–586.
18. 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.
19. Safford MM, Shewchuk R, Qu H, Williams JH, Estrada CA, Ovalle F, Allison JJ. Reasons for not intensifying medications: differentiating "clinical inertia" from appropriate care. *J Gen Intern Med*. 2007;22: 1648–1655.
20. Schmittdiel JA, Uratsu CS, Karter AJ, Heisler M, Subramanian U, Mangione CM, Selby JV. Why don't diabetes patients achieve recommended risk factor targets? Poor adherence versus lack of treatment intensification. *J Gen Intern Med*. 2008;23:588–594.
21. Heisler M, Hogan MM, Hofer TP, Schmittdiel JA, Pladevall M, Kerr EA. When more is not better: treatment intensification among hypertensive patients with poor medication adherence. *Circulation*. 2008;117: 2884–2892.
22. Grant R, Adams AS, Trinacty CM, Zhang F, Kleinman K, Soumerai SB, Meigs JB, Ross-Degnan D. Relationship between patient medication adherence and subsequent clinical inertia in type 2 diabetes glycemic management. *Diabetes Care*. 2007;30:807–812.
23. Grant RW, Singer DE, Meigs JB. Medication adherence before an increase in antihypertensive therapy: a cohort study using pharmacy claims data. *Clin Ther*. 2005;27:773–781.
24. Bolen SD, Samuels TA, Yeh HC, Marinopoulos SS, McGuire M, Abuid M, Brancati FL. Failure to intensify antihypertensive treatment by primary care providers: a cohort study in adults with diabetes mellitus and hypertension. *J Gen Intern Med*. 2008;23:543–550.
25. Lin ND, Martins SB, Chan AS, Coleman RW, Bosworth HB, Oddone EZ, Shankar RD, Musen MA, Hoffman BB, Goldstein MK. Identifying barriers to hypertension guideline adherence using clinician feedback at the point of care. *AMIA Annu Symp Proc*. 2006:494–498.
26. Steinman MA, Fischer MA, Shlipak MG, Bosworth HB, Oddone EZ, Hoffman BB, Goldstein MK. Clinician awareness of adherence to hypertension guidelines. *Am J Med*. 2004;117:747–754.

SUPPLEMENTAL MATERIAL

Appendix A: Logistic model to predict medication dose increases. Odds ratios above 1.0 indicate greater likelihood of a dose increase at a visit. While this model was derived and validated using a split dataset, the entire dataset was used for the analysis presented here (n = 9828 clinic visits).

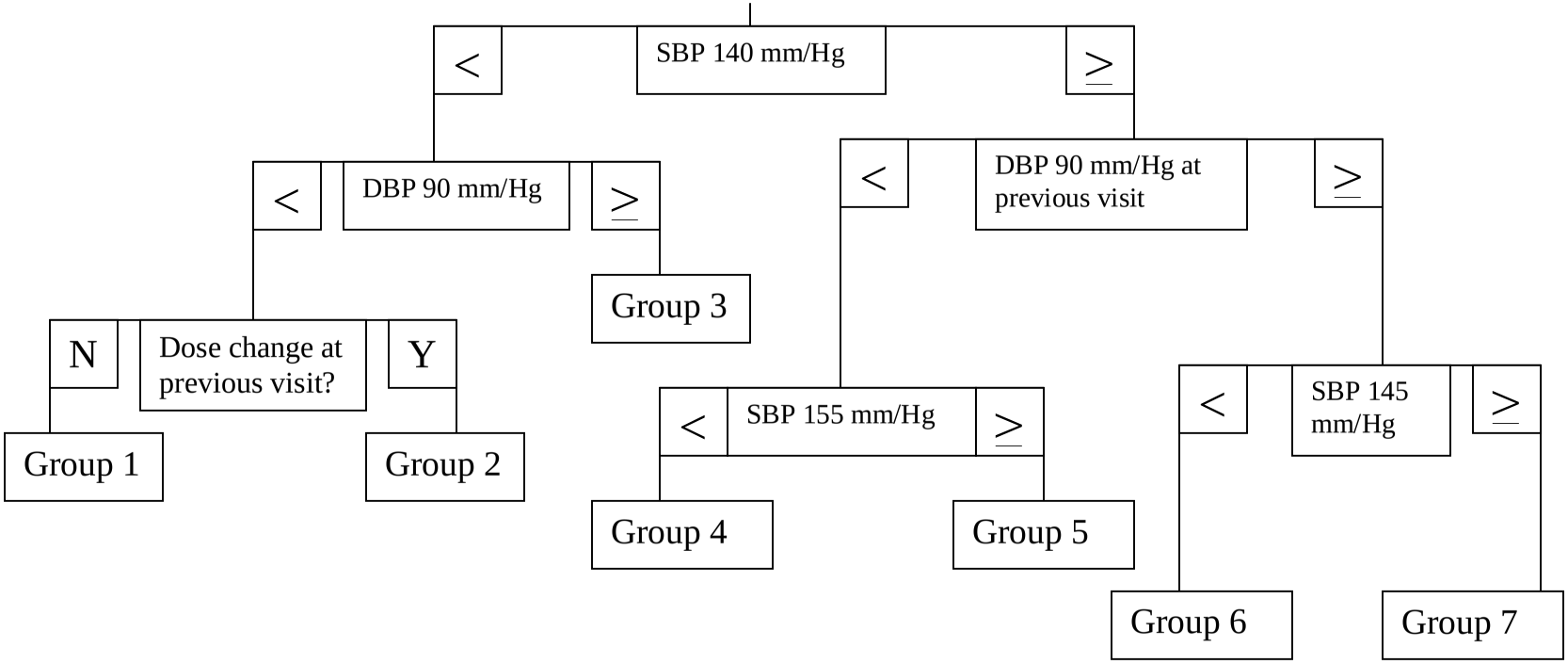
Predictor Variable	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
1. Systolic BP at current visit		
Below 130 mm/Hg	--	--
130-139 mm/Hg	1.5 (1.2 – 1.9)*	1.4 (1.1 – 1.7)*
140-149 mm/Hg	3.3 (2.6 – 4.1)*	2.6 (2.1 – 3.3)*
150 mm/Hg and above	6.3 (5.2 – 7.6)*	4.7 (3.8 – 5.7)*
2. Diastolic BP at current visit		
Below 90 mm/Hg	--	--
Above 90 mm/Hg	3.4 (2.9 – 4.0)*	1.7 (1.5 – 2.1)*
3. Diastolic BP at previous visit		
Below 90 mm/Hg	--	--
Above 90 mm/Hg	2.3 (1.9 – 2.7)*	1.4 (1.2 – 1.7)*
4. Days since previous visit		
Up to 119 days	--	--
120 days or more	1.5 (1.2 – 1.8)*	1.5 (1.2 – 1.8)*
5. Medication change at previous visit?		
No	--	--
Yes	2.7 (2.2 – 3.3)*	2.2 (1.8 – 2.7)*

*$p < 0.001$

Appendix B: CART-based model for predicting medication dose increases. The best model (presented here) had 7 terminal nodes, the same number as the model in the original paper by Berlowitz, et al. Cutoff BP values have been rounded to the nearest 5 mm/Hg for convenience and interpretability. This model had similar performance to our main model in terms of predicting which visits would have medication changes (c-statistic = 0.71, compared to 0.73 for the main model).

Group Definitions					Results	
Group	Current SBP	Current DBP	DBP at previous Visit	Dose Change at Previous Visit?	Number of Visits in Category	Chance of dose increase at each visit
1	<140 mm/Hg	< 90 mm/Hg		No	5700	3.8%
2	<140 mm/Hg	< 90 mm/Hg		Yes	399	8.4%
3	<140 mm/Hg	> 90 mm/Hg			442	11.0%
4	140 - 154 mm/Hg		< 90 mm/Hg		1727	11.6%
5	155+ mm/Hg		< 90 mm/Hg		792	18.9%
6	140 - 144 mm/Hg		> 90 mm/Hg		244	15.9%
7	145+ mm/Hg		> 90 mm/Hg		472	30.1%

Figure: Graphic Representation of CART-based model.



Intensifying Therapy for Hypertension Despite Suboptimal Adherence

Adam J. Rose, Dan R. Berlowitz, Meredith Manze, Michelle B. Orner, Nancy R. Kressin

Abstract—More intensive management can improve control blood pressure (BP) in hypertensive patients. However, many would posit that treatment intensification (TI) is not beneficial in the face of suboptimal adherence. We investigated whether the effect of TI on BP varies by adherence. We enrolled 819 patients with hypertension, managed in primary care at an academically-affiliated inner-city hospital. We used the following formula to characterize TI: (visits with a medication change−visits with elevated BP)/total visits. Adherence was characterized using electronic monitoring devices ("MEMS caps"). Patients who returned their MEMS caps (671) were divided into quartiles of adherence, whereas patients who did not return their MEMS caps (148) had "missing" adherence. We examined the relationship between TI and the final systolic blood pressure (SBP), controlling for patient-level covariates. In the entire sample, each additional therapy increase per 10 visits predicted a 2.0 mm Hg decrease in final SBP ($P<0.001$). After stratifying by adherence, in the "best" adherence quartile each therapy increase predicted a 2.1-mm Hg decrease in final SBP, followed by 1.8 for the "next-best" adherence quartile, 2.3 in the third quartile, and 2.4 in the "worst" adherence quartile. The effect size for patients with "missing" adherence was 1.6 mm Hg. The differences between the group with "best" adherence and the other 4 groups were not statistically significant. In this observational study, treatment intensification was associated with similar BP improvement regardless of the patient's level of adherence. A randomized trial could further examine optimal management of patients with suboptimal adherence. **(*Hypertension*. 2009;54:524-529.)**

Key Words: hypertension ■ adherence ■ medication therapy management ■ quality of care ■ ambulatory care

For almost 30 years, we have known that more intensive management of hypertension can improve blood pressure (BP) control, both in the setting of clinical trials[1] and in observational studies of routine clinical practice.[2,3] Similarly, it has long been appreciated that greater adherence to medication regimens can improve BP control.[4,5] More recently, there have been several efforts to understand the relationship between adherence and treatment intensity (TI) in the management of hypertension.[6–10] Some of these studies have addressed whether clinicians are more or less likely to increase therapy according to patient adherence,[6,7] whereas others have probed the relationship between TI and adherence in determining BP control over time.[8–10]

Explorations of the relationship between TI and adherence in determining BP control have been limited in their scope, mostly demonstrating that both TI and adherence have important effects on BP control.[8–10] However, a more important question has not yet been addressed, namely whether the effect of TI on BP control differs by adherence. This information would help inform the difficult clinical decision of how best to manage a patient suspected of suboptimal adherence to therapy. Despite the lack of evidence regarding this topic, there seems to be widespread agreement that it is not advisable to intensify therapy when a patient is nonadherent.[7,8,11] This may be because of a belief, on the part of clinicians, that nonadherent patients may not benefit from treatment intensification, and that it in fact may harm them by predisposing to hypotensive episodes when therapy is actually taken. However, the conviction that therapy should not be increased for nonadherent patients has not been subjected to empirical evaluation, and it seems to be based on a binary view of patients as completely adherent or completely nonadherent, when in fact most patients fall somewhere in between.[12]

We therefore set out to examine the association between TI, adherence, and BP control. Our study had 2 objectives: (1) to determine whether patient adherence to antihypertensive therapy predicts clinician decisions regarding therapy intensification, and (2) to determine whether the effect of TI on BP control differs among strata of adherence. We hypothesized that patients with suboptimal adherence would indeed have improved BP control with more intensive therapy, because a more potent regimen, even one taken less than 100% of the time, is likely to be more effective in controlling BP.

Received March 25, 2009; first decision April 7, 2009; revision accepted June 12, 2009.

From the Center for Health Quality, Outcomes, and Economic Research (A.J.R., D.R.B., M.B.O., N.R.K.), Bedford VA Medical Center; the Department of Medicine, Section of General Internal Medicine (A.J.R., D.R.B., M.M., N.R.K.), Boston University School of Medicine; the Department of Health Policy and Management (D.R.B., M.M.), Boston University School of Public Health; and the VA Boston Healthcare System (N.R.K.), Boston, Mass.

Correspondence to Adam J. Rose, MD, MSc, Center for Health Quality, Outcomes, and Economic Research, 200 Springs Road, Building #70, Bedford, MA 01730. E-mail adamrose@bu.edu

Hypertension **is available at http://hyper.ahajournals.org** **DOI: 10.1161/HYPERTENSIONAHA.109.133389**

Methods

Enrollment

This report is a secondary analysis of data from a randomized trial designed to test whether a clinician-directed curriculum about patient-centered counseling could improve doctor-patient communication, adherence to therapy, and blood pressure control (ClinicalTrials.gov Identifier: NCT00201149). Patients were enrolled from 7 outpatient primary care clinics at Boston Medical Center, an inner-city safety net hospital affiliated with the Boston University School of Medicine. The study was approved by the Institutional Review Board of Boston University Medical Center. We identified all patients of white or black race, age 21 and older, with outpatient diagnoses of hypertension on at least 3 separate occasions between August 2004 and June 2006. Because of this requirement for 3 previous outpatient diagnostic codes, our study enrolled only patients with prevalent as opposed to incident hypertension. Study staff then tracked the clinic visits of these 10 125 patients over a 19-month period, and, as they presented for care, approached 3526 of them to request participation in the study. Of those, 654 patients (19% of 3526) overtly refused to participate and 920 patients (26% of 3526) responded that they did not have time to participate, but we were unable to assess their eligibility before they declined. All willing respondents were then asked a series of questions and administered a cognitive screen to determine eligibility; 1083 patients (55% of the remaining 1952) were excluded, for reasons detailed in Figure S1 (please see http://hyper.ahajournals.org). Assuming a similar rate of ineligibility among patients whose eligibility was not assessed, we recruited 869 patients from a likely pool of 1578 eligible patients (55%).

Dependent Variable: Final Systolic Blood Pressure

The primary outcome was each patient's final SBP value, ie, the one immediately before study completion. These BP values were drawn from the clinical record of Boston Medical Center. We chose SBP rather than diastolic blood pressure as our primary outcome, because many more patients have poorly-controlled SBP.[13,14]

Categorizing Medication Increases

Automated data from Boston Medical Center's electronic medical record (EMR) were examined. Our database included all prescriptions written, as well as all clinical BP values recorded within the study period. The unit of analysis was a visit to the primary care clinic, as identified by a date on which a BP value was recorded. When there were multiple BP values recorded on one date, we chose the one with the lowest SBP; if two values were tied, we selected the one with the lower DBP.

We recorded the patient's initial regimen of antihypertensive medications, ie, the regimen before study inception. One of the authors (A.J.R.) manually reviewed all prescriptions for each patient to see when the BP regimen was increased. An increase in medication was defined as either a new medication being added to the regimen or an increase in the dose of an existing medication. The period between each 2 BP values was assigned a 1 if the regimen was increased during that period, or a 0 if it was not. Multiple increases during a single period were counted as a 1. A subset of 42 patients, representing 495 (5%) of all clinic visits, were randomly selected for blind reabstraction by another author (D.R.B.). Agreement between the 2 reviewers was excellent (κ=0.93, 95% CI 0.87 to 0.98).

Independent Variable: Treatment Intensity Score

We characterized TI using an observed-expected scoring system originally described by Okonofua et al.[3] We have shown that this scoring system is a valid predictor of BP control over time and is the preferred scoring system to measure TI in the care of hypertension.[15] One of the strengths of this measure is that it avoids confounding by severity, the tendency for patients with more severe disease to receive more intensive management.[15] Without accounting for confounding by severity, one can obtain the paradoxical result that more intensive management is associated with worse control of BP.[15] Because this TI measure inherently accounts for BP control, it is not necessary to also control for initial BP as a covariate.

For this TI measure, a medication increase is expected on each occasion when the recorded BP is 140/90 mm Hg or higher. Using this number, and the number of occasions on which the regimen was intensified, each patient was assigned a score between −1 and 1, using the following formula:

$$(\text{observed medication changes} - \text{expected medication changes})/\text{number of clinic visits}$$

As an example, over a period of 10 visits, 5 of which had an elevated BP value, a patient would have an expected proportion of visits with medication increases of 5/10. If this patient actually had 3 visits with medication increases, the score would be 3/10−5/10=−0.2, indicating that therapy was increased at 20% fewer visits than expected. If the patient had 6 visits with therapy increases, the score would be 6/10−5/10=0.1, indicating that therapy was increased at 10% more visits than expected.

We recognize that for patients with diabetes or chronic kidney disease, current guidelines set a lower BP target (ie, 130/80 mm Hg).[13] We therefore created an additional TI score only for patients with a low BP target. For this alternative TI score, a medication increase was expected on each occasion when the recorded BP is 130/80 mm Hg or higher, as opposed to 140/90 mm Hg for the main TI score. We conducted a sensitivity analysis, dividing the sample into patients with the higher and the lower BP thresholds, and repeating our analyses for each group separately using the appropriate TI score. Results were similar to our main analysis, and are not shown.

Stratification Variable: Adherence to Antihypertensive Therapy

We characterized adherence to antihypertensive therapy using Medication Events Monitoring System ([MEMS], AARDEX). These devices use a microchip to record all bottle openings. Adherence as measured by MEMS caps has been linked to improvements in numerous clinical outcomes,[16,17] including hypertension control.[18,19] Patients were each given one MEMS cap, corresponding to the antihypertensive medication that they took the most times per day. Clinicians were not given feedback about their patients' adherence as measured by MEMS caps.

When processing MEMS data into adherence scores, we began by identifying all patients who either did not return their MEMS cap or did not open it enough times to calculate an adherence score (for example, once). For all others, we used MEMS data from the first 90 days after they began using their MEMS cap, or a shorter period for patients who stopped using their MEMS cap sooner. We calculated the proportion of days in this period on which the patient took at least the number of doses prescribed. Patients who did not return their MEMS caps were considered to have "missing" adherence. The remaining patients were divided into quartiles by adherence; thus, there were 5 adherence groups included in the analysis: 4 quartiles and "missing."

Covariates

We collected patient demographic data, including self-reported race (black or white), sex, and age. Using both ICD-9 codes and problem lists from the EMR, we noted whether the patients had the following comorbid conditions, all of which could impact the blood pressure, the use of antihypertensive medications, or the perceived urgency of controlling hypertension: benign prostatic hypertrophy, cerebrovascular disease, chronic heart failure, chronic kidney disease, coronary artery disease, diabetes mellitus, hyperlipidemia, obesity (BMI >30), and peripheral vascular disease. We noted whether patients were actively using tobacco at any time during the study.

Finally, we controlled for assignment to the intervention or control arm of the parent randomized trial as a covariate. Clinicians treating the patients in the study arm received a one-time educational intervention designed to improve doctor-patient communication and

Table 1. Baseline Characteristics of the Study Population (n=819)

Characteristic	Percentage or Mean Value
Mean age	59.6
Male sex	34%
Black race	58%
Current smoker	7%
Obese	59%
Comorbid conditions	
Benign prostatic hypertrophy	4%
Cerebrovascular disease	6%
Chronic heart failure	4%
Chronic kidney disease	7%
Coronary artery disease	13%
Diabetes	33%
Hyperlipidemia	54%
Peripheral vascular disease	5%
Frequency of clinic visits	
Mean person-time, months	24.3
Mean clinic visits	12.0
Mean clinic visits per month	0.49
Medication classes at baseline	
ACE Inhibitors or ARBs	65%
Beta blockers	45%
Calcium channel blockers	36%
Diuretics, thiazide, or loop	65%
All other classes combined	12%
Baseline No. of medications	
None	1%
1	25%
2	37%
3	25%
4 or more	13%
Baseline blood pressure control	
Mean baseline blood pressure, mm Hg	134/80
Baseline blood pressure<140/90 mm Hg	55%

cultural competency; clinicians treating patients in the control arm did not receive the intervention.

Statistical Analyses

We compared baseline characteristics among the 5 adherence groups, using ANOVA and χ^2 tests as appropriate. We used a test of linear trend to compare TI scores among the 5 adherence strata. We examined the effect of TI on the final SBP using a generalized linear model, controlling for patient-level covariates. We then added interaction terms to our model to test whether the effect of TI on the final SBP differed among the adherence strata, controlling for patient-level covariates. Finally, we analyzed each adherence stratum separately, controlling for covariates, to confirm that the effect of TI on SBP remained statistically significant in all strata. For all analyses, we used SAS, version 9.1 (SAS Institute).

Results

Patient Characteristics

Of the 869 patients enrolled in the study, 50 were not analyzed because they had 2 or fewer BP values. Therefore, 819 patients with hypertension, all managed at Boston Medical Center, constituted our study population (Table 1). The mean follow-up time was 24 months; on average, patients visited the clinic once every 2 months. The mean age was 59.6 years, 34% of patients were male, and most (58%) were of black race. Considering their relatively young age, the population had a relatively high burden of comorbidity: 33% had diabetes, 13% had coronary artery disease, 7% had chronic kidney disease, and 59% were obese. Most patients (74%) were receiving 2 or more antihypertensive medications at the beginning of the study. The population was characterized by relatively well-controlled hypertension at baseline: the mean initial BP was 134/80 mm Hg, and 55% of patients had an initial BP below 140/90 mm Hg.

There were 5 adherence groups: 4 quartiles of adherence (98% and higher, 94% to 98%, 80% to 94%, below 80%) and patients who did not return their MEMS caps (missing adherence). Within the poor adherence quartile, the median adherence was 62% (Interquartile Range 42% to 73%). Comparison of baseline characteristics among these 5 adherence strata revealed several differences (Table S1, please see http://hyper.ahajournals.org). Most notably, black race was associated with poorer adherence or not returning the MEMS cap; the best adherence group contained 45% black patients, compared to the worst adherence group (69%) and the missing adherence group (76%, $P<0.001$ for χ^2 test). In addition, patients with poor or missing adherence had worse BP control at baseline. For example, 45% of patients with missing adherence and 50% of patients with the worst adherence had controlled BP at baseline, compared to 61% among patients with the best adherence (probability value for χ^2 test=0.03).

Table 2. Mean Treatment Intensity (TI) Score After Stratifying by Quartiles of Adherence to Therapy

Group (% of Days Adherent)	n	Mean TI Score*
Best adherence (>98%)	168	−0.24
Good adherence (93% to 98%)	168	−0.26
Fair adherence (80% to 93%)	173	−0.26
Worst adherence (<80%)	162	−0.33
Missing adherence	148	−0.33
Test of linear trend	…	0.002

*Mean TI score for entire sample (n=819) was −0.28. A difference of 0.1 in the TI score indicates one more therapy increase than predicted per 10 visits.

Treatment Intensity, Adherence, and Blood Pressure Control

Blood pressure was elevated at 4894 of 11 530 clinic visits (42%), and therapy was increased at 7.4% of 11,530 visits. The median TI score was −0.25 (IQR −0.06, −0.50); the mean was −0.28 (SD 0.29). Among the 671 patients with complete adherence data, the average patient was adherent on 85% of days (median 94%, interquartile range 80% to 98%). Patients with better adherence received more intensive management (Table 2). The difference in the mean TI between the best and worst adherence quartiles was 0.09, approximately equivalent to 1 extra therapy increase per 11 clinic visits.

Table 3. Effect of Treatment Intensity Score on Final Systolic Blood Pressure

Adherence Group	Adjusted Effect*	*P* Value†
Best adherence (>98%)	−2.1	…
Good adherence (93% to 98%)	−1.8	0.49
Fair adherence (80% to 93%)	−2.3	0.73
Worst adherence (<80%)	−2.4	0.55
Missing adherence	−1.6	0.22

Interaction terms were used to test whether the effect sizes in patients with suboptimal adherence differed from the effect size among patients in the top quartile of adherence (n=819).

*Analyses adjusted for demographics, comorbid conditions, and treatment assignment (intervention vs control). All beta coefficients are expressed in mm Hg. Effect of TI is per change of 0.1 in the treatment intensity score (equivalent to one additional therapy increase per 10 visits). For example, a beta coefficient of −2.0 means that for every additional therapy increase per 10 visits, the mean final systolic blood pressure will be 2.0 mm Hg lower.

†*P* values for adherence strata test for a difference from the excellent adherence group. The effect of the entire TI variable was statistically significant ($P<0.001$). In addition, when each adherence stratum was analyzed separately, the effect of TI was statistically significant.

In the entire sample of 819 patients, each additional therapy increase per 10 visits predicted a 2 mm Hg decrease in the final SBP, after adjusting for covariates ($P<0.001$). We added interaction terms (Table 3) to reflect membership in the other adherence groups, compared to the reference category (best adherence). The effect size in the best adherence group was a 2.1 mm Hg decrease in SBP for each additional therapy increase per 10 visits. The effect sizes in the other adherence groups were 1.8 mm Hg in the second quartile, 2.3 mm Hg in the third quartile, 2.4 mm Hg in the fourth (worst) adherence quartile, and 1.6 mm Hg among patients with missing adherence. These effect sizes did not differ from that of the best adherence group at the 0.05 level of significance. In addition, we reran the multivariate regression separately for each adherence stratum; the effect of TI on final SBP remained statistically significant for each stratum ($P=0.01$ for missing adherence and $P<0.001$ for all other groups).

We also explored the effect of TI for patients with even lower adherence to therapy than the worst adherence quartile: less than 60% adherence (n=75). The effect of TI in that group, controlling for covariates, was similar to our other analyses (final SBP 2.0 mm Hg lower for each additional therapy increase per 10 visits, $P=0.006$).

Discussion

In this observational study, we investigated the interaction of adherence and TI in determining BP control. We found that more adherent patients received somewhat more intensive management, suggesting that clinicians may hesitate to intensify therapy in the face of suspected nonadherence. We also found that greater TI was associated with improved BP control over time, and that this effect was similar in size for patients with varying levels of adherence. This is a nonintuitive finding, and one which may surprise many. We would suggest that the key to understanding this finding is to remember that adherence is not a binary concept, with patients divided into those who are "adherent" and those who are "nonadherent." In our study, even patients with the worst adherence generally took approximately half their doses of medication. Many antihypertensive medications have long half-lives, and drugs with long half-lives may have a degree of "forgiveness" when some doses are missed.[20] Previous studies have shown that blood pressure response to many antihypertensives persists for several days after the last dose was taken, although the period of "forgiveness" varies among drugs.[21]

Many clinicians address suspected nonadherence by asking the patient to improve adherence, and then rechecking the BP at the next visit. This strategy may well reduce treatment intensity over time, especially if another reason not to intensify therapy is found at the following visit.[2,3,22] Our results suggest that, whereas clinicians in our study were less likely to intensify therapy in patients with suboptimal adherence, they could have improved these patients' BP control considerably by intensifying therapy. We do not mean to suggest that it is not worthwhile to address suboptimal adherence—the evidence is quite clear that greater adherence improves BP control.[4,5] However, it is notoriously difficult and effort-intensive to improve adherence, and not all patients will respond to such efforts.[23,24] Indeed, we know that clinicians often are not even aware of issues with adherence.[25–27] Although improving adherence remains an important priority, our results suggest that clinicians need not reserve therapy increases for patients with ideal adherence to therapy.

Our study population, in general, had a relatively high degree of adherence to therapy, which some might find surprising among an urban safety net population. It is important to note, however, that previous studies have recorded similar degrees of adherence to antihypertensive medications. For example, Choo et al studied patients in a managed care organization in Massachusetts and found that the mean percentage of days with adherence was 86%, and the median was 92% (IQR 0.77 to 0.98).[28] By comparison, we found a mean adherence of 84% and a median of 94% (IQR 0.80 to 0.98). In another study, Fung et al found that 27% of Medicare+Choice beneficiaries were poorly adherent, defined as taking less than 80% of their medication[29]; in our study, 24% of patients were less than 80% adherent. These comparisons remind us that divergent patient populations can have very similar patterns of adherence, and suggest that our results may be broadly generalizable to other populations.

Our study has several limitations. First, although MEMS caps have strengths as a measure of adherence,[26–28,30–32] they also have weaknesses.[28,33] Patients may take their medication more often than MEMS data would suggest, particularly if they are using some other sort of pill box rather than the bottle used for the MEMS cap.[33] We made efforts to minimize this effect, excluding patients from our study who stated that they use a pill organizer, but it is still possible that some patients identified as very poorly adherent in our study were actually quite adherent to their medication, but not to using the MEMS cap. Similarly, we cannot fully characterize adherence among patients who did not return their MEMS caps. However, the fact that these patients had higher BP at baseline than those with complete MEMS data supports the contention that these patients may have had the worst adherence of all. In any

event, patients with incomplete MEMS data also benefited from TI.

Second, this study did not examine definitive outcomes of care such as cardiovascular events or mortality. However, improved BP control (an intermediate outcome) has robustly been tied to improvements in morbidity and mortality.[13] In addition, it is possible that patients whose therapy was intensified despite nonadherence experienced some episodes of hypotension, a commonly raised concern in such a situation. This would raise concerns that, although more intensive management of hypertension in suboptimally adherent patients might lower BP, it might also increase risk for adverse events. However, there were no hypotensive episodes reported to study staff by patients or clinicians.

Third, this study shares the limitations of any observational study. Although our results suggest that patients with less-than-ideal adherence do benefit from intensification of the antihypertensive regimen, it cannot determine the ideal management for a nonadherent patient with hypertension. A randomized trial could assign nonadherent patients to intensification, adherence interventions, both, or neither, and would be ideally suited to answer this question. Fourth, we had few, if any, patients in our study who took none of their medication at all. Our results may not apply to such uncommon patients, and we would agree that intensifying antihypertensive therapy for such a patient would not be beneficial. Fifth, our study enrolled only patients with prevalent as opposed to incident hypertension. Therefore, our findings may not be generalizable to patients with newly diagnosed hypertension, who may have different patterns of adherence. Sixth, this study relies on data from one medical center, which may not be representative of other settings. Boston Medical Center is an academic, inner-city safety net hospital. Its academically oriented clinicians and largely immigrant and poor patient population are a somewhat unique combination. These results remain to be confirmed in other settings.

Finally, there are many legitimate reasons why a clinician-patient dyad might decide not to intensify therapy, including competing priorities, medication side effects, and patient unwillingness to accept a more intensive regimen. We do not mean to suggest that intensifying therapy is always the correct response to an elevated BP value. Rather, our study suggests that, when therapy intensification is mutually acceptable to the patient and the clinician, and there are no other reasons not to intensify, then suboptimal adherence alone is not a sufficient reason to forego intensification. Although it is important to communicate effectively about adherence and to try to improve it, it is not necessary to await proof of perfect adherence before intensifying therapy for hypertension.

Perspectives

In this observational study, more intensive management of hypertension improved blood pressure control to a similar extent regardless of the patient's level of adherence. The findings of this study do not diminish the importance of identifying patients with suboptimal adherence and trying to help them improve their adherence, because adherence remains an unquestioned determinant of control for hypertension and numerous other conditions. However, this study does call into question the widely held assumption that "nonadherent" patients cannot benefit from therapy intensification. Indeed, one of the major contributions of this study is to remind us that adherence is not a binary concept, with patients divided into those who are "adherent" or "nonadherent." Instead, all patients should be viewed as somewhere on a spectrum of adherence. The issue that we examined (ie, whether patients with uncontrolled hypertension and suboptimal adherence benefit from therapy intensification) has not previously been subjected to investigation because the answer was widely assumed. Now that this assumption has been challenged, we think it is time for further studies, particularly randomized trials, to determine the most effective management strategy for patients with uncontrolled hypertension and suboptimal adherence.

Acknowledgments

The authors thank Arlene Ash, PhD, and Allen Gifford, MD, for their valuable comments on earlier drafts of this manuscript.

Sources of Funding

This research was supported by a grant from the National Institutes of Health (HL072814, N.R. Kressin, PI). Dr Rose is supported by a career development award from the Department of Veterans Affairs, Health Services Research and Development Service. Dr Kressin is supported in part by a Research Career Scientist award from the Department of Veterans Affairs, Health Services Research & Development Service (RCS 02-066-1). The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

Disclosures

None.

References

1. Hypertension Detection and Follow-up Program Cooperative Group. Five-year findings of the hypertension detection and follow-up program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension. *JAMA*. 1979;242:2562–2571.
2. Berlowitz DR, Ash AS, Hickey EC, Friedman RH, Glickman M, Kader B, Moskowitz MA. Inadequate management of blood pressure in a hypertensive population. *N Engl J Med*. 1998;339:1957–1963.
3. 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;47:345–351.
4. DiMatteo MR, Giordani PJ, Lepper HS, Croghan TW. Patient adherence and medical treatment outcomes: a meta-analysis. *Med Care*. 2002;40:794–811.
5. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care*. 1986;24:67–74.
6. Grant RW, Singer DE, Meigs JB. Medication adherence before an increase in antihypertensive therapy: a cohort study using pharmacy claims data. *Clin Ther*. 2005;27:773–781.
7. Heisler M, Hogan MM, Hofer TP, Schmittdiel JA, Pladevall M, Kerr EA. When more is not better: treatment intensification among hypertensive patients with poor medication adherence. *Circulation*. 2008;117:2884–2892.
8. Ho PM, Magid DJ, Shetterly SM, Olson KL, Peterson PN, Masoudi FA, Rumsfeld JS. Importance of therapy intensification and medication nonadherence for blood pressure control in patients with coronary disease. *Arch Intern Med*. 2008;168:271–276.
9. Rose AJ, Berlowitz DR, Orner MB, Kressin NR. Understanding uncontrolled hypertension: is it the patient or the provider? *J Clin Hypertens (Greenwich)*. 2007;9:937–943.
10. Schmittdiel JA, Uratsu CS, Karter AJ, Heisler M, Subramanian U, Mangione CM, Selby JV. Why don't diabetes patients achieve recom-

mended risk factor targets? Poor adherence versus lack of treatment intensification. *J Gen Intern Med*. 2008;23:588–594.
11. Valderas JM. Treatment intensification does not always lead to better quality of care in patients with hypertension. *Ann Intern Med*. 149:362, 2008; author reply 362–363.
12. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med*. 2005;353:487–497.
13. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr, Jones DW, Materson BJ, Oparil S, Wright JT Jr, Roccella EJ. Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension*. 2003;42:1206–1252.
14. Rose AJ, Shimada SL, Rothendler JA, Reisman JI, Glassman PA, Berlowitz DR, Kressin NR. The accuracy of clinician perceptions of "usual" blood pressure control. *J Gen Intern Med*. 2008;23:180–183.
15. Rose AJ, Berlowitz DR, Manze M, Orner MB, Kressin NR. Comparing methods of measuring treatment intensification in hypertension care. *Circ Cardiovasc Qual Outcomes*. 2009; epub June 16, 2009. DOI: 10.1161/CIRCOUTCOMES.108.838649.
16. Dobbels F, De Geest S, van Cleemput J, Droogne W, Vanhaecke J. Effect of late medication non-compliance on outcome after heart transplantation: a 5-year follow-up. *J Heart Lung Transplant*. 2004;23:1245–1251.
17. Kimmel SE, Chen Z, Price M, Parker CS, Metlay JP, Christie JD, Brensinger CM, Newcomb CW, Samaha FF, Gross R. The influence of patient adherence on anticoagulation control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Arch Intern Med*. 2007;167:229–235.
18. Lee JY, Kusek JW, Greene PG, Bernhard S, Norris K, Smith D, Wilkening B, Wright JT Jr. Assessing medication adherence by pill count and electronic monitoring in the African Am Study of Kidney Disease and Hypertension (AASK) Pilot Study. *Am J Hypertens*. 1996;9:719–725.
19. Qureshi NN, Hatcher J, Chaturvedi N, Jafar TH. Effect of general practitioner education on adherence to antihypertensive drugs: cluster randomised controlled trial. *BMJ*. 2007;335:1030.
20. Urquhart J. Pharmacodynamics of variable patient compliance: implications for pharmaceutical value. *Adv Drug Deliv Rev*. 1998;33:207–219.
21. Girvin BG, Johnston GD. Comparison of the effects of a 7-day period of non-compliance on blood pressure control using three different antihypertensive agents. *J Hypertens*. 2004;22:1409–1414.
22. Phillips LS, Branch WT, Cook CB, Doyle JP, El-Kebbi IM, Gallina DL, Miller CD, Ziemer DC, Barnes CS. Clinical inertia. *Ann Intern Med*. 2001;135:825–834.
23. Krousel-Wood M, Hyre A, Muntner P, Morisky D. Methods to improve medication adherence in patients with hypertension: current status and future directions. *Curr Opin Cardiol*. 2005;20:296–300.
24. Kressin NR, Wang F, Long J, Bokhour BG, Orner MB, Rothendler J, Clark C, Reddy S, Kozak W, Kroupa LP, Berlowitz DR. Hypertensive patients' race, health beliefs, process of care, and medication adherence. *J Gen Intern Med*. 2007;22:768–774.
25. Bokhour BG, Berlowitz DR, Long JA, Kressin NR. How do providers assess antihypertensive medication adherence in medical encounters? *J Gen Intern Med*. 2006;21:577–583.
26. Byerly M, Fisher R, Whatley K, Holland R, Varghese F, Carmody T, Magouirk B, Rush AJ. A comparison of electronic monitoring vs. clinician rating of antipsychotic adherence in outpatients with schizophrenia. *Psychiatry Res*. 2005;133:129–133.
27. Waterhouse DM, Calzone KA, Mele C, Brenner DE. Adherence to oral tamoxifen: a comparison of patient self-report, pill counts, and microelectronic monitoring. *J Clin Oncol*. 1993;11:1189–1197.
28. Choo PW, Rand CS, Inui TS, Lee ML, Cain E, Cordeiro-Breault M, Canning C, Platt R. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. *Med Care*. 1999;37:846–857.
29. Fung V, Huang J, Brand R, Newhouse JP, Hsu J. Hypertension treatment in a Medicare population: adherence and systolic blood pressure control. *Clin Ther*. 2007;29:972–984.
30. Burnier M, Schneider MP, Chiolero A, Stubi CL, Brunner HR. Electronic compliance monitoring in resistant hypertension: the basis for rational therapeutic decisions. *J Hypertens*. 2001;19:335–341.
31. Guerrero D, Rudd P, Bryant-Kosling C, Middleton B, Middleton BF. Antihypertensive medication-taking. Investigation of a simple regimen. *Am J Hypertens*. 1993;6:586–592.
32. Hamilton GA. Measuring adherence in a hypertension clinical trial. *Eur J Cardiovasc Nurs*. 2003;2:219–228.
33. Samet JH, Sullivan LM, Traphagen ET, Ickovics JR. Measuring adherence among HIV-infected persons: Is MEMS consummate technology? *AIDS Behav*. 2001;5:21–30.

Online Supplement

Intensifying Therapy for Hypertension despite Suboptimal Adherence

Adam J. Rose, MD MSc[1,2]
Dan R. Berlowitz, MD MPH[1,2,3]
Meredith Manze, MPH[2,3]
Michelle B. Orner, MPH[1]
Nancy R. Kressin, PhD[1,2,4]

From the [1]Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Bedford, MA; [2]Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA; [3]Department of Health Policy and Management, Boston University School of Public Health, Boston, MA; [4]VA Boston Healthcare System, Boston, MA.

Corresponding Author and Requests for Reprints:

Adam J. Rose, MD MSc
Center for Health Quality, Outcomes, and Economic Research
200 Springs Road, Building #70
Bedford, MA 01730
Telephone: (617) 990-2530
Fax: (917) 591-3104
Email: adamrose@bu.edu

Figure S1. Flow chart of patient recruitment.

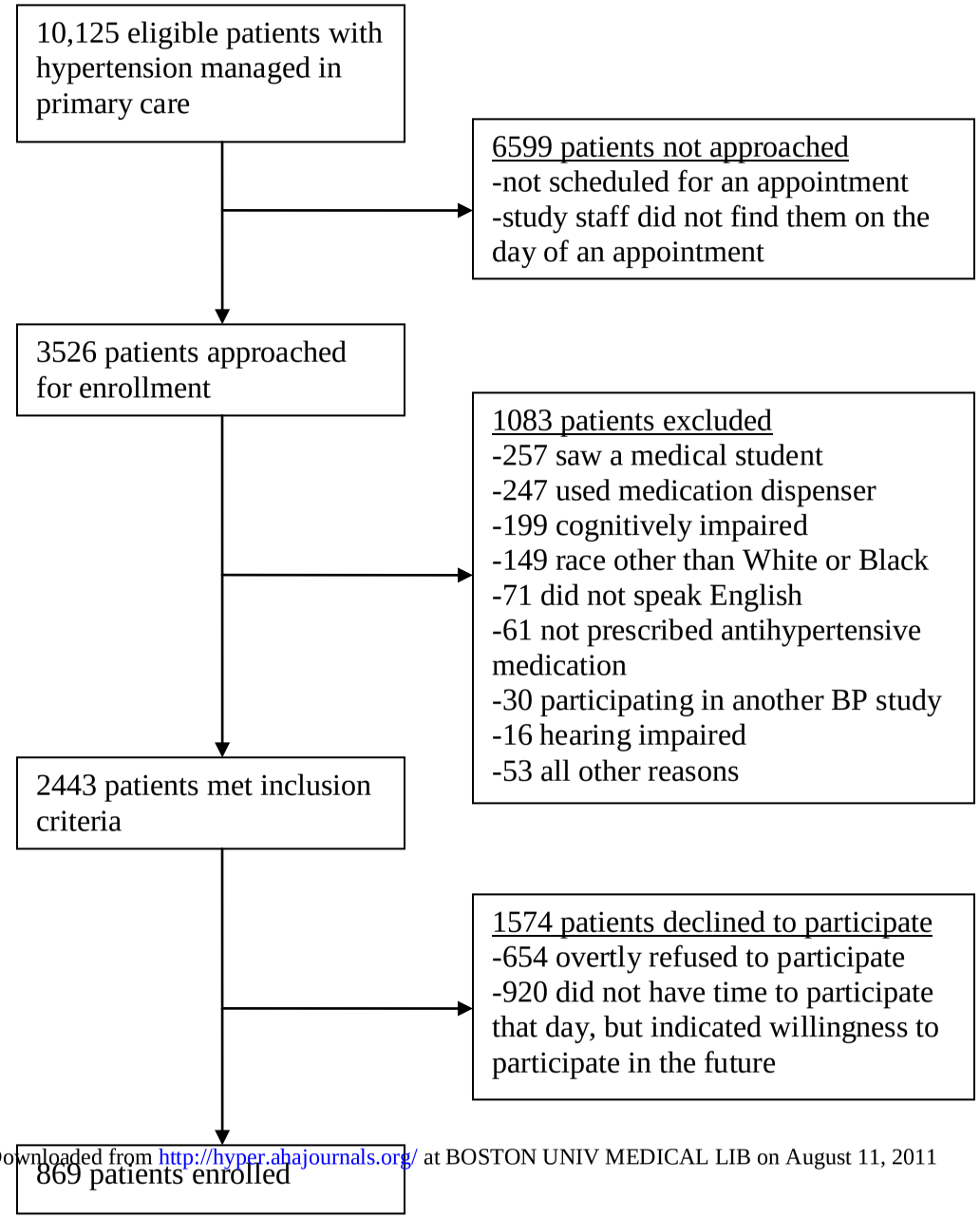


Table S1. Comparison of baseline characteristics among the 5 adherence strata (total n = 819).

Characteristic	Adherence > 98% (n =168)	Adherence 94% - 98% (n = 168)	Adherence 80% - 94% (n = 173)	Adherence < 80% (n = 162)	Missing Adherence (n = 148)	p-value*
Mean Age	62.4	62.3	58.2	57.5	57.3	<0.001
Male Gender	38%	36%	30%	32%	33%	0.53
Black Race	45%	46%	57%	69%	76%	< 0.001
Current Smoker	2%	7%	6%	11%	12%	0.003
Obese	60%	52%	57%	67%	59%	0.08
Comorbid Conditions						
Benign Prostatic Hypertrophy	7%	4%	3%	2%	3%	0.18
Cerebrovascular Disease	5%	5%	6%	4%	7%	0.76
Chronic Heart Failure	2%	1%	3%	7%	6%	0.01
Chronic Kidney Disease	4%	6%	7%	8%	9%	0.50
Coronary Artery Disease	12%	12%	15%	13%	11%	0.89
Diabetes	26%	32%	35%	41%	31%	0.054
Hyperlipidemia	57%	58%	54%	51%	47%	0.21
Peripheral Vascular Disease	7%	7%	5%	6%	3%	0.52
Frequency of Clinic Visits						
Mean Person-Time, Months	24.1	24.6	24.6	23.3	24.7	0.45
Mean Clinic Visits	11.9	11.5	11.7	12.2	12.7	0.87
Mean Clinic Visits/Month	0.48	0.47	0.47	0.52	0.52	0.62
Medication Classes at Baseline						
ACE Inhibitors or ARBs	67%	64%	66%	67%	60%	0.73
Beta Blockers	36%	46%	47%	52%	45%	0.08
Calcium Channel Blockers	39%	27%	35%	44%	37%	0.02
Diuretics, Thiazide or Loop	65%	61%	65%	69%	66%	0.67
All Other Classes Combined	7%	10%	11%	17%	14%	0.047
Baseline Number of Medications						0.01
None	1%	1%	1%	1%	0%	
1	24%	33%	23%	22%	25%	

2	42%	34%	41%	28%	37%	
3	26%	21%	24%	28%	25%	
4 or more	8%	11%	12%	22%	13%	
Baseline BP Control						
Baseline SBP, mm/Hg	132	133	132	135	137	0.04
Baseline DBP, mm/Hg	77	79	81	82	83	< 0.001
Baseline BP < 140/90 mm/Hg	61%	57%	58%	50%	45%	0.03

BP: Blood Pressure
SBP: Systolic Blood Pressure
DBP: Diastolic Blood Pressure
*Comparisons of continuous variables are by ANOVA test. Comparisons of dichotomous variables are by chi-square test.

Does Opioid Therapy Affect Quality of Care for Diabetes Mellitus?

Adam J. Rose, MD, MSc; John A. Hermos, MD; Susan M. Frayne, MD, MPH; Leonard M. Pogach, MD, MBA; Dan R. Berlowitz, MD, MPH; and Donald R. Miller, ScD

Objective: To examine whether veterans who received chronic opioid therapy had worse diabetes performance measures than patients who did not receive opioids.

Study Design: Retrospective cohort study.

Methods: We identified all patients with diabetes mellitus receiving care in US Department of Veterans Affairs facilities during 2004. Cases received at least 6 prescriptions for chronic opioids during 2004, while controls were randomly selected from among patients with diabetes who received no opioids. We compared process measures (glycosylated hemoglobin and low-density lipoprotein cholesterol levels tested and an eye examination performed) and outcome measures (glycosylated hemoglobin level ≤9.0% and low-density lipoprotein cholesterol level ≤130 mg/dL) between groups.

Results: Cases (n = 47,756) had slightly worse diabetes performance measures than controls (n = 220,912) after adjustment for covariates. For example, 86.4% of cases and 89.0% of controls had a glycosylated hemoglobin test during fiscal year 2004 (adjusted odds ratio, 0.69; P <.001). Among cases, receipt of higher-dose opioids was associated with additional decrement in diabetes performance measures, with a dose-response relationship.

Conclusions: Chronic opioid therapy among patients within the Veterans Affairs system is associated with slightly worse diabetes performance measures compared with patients who do not receive opioids. However, patients receiving higher dosages of opioids had additional decrements in diabetes performance measures; these patients may be appropriate targets for interventions to improve their care for pain and diabetes.

(Am J Manag Care. 2009;15(4):217-224)

For author information and disclosures, see end of text.

Understanding the role of clinical complexity as a determinant of quality of care is a major research goal.[1] In previous studies,[2-9] the effect of clinical complexity on quality of care has varied depending on the diagnoses, the clinician and patient, and the clinical environment. Seeking to harmonize these mixed results into a unifying theory, Piette and Kerr[10] proposed that symptomatic conditions may have a greater effect on quality of care than asymptomatic conditions and that conditions with dissimilar management goals ("discordant conditions") may have a greater effect than those with similar goals ("concordant conditions").

By this reasoning, chronic pain could have a considerable adverse effect on quality of care for unrelated conditions. Pain is highly symptomatic, and pain management is discordant with the management of other conditions.[11] While the use of opioids to treat chronic noncancer pain is increasingly accepted,[12] opioid therapy may present additional challenges due to the potential for abuse, dependence, and diversion and due to conflicts over appropriate dosages.[13-19] However, opioid therapy could also facilitate care for unrelated conditions. Patients receiving opioids may visit the clinic more often, allowing more opportunities for medical management.[10] Adequate treatment of pain may improve the patient's functional status and quality of life,[12] allowing greater focus on self-care activities.

Diabetes mellitus, a common, costly, and highly morbid condition,[20,21] is a good condition in which to examine this possibility. Adequate management of diabetes requires collaboration among clinicians and the patient within a system of care,[22-27] and explicit guidelines and diabetes performance targets exist with which to examine the adequacy of diabetes care.[28-30] Krein et al[31] showed that among patients with diabetes, chronic pain is a barrier to the completion of self-care activities such as taking medications, exercising, and pursuing a prudent diet. However, the effect of pain on process and outcome measures of diabetes care is unknown. In addition, no study has specifically examined the effect of opioid therapy on the quality of care for unrelated chronic conditions, but there is reason to believe that opioid therapy may impart more complexity and challenge than pain alone.[32]

To clarify whether the net effect of opioid therapy is to promote or impede care for diabetes, we analyzed a large database of patients with diabetes in the US Department of Veterans

In this issue
Take-Away Points / p222
www.ajmc.com
Full text and PDF

Affairs (VA) and identified those receiving chronic opioid therapy. We compared patients receiving chronic opioids versus patients not receiving opioids regarding selected diabetes performance measures. We hypothesized that the distractions and concerns associated with chronic opioid therapy, as well as perhaps other characteristics of patients with chronic pain, would be reflected in worse diabetes performance measures. We also hypothesized that among those receiving opioids there would be a dose-response relationship between higher opioid dosages and decrements in diabetes performance measures.

METHODS

Study Sample

We identified subjects from the Diabetes Epidemiology Cohort, which comprises all patients with diabetes seen in the VA.[21] The Diabetes Epidemiology Cohort links administrative, laboratory, and pharmacy data from the VA with Medicare claims, providing a rich data set for analysis.[21,33] We first looked at all veterans treated for diabetes during fiscal year (FY) 2004 whose diabetes had been diagnosed before the start of FY 2002. Based on earlier work,[21] we defined patients as having diabetes if they had at least 2 *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) codes for diabetes or any prescriptions for antiglycemic medications within a 2-year period.

We excluded patients who had an *ICD-9-CM* diagnosis of malignant neoplasm (other than basal or squamous carcinoma of the skin) within 2 years of study inception. The management of cancer-related pain is qualitatively different; moreover, diabetes performance measures may not apply to patients with active malignant neoplasms. We also excluded all patients receiving methadone hydrochloride or buprenorphine hydrochloride–naloxone hydrochloride for treatment of opioid dependence. Finally, we excluded patients who had fewer than 2 VA primary care visits in FY 2004, as a large portion of their diabetes care may not appear in our database.

This study was approved by the Institutional Review Board of Bedford VA Medical Center.

Independent Variable: Chronic Opioid Therapy

Our independent variable was the prescription of chronic opioids. We considered the following "major" opioids: codeine, fentanyl citrate, hydrocodone, hydromorphone hydrochloride, methadone, morphine sulfate, and oxycodone; all are Schedule II or III controlled substances according to the US Drug Enforcement Administration.[34] Any formulation suitable for outpatient administration was considered, including tablets, patches, elixirs, and sprinkles. We also included formulations that combine opioids with other drugs such as acetaminophen. Buprenorphine, butorphanol, nalbuphine hydrochloride, pentazocine, and propoxyphene, which are less potent, were considered "minor" opioids.[34]

Patients who received at least 6 prescriptions for major opioids during FY 2004, with or without additional minor opioids, constituted the chronic opioid group (cases). This cutoff of 6 prescriptions was chosen to distinguish treatment for chronic pain from treatment for acute pain and is consistent with previous definitions of chronic pain.[17,18] Patients who received any major or minor opioids during FY 2004 but did not meet criteria for the case group were excluded from the study. We randomly selected controls from among the remaining patients, who had received no opioids during FY 2004, to achieve a control group approximately 4 times as numerous as the case group.

Dependent Variables: Diabetes Performance Measures

Our 3 process measures, which could be completed at any time during FY 2004, were testing of glycosylated hemoglobin (A1C) level, testing of low-density lipoprotein cholesterol (LDL-C) level, and a dilated eye examination. Our 2 outcome measures were at least 1 A1C level of 9.0% or less and at least 1 LDL-C level of 130 mg/dL or less during FY 2004 (to convert A1C level to proportion of total hemoglobin, multiply by 0.01; to convert cholesterol level to millimoles per liter, multiply by 0.0259). If no test results were available among VA data, patients were considered to have levels above these thresholds. These diabetes performance measures are based on VA clinical practice guidelines for diabetes and reflect a minimal standard of care.[28,29] We also examined lower targets for glycemic and lipemic control (ie, A1C level ≤8.0% and LDL-C level ≤100 mg/dL).

Covariates

Age was divided into the following 4 categories: 54 years or younger, 55 to 64 years, 65 to 74 years, and 75 years or older. Race/ethnicity was categorized into the following 4 groups: white non-Hispanic, black non-Hispanic, all others, and missing. The VA priority status, which characterizes the degree of entitlement to VA care, was defined as follows: poverty, full disability, partial disability, or none of the above.

More or less intensive management of diabetes may be indicated depending on life expectancy and comorbidities.[29] We focused on the following complications of diabetes by identifying conditions with at least 1 *ICD-9-CM* code during FYs 1997 through 2004: cellulitis, gangrene/ulcer, other diabetic infections, congestive heart failure, other heart diseases, cerebrovascular disease, peripheral vascular disease, renal disease, and diabetic eye disease. Mental health conditions may also

affect diabetes care.[2-5] Using similar *ICD-9-CM* code–based definitions, we identified the following mental health conditions: major depression, bipolar disorder, anxiety disorders, posttraumatic stress disorder, and schizophrenia. We also recorded the number of VA primary care visits; more visits might allow more opportunities to complete diabetes performance measures. We also examined pain diagnoses, dividing them into the following 4 broad categories: neuropathic pain, musculoskeletal pain, chronic headache, and psychogenic pain. Using *ICD-9-CM* codes, we categorized patients according to whether or not they had any diagnoses in each category (vs none).

We hypothesized that patients receiving higher daily doses of opioids might be at risk for additional decrements in diabetes performance measures, as the receipt of higher dosages suggests difficulties in pain management and possibly physiologic tolerance and an increased risk of dependence.[13,17,35] We used a standard equivalency table[36] to convert all opioid dosages to oral morphine equivalents. We calculated a mean daily dose of opioid therapy in FY 2004 for each patient in the study and categorized patients into quartiles based on their daily opioid doses.

Finally, we assigned each patient to 1 VA medical center so that we could control for site of care. Our assignment was based on the site the patient visited most often for diabetes care during FY 2004. If 2 sites were visited equally, we selected the site visited closest to the end of the year.

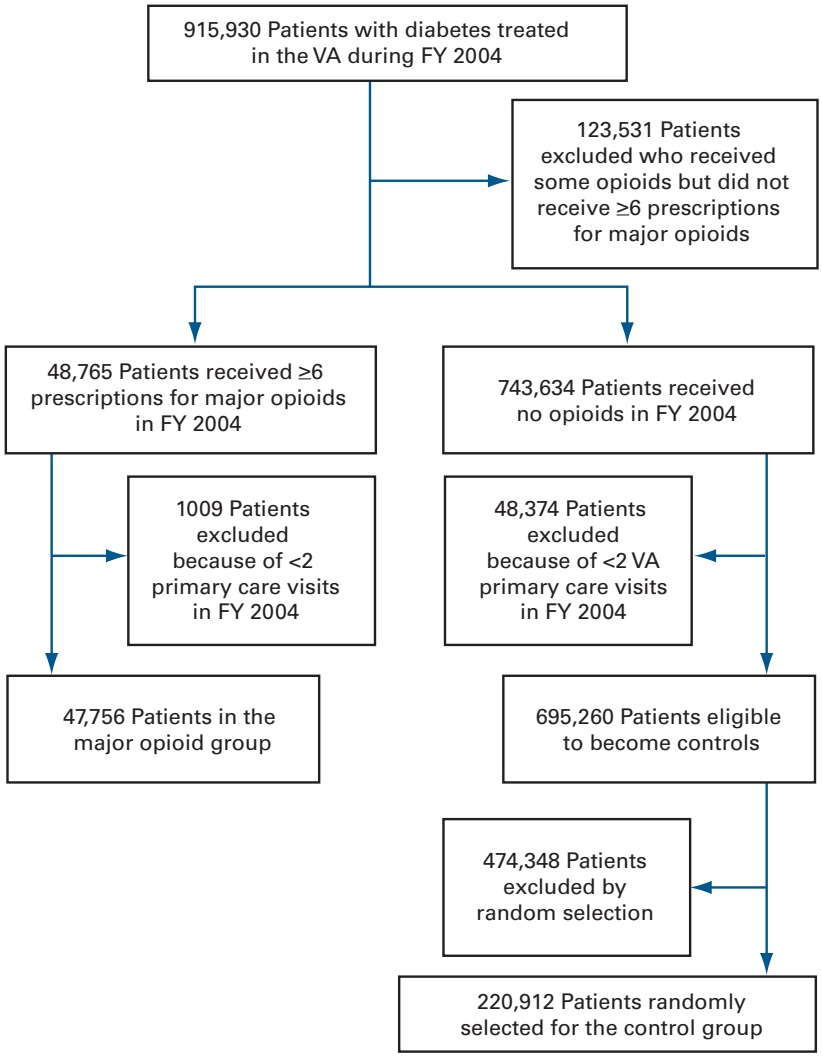
Statistical Analysis

We began our analysis with bivariate comparisons of demographics, comorbidities, and healthcare utilization between cases and controls. Using χ^2 tests, we then performed unadjusted comparisons of the proportions fulfilling each of the 5 diabetes performance measures. We performed adjusted analyses using generalized estimating equations to account for the clustering of outcomes by site of care, while adjusting for other covariates (sex, age, race/ethnicity, VA priority status, pain diagnoses, diabetic complications [including neuropathic pain], mental health conditions, and the number of VA primary care visits during the study). We did not adjust for eye disease when studying the eye examination process measure.

To investigate the possible effect of missing data on our results, we repeated key analyses among subsets of patients who were likely to have complete data. We restricted process measures to patients 65 years or older, who would presumably use Medicare when not using the VA and thus would have complete data for process measures. We restricted outcome measures to patients who had an A1C or LDL-C test within the VA at least once during the study (ie, those for whom laboratory values were available).

Finally, we added the mean daily opioid dose to our models and examined its ability to risk stratify the cases regarding diabetes performance measures. Our analyses were conducted us-

Figure. Inclusions and Exclusions for the Case and Control Groups



FY indicates fiscal year; VA, US Department of Veterans Affairs.

Table 1. Demographics of the Study Groups[a]

Demographic	% Cases (n = 47,756)	% Controls (n = 220,912)
Sociodemographics		
Male sex	96.8	98.0
Age group, y		
≤54	22.9	11.0
55-64	37.4	24.8
65-74	23.5	32.6
≥75	16.2	31.6
Race/ethnicity		
White non-Hispanic	75.9	68.4
Black non-Hispanic	11.8	12.0
All others	4.6	6.6
Missing	7.7	13.0
VA priority status		
Poverty	36.2	39.0
Full disability	41.2	20.4
Partial disability	15.1	17.0
None of the above	7.5	23.5
Comorbidities		
Pain conditions		
Musculoskeletal pain	85.6	48.9
Neuropathic pain	8.0	4.6
Chronic headache	0.9	0.3
Psychogenic pain	1.5	0.2
Diabetic complications		
Any	70.9	60.5
Cellulitis	26.2	16.0
Gangrene/ulcer	12.9	10.0
Other diabetic infections	9.9	4.4
Congestive heart failure	16.9	13.9
Other heart diseases	20.8	17.2
Cerebrovascular disease	12.9	10.5
Peripheral vascular disease	24.0	20.4
Renal disease	9.7	7.2
Diabetic eye disease	23.0	22.1
Mental health diagnoses		
Any	36.8	15.9
Major depression	22.6	10.0
Anxiety disorders	14.1	5.1
Posttraumatic stress disorder	6.6	2.3
Substance abuse disorders	6.2	1.7
Bipolar disorder	4.3	1.7
Schizophrenia	2.6	1.2
Healthcare utilization		
No. of primary care visits in FY 2004		
2-3	33.3	56.6
4-5	25.6	25.4
6-8	21.1	12.0
≥9	20.0	6.0
Any Medicare utilization in FY 2004	71.9	73.4

FY indicates fiscal year; VA, US Department of Veterans Affairs.
[a] P <.001 for all comparisons.

ing SAS version 9.1 (SAS Inc, Cary, NC).

RESULTS

Demographics

The Figure shows the exclusions that led to our case and control groups. We compared summary statistics between cases and controls (Table 1). Cases were younger (eg, 60.3% were <65 years vs 35.8% of controls). Despite their younger age, they were more likely to have at least 1 diabetic complication (70.9% vs 60.5%) and to have each particular diabetic complication. They were also more than twice as likely to have at least 1 mental health condition (36.8% vs 15.9%) and to have each specific mental health condition. They had more VA primary care visits than controls (eg, 41.1% vs 18.0% had ≥6 primary care visits). Despite the differences in age and VA primary care utilization, the 2 groups used Medicare at similar rates.

Among the cases (Table 2), most (67.4%) received only short-acting opioid formulations. The most commonly prescribed opioid was hydrocodone, followed by short-acting oxycodone, codeine, and long-acting morphine. Thirty-nine percent received more than 1 opioid during FY 2004. Among 89.4% of cases for whom dosage information was available, the mean total daily dose (in milligrams of morphine) was 88.9 mg. The median total daily dose was much lower (22.7 mg), indicating a rightward skew to this distribution. Opioid dose quartile was found to be highly

collinear with the duration of action of the drugs received. For example, patients receiving only short-acting drugs were unlikely to be in the highest-dose quartile (3.2%) compared with patients receiving only long-acting drugs (77.0%) or both long-acting and short-acting drugs (56.1%). All patients for whom dose quartile was unavailable were receiving only short-acting drugs; therefore, it seems likely that most of them were also receiving lower total daily doses.

Comparison of Diabetes Performance Measures

An unadjusted comparison of diabetes performance measures between cases and controls is given in Table 3. There were small differences between groups, all of which attained statistical significance because of large sample size. Among process measures, cases were less likely to have their A1C level tested (86.4% vs 89.0%) and to have their LDL-C level tested (75.9% vs 80.3%) but were more likely to have an eye examination performed (67.0% vs 66.3%). Among outcome measures, cases were slightly less likely to have A1C control (75.9% vs 76.5%) and LDL-C control (65.2% vs 66.1%).

To evaluate the effect of missing data on our results, we repeated our analysis of process measures among patients 65 years or older, whose data would presumably be more complete. While diabetes performance measures were slightly improved in both groups, between-group differences were unaffected. Similarly, for outcome measures, we restricted our analysis to patients who had at least 1 test within the VA. While diabetes performance measures improved considerably in both groups, between-group differences did not change appreciably (eg, A1C level ≤9.0% increased from 75.9% in cases and 76.5% in controls to 91.8% and 92.6%, respectively). Examination of stricter targets for glycemic control (A1C level ≤8.0%) and for lipemic control (LDL-C level ≤100 mg/dL) worsened diabetes performance measures, but between-group differences remained slight.

In unadjusted analyses that accounted for the clustering of outcomes by site of care (Table 3), odds ratios echoed the unadjusted proportions. After adjustment for covariates, cases now lagged behind controls for each of our 5 diabetes performance measures, with effect sizes somewhat widened (eg, the odds ratio for A1C level measurement was 0.78 before adjustment and 0.69 after adjustment). Between-group differences remained small for outcome measures even after adjustment for covariates (adjusted odds ratios, 0.90 and 0.87 for A1C level ≤9.0% and LDL-C level ≤130 mg/dL, respectively).

After stratification by opioid dosage, higher daily doses predicted worse performance on all 5 diabetes measures we studied (Table 4). For example, patients in the highest-dose quartile had an odds ratio of 0.55 for having their A1C level tested and an odds ratio of 0.79 for having an A1C level of 9.0% or less compared with controls.

Table 2. Opioid Therapy Among 47,756 Cases

Opioid Therapy	Value
Long-acting vs short-acting opioid, %	
Short acting only	67.4
Long acting only	9.5
Both long and short acting	23.0
Specific drugs, %[a]	
Short acting	
Hydrocodone	52.4
Oxycodone, short acting	34.1
Codeine	23.4
Morphine sulfate, short acting	4.8
Hydromorphone hydrochloride	1.1
Long acting	
Morphine, long acting	17.3
Methadone hydrochloride	8.4
Fentanyl citrate patch	7.2
Oxycodone, long acting	4.5
No. of different drugs received, %	
1	60.7
2	28.4
3	8.4
≥4	2.5
Total daily dose, mg of morphine[b]	
Mean [SD]	88.9 [824.4]
Median (interquartile range), mg of morphine	22.7 (11.0-55.7)

[a]Percentages exceed 100% because some patients received more than 1 kind of opioid.
[b]Dosage information was missing for 10.6% of patients.

DISCUSSION

Patients receiving chronic opioid therapy had only slightly worse diabetes performance measures than those not receiving opioids; the difference was smaller than we had anticipated. Within the opioid group, the receipt of higher daily opioid doses predicted further decrements in all of our diabetes performance measures, with a dose-response relationship. This suggests that the small difference in diabetes performance measures between cases and controls is largely attributable to patients receiving higher dosages of opioids. Resources should be focused on improving care for patients receiving high dosages of opioids. For example, a mean daily dose exceeding 60 mg of morphine (our highest-dose

Table 3. Comparison of Diabetic Performance Measures Between Cases and Controls

Variable	% Cases (n = 47,756)	% Controls (n = 220,912)	*P*	Odds Ratio (95% Confidence Interval)[a] Unadjusted	Odds Ratio (95% Confidence Interval)[a] Adjusted
Process measures					
A1C level measured	86.4	89.0	<.001	0.78 (0.72-0.86)	0.69 (0.63-0.76)
LDL-C level measured	75.9	80.3	<.001	0.77 (0.71-0.84)	0.71 (0.66-0.78)
Eye examination performed	67.0	66.3	.001	1.03 (0.99-1.08)	0.80 (0.77-0.84)
Outcome measures					
A1C level ≤9.0%	75.9	76.5	.006	0.97 (0.91-1.04)	0.90 (0.84-0.96)
LDL-C level ≤130 mg/dL	65.2	66.1	<.001	0.96 (0.89-1.04)	0.87 (0.82-0.94)
Stricter outcome measure targets					
A1C level ≤8.0%	67.7	68.4	.007	0.97 (0.92-1.02)	0.97 (0.91-1.02)
LDL-C level ≤100 mg/dL	50.5	49.4	<.001	1.05 (0.99-1.10)	0.98 (0.93-1.04)

A1C indicates glycosylated hemoglobin; LDL-C, low-density lipoprotein cholesterol.
SI conversion factors: To convert A1C level to proportion of total hemoglobin, multiply by 0.01; to convert cholesterol level to millimoles per liter, multiply by 0.0259.
[a]Odds ratios are the odds of completing diabetes performance measures among cases compared with controls. An odds ratio of less than 1 indicates that cases are less likely than controls to complete a diabetes performance measure. Unadjusted odds ratios account for the clustering of outcomes by site of care using general estimating equations. Adjusted odds ratios also account for sociodemographics, comorbidities, and number of primary care visits in FY 2004.

quartile) could trigger an automatic consultation with a nurse care manager.

The receipt of chronic opioid therapy is a compound concept that includes elements of pain, provider prescribing patterns, and medication utilization. Krein et al[31] previously showed that chronic pain distracts patients with diabetes from self-care tasks, including adherence to diet, exercise, and medication use. In that study, taking a medication for pain seemed to mitigate the negative effect of pain on some self-care activities, possibly because well-treated pain is less all-consuming.[31] To our knowledge, our study is the first to examine the relationship between a specific therapy for pain (opioid use) and process and outcome measures of diabetes care. Our study reminds us of the potential to use large clinical databases (such as those available within the VA) to answer meaningful questions about the care received by previously understudied groups of patients.

Our study has some limitations. First, we did not control for pain scores or for the severity of pain. It may be that our finding of a dose-response relationship for opioid therapy may partly or wholly reflect the effect of increasing severity of pain. Future research might be able to separate the effect of opioid therapy from that of pain severity, but given the subjective nature of any measure of pain, this would require detailed data and might still be of questionable validity.

Second, the highly integrated nature of VA care and the standardized clinical care provided throughout the VA may have minimized the detrimental effects of opioid therapy in our study. In non-VA settings, opioid therapy may affect diabetes care more negatively.

Third, our VA cohort was mostly male and had a high incidence of poverty, comorbidity, and disability. These factors limit the generalizability of our findings to other populations.

Fourth, we were unable to capture some elements of care that occurred outside the VA. However, our reanalysis of subsets of patients with complete data suggested that, while incomplete data capture affected rates of diabetes performance measure completion, it did not greatly alter between-group comparisons.

Fifth, our data set was insufficiently detailed to identify patients who were abusing prescription opioids. Defining prescription drug abuse in clinical practice is challenging,[37] identifying it from paper medical record review can be difficult,[38] and identifying it from automated data is even more problematic. It is plausible that our finding of a dose-response curve for opioid therapy is partially due to an

Take-Away Points

Within the Veterans Affairs system, patients who received opioids for chronic pain had slightly worse diabetes performance measures than patients who did not receive opioids.

- Comparisons included measurement of glycemic and lipemic control, achievement of moderate or better glycemic and lipemic control, and a yearly eye examination.
- Among the group receiving opioids, the receipt of higher daily opioid doses predicted worse results for all 5 diabetes performance measures. A dose-response relationship was observed, lending additional credibility to this finding.

Table 4. Completion of Diabetes Performance Measures Among 268,668 Subjects for Whom Opoid Dosage Information Was Available[a]

Variable	Opioid Dose Quartile (95% Confidence Interval)[b]					Test for Linear Trend *P*
	Controls (n = 220,912)	Lowest (n = 10,670)	Lower (n = 10,675)	Higher (n = 10,677)	Highest (n = 10,678)	
Process measures						
A1C level measured	1 [Reference]	0.82 (0.72-0.93)	0.74 (0.66-0.83)	0.63 (0.56-0.71)	0.55 (0.49-0.62)	<.001
LDL-C level measured	1 [Reference]	0.76 (0.68-0.84)	0.73 (0.66-0.81)	0.67 (0.61-0.74)	0.64 (0.57-0.72)	<.001
Eye examination performed	1 [Reference]	0.90 (0.84-0.96)	0.80 (0.75-0.85)	0.77 (0.73-0.83)	0.72 (0.67-0.77)	<.001
Outcome measures						
A1C level ≤9.0%	1 [Reference]	1.03 (0.94-1.11)	0.92 (0.85-1.00)	0.82 (0.75-0.90)	0.79 (0.71-0.88)	.002
LDL-C level ≤130 mg/dL	1 [Reference]	0.93 (0.85-1.02)	0.86 (0.79-0.93)	0.83 (0.76-0.89)	0.82 (0.74-0.90)	.003

A1C indicates glycosylated hemoglobin; LDL-C, low-density lipoprotein cholesterol.
SI conversion factors: To convert A1C level to proportion of total hemoglobin, multiply by 0.01; to convert cholesterol level to millimoles per liter, multiply by 0.0259.
[a]Results are adjusted for sociodemographics, pain diagnoses, comorbidities, and number of primary care visits in fiscal year 2004. Regression analyses were performed using generalized estimating equations to account for the clustering of outcomes by site of care.
[b]Opioid dosage information was missing for 10.6% of subjects. The other 89.4% were categorized into quartiles by dosage.

increasing prevalence of prescription drug abuse in the higher-dosage categories. In addition, quartiles of opioid dosage may reflect the physical tolerance that naturally develops over time with chronic opioid therapy, necessitating higher dosages of this medication.[39] A more complete examination of the effect of prescription drug abuse on the quality of care for unrelated conditions would ideally be conducted with a detailed paper medical record review rather than with automated data.

Sixth, our comorbidity data were not sufficiently detailed to identify patients for whom tight glycemic or lipemic control would not be indicated because of limited life expectancy. We addressed this by having modest expectations for glycemic and lipemic control (ie, an A1C level of ≤9.0% and an LDL-C level of ≤130 mg/dL). Although the application of more stringent standards to some patients may be of questionable value,[5,29,40-42] these targets should apply to most, if not all, patients.

In summary, patients receiving chronic opioids to treat pain had slightly worse diabetes performance measures than patients who did not receive opioids. However, stratification by opioid dosage revealed that patients receiving high dosages had additional decrements in diabetes performance measures. Efforts should be focused on improving the quality of care in such patients for pain and for diabetes.

Acknowledgment

We thank Qing Shao for his expert SAS programming.

Author Affiliations: Center for Health Quality, Outcomes, and Economic Research (AJR, DRB, DRM), Bedford VA Medical Center, Bedford, MA; Department of Medicine (AJR, JAH, DRB), Boston University, Boston, MA; Massachusetts Veterans Epidemiology Research and Information Center (JAH), VA Boston Healthcare System, Boston, MA; Center for Health Care Evaluation (SMF), VA Palo Alto Healthcare System, Palo Alto, CA; Division of General Internal Medicine (SMF), Stanford University, Stanford, CA; Center for Healthcare Knowledge Management (LMP), New Jersey Veterans Healthcare System, East Orange; University of Medicine and Dentistry of New Jersey (LMP), Newark; and Department of Health Policy and Management (DRB, DRM), Boston University School of Public Health, Boston, MA.

Funding Source: This study was supported by a grant from the US Department of Veterans Affairs Clinical Sciences Research and Development Service (principal investigator DRM and coprincipal investigator LMP). Dr Rose is supported by a career development grant from the US 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 views of the US Department of Veterans Affairs.

Author Disclosure: The authors (AJR, JAH, SMF, LMP, DRB, DRM) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (AJR, JAH, DRB, DRM); acquisition of data (DRB, DRM); analysis and interpretation of data (AJR, JAH, SMF, LMP, DRB, DRM); drafting of the manuscript (AJR, JAH, SMF, LMP, DRM); critical revision of the manuscript for important intellectual content (AJR, JAH, SMF, LMP, DRB); statistical analysis (AJR, DRM); administrative, technical, or logistic support (DRB); and supervision (JAH, DRB, DRM).

Address correspondence to: Adam J. Rose, MD, MSc, Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, 200 Springs Rd, Bldg 70, Bedford, MA 01730. E-mail: adamrose@bu.edu.

REFERENCES

1. Werner RM, Greenfield S, Fung C, Turner BJ. Measuring quality of care in patients with multiple clinical conditions: summary of a conference conducted by the Society of General Internal Medicine. *J Gen Intern Med.* 2007;22(8):1206-1211.

2. 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(9):1584-1590.

3. Dixon LB, Kreyenbuhl JA, Dickerson FB, et al. A comparison of type 2 diabetes outcomes among persons with and without severe mental illnesses [correction appears in *Psychiatr Serv.* 2004;55(9):1065]. *Psychiatr Serv.* 2004;55(8):892-900.

4. Frayne SM, Halanych JH, Miller DR, et al. Disparities in diabetes care: impact of mental illness. *Arch Intern Med.* 2005;165(22):2631-2638.

5. Meduru P, Helmer D, Rajan M, Tseng CL, Pogach L, Sambamoorthi U. Chronic illness with complexity: implications for performance

measurement of optimal glycemic control. *J Gen Intern Med.* 2007;22(suppl 3):408-418.

6. Redelmeier DA, Tan SH, Booth GL. The treatment of unrelated disorders in patients with chronic medical diseases. *N Engl J Med.* 1998;338(21):1516-1520.

7. Higashi T, Wenger NS, Adams JL, et al. Relationship between number of medical conditions and quality of care. *N Engl J Med.* 2007;356(24):2496-2504.

8. Lagu T, Weiner MG, Hollenbeak CS, et al. The impact of concordant and discordant conditions on the quality of care for hyperlipidemia. *J Gen Intern Med.* 2008;23(8):1208-1213.

9. Turner BJ, Hollenbeak CS, Weiner M, Ten Have T, Tang SS. Effect of unrelated comorbid conditions on hypertension management. *Ann Intern Med.* 2008;148(8):578-586.

10. Piette JD, Kerr EA. The impact of comorbid chronic conditions on diabetes care. *Diabetes Care.* 2006;29(3):725-731.

11. Bertakis KD, Azari R, Callahan EJ. Patient pain: its influence on primary care physician-patient interaction. *Fam Med.* 2003;35(2):119-123.

12. The Use of Opiates for the Treatment of Chronic Pain: A Consensus Statement From the American Academy of Pain Medicine and the American Pain Society. Glenview, IL: American Academy of Pain Medicine and American Pain Society; 1996.

13. 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:e46.

14. 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 Med.* 2002;17(5):327-333.

15. Parran T Jr. Prescription drug abuse: a question of balance. *Med Clin North Am.* 1997;81(4):967-978.

16. Portenoy RK, Dole V, Joseph H, et al. Pain management and chemical dependency: evolving perspectives. *JAMA.* 1997;278(7):592-593.

17. 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.* 2002;17(3):173-179.

18. Reuler JB, Girard DE, Nardone DA. The chronic pain syndrome: misconceptions and management. *Ann Intern Med.* 1980;93(4): 588-596.

19. Savage SR. Long-term opioid therapy: assessment of consequences and risks. *J Pain Symptom Manage.* 1996;11(5):274-286.

20. American Diabetes Association. Economic costs of diabetes in the U.S. in 2007 [published correction appears in *Diabetes Care.* 2008;31(6):1271]. *Diabetes Care.* 2008;31(3):596-615.

21. 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.

22. Fisher EB, Brownson CA, O'Toole ML, Anwuri VV, Shetty G. Perspectives on self-management from the Diabetes Initiative of the Robert Wood Johnson Foundation. *Diabetes Educ.* 2007;33(suppl 6):216S-224S.

23. Fisher EB, Brownson CA, O'Toole ML, et al. The Robert Wood Johnson Foundation Diabetes Initiative: demonstration projects emphasizing self-management. *Diabetes Educ.* 2007;33(1):83-84, 86-88, 91-92.

24. Dalewitz J, Khan N, Hershey CO. Barriers to control of blood glucose in diabetes mellitus. *Am J Med Qual.* 2000;15(1):16-25.

25. Drass J, Kell S, Osborn M, et al. Diabetes care for Medicare beneficiaries: attitudes and behaviors of primary care physicians. *Diabetes Care.* 1998;21(8):1282-1287.

26. Glasgow RE, Hampson SE, Strycker LA, Ruggiero L. Personal-model beliefs and social-environmental barriers related to diabetes self-management. *Diabetes Care.* 1997;20(4):556-561.

27. Golin CE, DiMatteo MR, Gelberg L. The role of patient participation in the doctor visit: implications for adherence to diabetes care. *Diabetes Care.* 1996;19(10):1153-1164.

28. Pogach LM, Brietzke SA, Cowan CL Jr, Conlin P, Walder DJ, Sawin CT. Development of evidence-based clinical practice guidelines for diabetes: the Department of Veterans Affairs/Department of Defense guidelines initiative. *Diabetes Care.* 2004;27(suppl 2):B82-B89.

29. US Department of Veterans Affairs. Clinical practice guidelines: diabetes. Washington, DC; 2009.

30. Fleming BB, Greenfield S, Engelgau MM, Pogach LM, Clauser SB, Parrott MA. The Diabetes Quality Improvement Project: moving science into health policy to gain an edge on the diabetes epidemic. *Diabetes Care.* 2001;24(10):1815-1820.

31. Krein SL, Heisler M, Piette JD, Makki F, Kerr EA. The effect of chronic pain on diabetes patients' self-management. *Diabetes Care.* 2005;28(1):65-70.

32. Mitchinson AR, Kerr EA, Krein SL. Management of chronic noncancer pain by VA primary care providers: when is pain control a priority? *Am J Manag Care.* 2008;14(2):77-84.

33. 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):377-391.

34. US Department of Justice Web site. Drug Enforcement Administration: drug scheduling. http://www.usdoj.gov/dea/pubs/scheduling.html. Accessed October 7, 2008.

35. Hermos JA, Young MM, Gagnon DR, Fiore LD. Characterizations of long-term oxycodone/acetaminophen prescriptions in veteran patients. *Arch Intern Med.* 2004;164(21):2361-2366.

36. US Department of Veterans Affairs. Clinical practice guidelines: opioid therapy for chronic pain. Washington, DC; 2009.

37. Butler SF, Budman SH, Fernandez KC, et al. Development and validation of the Current Opioid Misuse Measure. *Pain.* 2007;130(1-2): 144-156.

38. Clark JD. Chronic pain prevalence and analgesic prescribing in a general medical population. *J Pain Symptom Manage.* 2002;23(2): 131-137.

39. American Psychiatric Association Task Force on DSM-IV. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR.* Washington, DC: American Psychiatric Association; 2000.

40. Hayward RA, Hofer TP, Vijan S. Narrative review: lack of evidence for recommended low-density lipoprotein treatment targets: a solvable problem. *Ann Intern Med.* 2006;145(7):520-530.

41. Pogach LM, Tiwari A, Maney M, Rajan M, Miller DR, Aron D. Should mitigating comorbidities be considered in assessing healthcare plan performance in achieving optimal glycemic control? *Am J Manag Care.* 2007;13(3):133-140.

42. Action to Control Cardiovascular Risk in Diabetes Study Group. Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med.* 2008;358(24):2545-2559. ■

Journal of Thrombosis and Haemostasis, **7**: 94–101 DOI: 10.1111/j.1538-7836.2008.03199.x

ORIGINAL ARTICLE

Warfarin dose management affects INR control

A. J. ROSE,*† A. OZONOFF,‡ D. R. BERLOWITZ,*†§ L. E. HENAULT† and E. M. HYLEK†

**Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, MA; †Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA; ‡Department of Biostatistics; and §Department of Health Policy and Management, Boston University School of Public Health, Boston, MA, USA*

To cite this article: Rose AJ, Ozonoff A, Berlowitz DR, Henault LE, Hylek EM. Warfarin dose management affects INR control. *J Thromb Haemost* 2009; **7**: 94–101.

Summary. *Background:* Little is known about how patterns of warfarin dose management contribute to percentage time in the therapeutic International Normalized Ratio (INR) range (TTR). *Objectives:* To quantify the contribution of warfarin dose management to TTR and to define an optimal dose management strategy. *Patients/methods:* We enrolled 3961 patients receiving warfarin from 94 community-based clinics. We derived and validated a model for the probability of a warfarin dose change under various conditions. For each patient, we computed an observed minus expected (O − E) score, comparing the number of dose changes predicted by our model to the number of changes observed. We examined the ability of O − E scores to predict TTR, and simulated various dose management strategies in the context of our model. *Results:* Patients were observed for a mean of 15.2 months. Patients who deviated the least from the predicted number of dose changes achieved the best INR control (mean TTR 70.1% unadjusted); patients with greater deviations had lower TTR (65.8% and 62.0% for fewer and more dose changes respectively, Bonferroni-adjusted $P < 0.05/3$ for both comparisons). On average, clinicians in our study changed the dose when the INR was 1.8 or lower/3.2 or higher (mean TTR: 68%); optimal management would have been to change the dose when the INR was 1.7 or lower/3.3 or higher (predicted TTR: 74%). *Conclusions:* Our observational study suggests that INR control could be improved considerably by changing the warfarin dose only when the INR is 1.7 or lower/3.3 or higher. This should be confirmed in a randomized trial.

Keywords: anticoagulants, medication therapy management, quality of healthcare, warfarin.

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.: +1 781 687 3198; fax: +1 917 591 3104.
E-mail: adamrose@bu.edu

Received 12 August 2008, accepted 13 October 2008

Warfarin is highly efficacious for the prevention of strokes [1–3], the treatment of venous thromboembolism [4], and other indications. Numerous studies have focused on the underutilization of warfarin for patients with atrial fibrillation, the most common indication for warfarin [5–8]. However, relatively little is known about how best to manage warfarin once it is initiated. This lack of evidence regarding optimal management strategies probably contributes to our limited success in maintaining patients within the target International Normalized Ratio (INR) range. Even the highly selected and carefully managed patients in clinical trials only achieve a mean percentage time in the therapeutic INR range (TTR) of 66.4% [9], a figure that leaves much room for improvement. Suboptimal INR control has important clinical implications: in a recent study of patients in Ontario, time spent above the target INR range accounted for 26% of hemorrhagic events in patients receiving warfarin, and time below the target range accounted for 11% of thromboemboli [10]. For patients to derive maximal benefit from warfarin therapy, we must find ways to improve INR control.

Many factors might contribute to suboptimal INR control, including inadequate adherence to warfarin therapy [11], inadequate or erratic dietary intake of vitamin K [12], interactions with other medications [13], and genetic differences between patients [14]. Clinician variation in the decision to change the warfarin dose may likewise affect INR control, but this has not yet been studied. Clinicians are likely to differ in their probability of changing the warfarin dose in response to a given INR value, especially when the value is only slightly out of range. Studies have shown that managing warfarin doses via standardized computer-based algorithms results in higher TTR than management without computer assistance [15–17]. Even in settings where such computer programs are not utilized, these studies suggest that reducing variation in the management of warfarin doses can improve the stability of INR control.

We therefore used a large, nationally representative database of patients receiving warfarin without the aid of a computerized dosing algorithm to examine variations in the management of warfarin doses. In order to do this, we addressed the following four issues. First, we derived and

validated a model to predict when clinicians would change warfarin doses. This model describes what the 'average' clinician would do under various circumstances. Second, we documented deviations from this model, to demonstrate that the management of warfarin varies among patients and among sites of care, even in similar clinical situations. Third, we examined variations in practice as a predictor of TTR, on both a patient and a site level. Finally, we simulated the effect of different warfarin dose management strategies on TTR. Through this study, we sought to provide insights into optimal dose management strategies to improve the stability of INR control.

Methods

Collection of data/study cohort

Data collection for this study has been described in detail elsewhere [18–21]. Physician practices that were registered users of CoumaCare® software (Bristol-Myers Squibb, New York, NY, USA) were invited to participate. CoumaCare® was freely available and was used by many anticoagulation management services for patient tracking, data entry, and record keeping. It does not include dosing algorithms or other forms of decision support. This uniformity of data structure among sites made our study possible at a time when only 18% of US medical practices had an electronic medical record [22]. In total, 174 practices registered online to participate, and 101 sites had the technological capability and the review board approval necessary to proceed. All sites had at least one dedicated provider managing warfarin (most often a nurse), usually within the setting of a community-based, physician group practice. Enrollment began in April 2000 and follow-up ended in March 2002. To be eligible, patients had to be 18 years of age or older and provide written informed consent. The study protocol was approved by the Western Institutional Review Board (WIRB) of Olympia, WA, and by local review boards where they existed.

Database

To create the database for this study, we excluded patients who had a target INR range other than 2.0–3.0 for any part of the study period. We wished to study variations in the management of warfarin doses under similar clinical circumstances; therefore, we only included patients with the most common INR target range [13]. To limit our study to experienced users rather than patients in the inception phase of warfarin therapy, we also excluded all patients who had received less than a full month of warfarin at the time when they were enrolled in our study. Among the patients remaining in the database, we excluded all INR values recorded within 14 days before and after an intentional interruption of warfarin therapy for a procedure. The management of warfarin proximal to an intentional interruption is a different issue from routine warfarin management.

Variables

Patient age and gender were recorded. Patient race was recorded by site clinicians; and was collapsed to three categories: White, Black, and all others. The following comorbid conditions were recorded: diabetes, hypertension, congestive heart failure, history of prior stroke, and coronary artery disease. All INR values were recorded, with the dates on which they were obtained. From the INR values, we calculated TTR [23] for each patient. The total weekly warfarin dose was updated for each patient at each clinical visit; therefore, we were able to determine when the dose of warfarin was increased or decreased between two visits. It is of note that our analysis considered changes in the weekly dose of warfarin, but did not identify occasions when the patient was asked to omit doses of warfarin. In addition, the relative sizes of dose changes were not considered.

Statistical analyses – overview

Analyses consisted of four steps. First, we derived and validated two models to estimate, for each clinical encounter, the probability that the warfarin dose would be increased or decreased respectively. Second, we used these two models to predict the total number of expected dose changes for each patient throughout the study period. By comparing the expected dose changes to observed dose changes over the course of our study, we were able to characterize each patient as having more or fewer dose changes than predicted by our model, and to characterize the extent of the deviation. Third, we used these observed minus expected (O − E) scores to characterize the warfarin dose management received by individual patients and the dose management styles at the sites of care in our database. These scores quantify the deviation from usual management in a given clinical situation, as defined by our models. We examined the relationship of these O − E scores to the stability of INR control, as measured by TTR. Fourth, we simulated the effect on TTR of several possible INR cutoffs for adjusting the warfarin dose, using the models that we had developed. Each cutoff was associated with an expected O − E score and an expected mean TTR for the entire cohort.

Statistical analyses – details

We examined the following predictors of warfarin dose changes: INR at the current visit, INR at the previous visit, dose change at the previous visit, number of days since the previous visit (0–7, 8–14 or 15+ days), and patient characteristics (age, gender, race, hypertension, diabetes, coronary artery disease, prior stroke, and congestive heart failure). Because of the need for at least two previous visits to calculate dose changes at the previous visit, the first two INR values for each patient were excluded from this and later predictive models. We used separate logistic regressions to model dose increase and dose decrease, respectively. The dataset was split 60/40 for

derivation and validation of the models, respectively. Model performance was assessed by the *c*-statistic for discrimination and the Hosmer–Lemeshow test for calibration. Multiple selection procedures were tried to achieve models with the best possible fit, including using all of the variables together, best subset selection, backward and forward selection, and sequential trials of adding variables that had strong effects in bivariate analysis.

The total probability of a dose change at each visit was the sum of the predicted probability of a dose increase and of a dose decrease, with a maximum combined probability of 100%. We subtracted the total number of expected dose changes for each patient from the observed number of dose changes; we divided by the number of visits for that patient. Each patient received an O − E score, with a score of 0 indicating that the observed and expected dose changes were the same, negative scores indicating fewer dose changes than expected, and positive scores indicating more dose changes than expected. Each unit of 0.1 indicates either one more or one fewer dose change than expected per 10 visits. The O − E score largely examines decision-making in marginal situations, when clinicians might reasonably make either choice (i.e. to change the dose or refrain from doing so). An example of a marginal situation might be an INR of 3.1, which is only slightly out of range. An INR of 5.0, by comparison, would prompt most clinicians to reduce the dose.

We described the distribution of O − E scores among patients and of mean O − E scores among sites of care. We compared O − E scores to TTR on the patient level and the site level. For patient-level analyses, we used generalized estimating equations, to account for the intraclass correlation of patient outcomes among sites of care. We have previously shown, using this database, that gender, race, coronary artery disease and congestive heart failure predict TTR on a patient level [21]; we therefore controlled for these variables in our patient-level analyses. For site-level analyses, we used correlation, linear regression, and analysis of variance. We also compared actual management during the study to the management that would have occurred using various INR cutoffs for adjusting the warfarin dose. We computed the O − E score that would have resulted from each of these cutoffs, as well as the predicted TTR for each cutoff. We derived and validated our models using SAS, Version 9.1 (SAS Institute, Cary, NC, USA) and performed other analyses using the R statistical package, version 2.2 (R Foundation, 2007).

Results

There were 6761 patients in the entire database; after exclusions for INR target range and patients new to warfarin, 4248 patients remained, with a total of 67 199 INR values. We excluded 1206 INR values, due to proximity to an interruption of therapy, and 8408 INR values because they were among the first two values for a patient, leaving a total of 58 791 INR values for analysis. An additional 650 INRs that were more than 56 days from the previous INR measurement were excluded because of difficulty in computing TTR, leaving 58 141 INR values. Finally, after these maneuvers, 287 patients had fewer than three INR values and were excluded from analysis; 3961 patients remained. Characteristics of the 3961 patients in our database appear in Table 1. Their mean age was 71.9 years, and 42% were female. Atrial fibrillation was the most common indication for anticoagulation (62%); valvular heart disease was relatively uncommon, due to the requirement for an INR target range of 2–3. The mean TTR was 68.4%.

Table 1 Demographics, clinical characteristics, warfarin management and International Normalized Ratio (INR) control for patients in the study (n = 3961)

Parameter	Number (%) or mean (SD)
Mean age (SD)	71.9 (11.1)
Female gender (%)	1678 (42%)
Race/ethnicity (%)	
White	3673 (93%)
Black	101 (2%)
Other	187 (5%)
Primary indication for anticoagulation (%)	
Atrial fibrillation	2467 (62%)
Venous thromboembolism	527 (13%)
Valvular heart disease/prosthetic valve	120 (3%)
Previous stroke/embolus	448 (11%)
All other indications combined	399 (10%)
Comorbid conditions (%)	
Coronary artery disease	969 (24%)
Diabetes mellitus	629 (16%)
Hypertension	1699 (43%)
Heart failure	692 (18%)
Prior stroke	399 (10%)
Mean number of INR values* (SD)	14.6 (7.8)
Mean interval between INR tests*, days (SD)	22.1 (6.2)
Mean total months in database* (SD)	15.2 (5.6)
Mean weekly warfarin dose*, mg (SD)	30.6 (13.4)
Mean number of interruptions of warfarin* (SD)	0.068 (0.25)
Percentage time in therapeutic INR range (TTR)*	
Mean TTR (SD)	68.4% (19.3%)
Median TTR (IQR)	70.2% (26.8%)

IQR, interquartile range. *For these calculations, we excluded the first two INR measurements for each patient, because of the need for two prior visits in our prediction models.

The dose of warfarin was decreased at 14.4% of visits, increased at 16.4% of visits, and remained the same at 69.3% of visits. Both dose decreases and dose increases were best predicted by three-variable models (Appendix A); different selection procedures gave the same result. Dose decrease was best predicted by a model containing INR at the current visit, number of days since the previous visit, and dose increase at the previous visit (*c*-statistic = 0.90). Dose increase was best predicted by a model containing INR at the current visit, number of days since the previous visit, and dose decrease at the previous visit (*c*-statistic = 0.88). In both models, the INR value was by far the most important predictor. For both models, a shorter follow-up interval predicted a higher

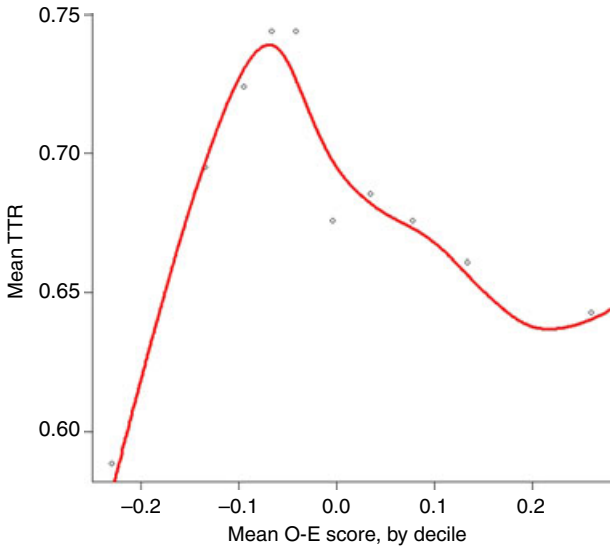


Fig. 1. Relationship between percentage time in the therapeutic International Normalized Ratio (INR) range (TTR) and the observed minus expected score for warfarin dose management style (O − E score). The curve is a cubic smoothing spline fit to the patient data [25]; points denote mean values grouped by deciles of patient data (approximately 390 patients per decile).

likelihood of a dose change, and a dose change in one direction at the previous visit was predictive of a dose change in the opposite direction at the current visit. Other variables were not meaningful predictors of dose changes, including age, gender, race, hypertension, diabetes, congestive heart failure, and prior stroke.

The mean O − E score for the 3961 patients was − 0.006 (SD 0.13); O − E scores were approximately normally distributed. Fifty-seven per cent of patients had an O − E score between − 0.1 and 0.1 (i.e. within one dose change of predicted per 10 visits), and 88% had an O − E score between − 0.2 and 0.2. Using a cubic smoothing spline function [24], we fitted a curve to describe the relationship between O − E scores and TTR on the patient level (Fig. 1). In general, extremes of O − E score were associated with lower TTR than O − E scores closer to zero. The maximum of the fitted curve occurs at an O − E score of − 0.083, or approximately one fewer dose change than predicted per 12 visits. To further evaluate the relationship between the O − E score and TTR, we divided patients into three groups based on their O − E scores: group A (considerably fewer dose changes than predicted), group C (considerably more dose changes than predicted), and group B (all others). Group B, the group whose management conformed most closely to the norm, had the highest TTR in both unadjusted and adjusted analyses, a statistically significant result (Table 2).

We also simulated the possible impact of four INR cutoffs for change of the warfarin dose. The first strategy would be to adjust the warfarin dose whenever the INR is 1.9 or lower/3.1 or higher. This strategy, applied to our database, would have yielded a mean O − E score of 0.071 and a mean predicted TTR of 67% (see Fig. 1). The second strategy would be to adjust the dose whenever the INR is 1.8 or lower/3.2 or higher; this would have yielded a mean O − E score of 0.003 and a mean TTR of 69%. The third strategy would be to adjust the dose whenever the INR is 1.7 or lower/3.3 or higher; this would have yielded a mean O − E score of − 0.064 and a mean predicted TTR of 74%. The fourth strategy would be to adjust the dose whenever the INR is 1.6 or lower/3.4 or higher; this would have yielded a mean O − E score of − 0.115 and a mean predicted TTR of 71%. According to this simulation, the average management in this study was extremely similar to the second strategy (change dose when the INR is 1.8 or lower/3.2 or higher), whereas the management strategy expected to produce the highest TTR would be extremely similar to the third strategy (change dose when the INR is 1.7 or lower/3.3 or higher).

Table 2 Comparison of mean percentage time in therapeutic INR range (TTR) among patients with fewer dose changes than predicted (group A), about as many dose changes as predicted (group B), and more dose changes than predicted (group C) (total n = 3961 patients)

	Number of patients	Mean TTR (unadjusted)	Mean TTR (adjusted)
Group A (O − E < 0.1)	919	65.8%	67.9%
Group B (O − E − 0.1–0.1)	2242	70.1%	72.2%
Group C (O − E > 0.1)	800	62.0%	64.3%
P-value	–	< 0.001	< 0.001

O − E, observed minus expected score. All analyses account for the site of care using generalized estimating equations. Adjusted analyses also control for gender, race, coronary artery disease, and congestive heart failure. In adjusted analyses, the TTR presented is for a white male without coronary artery disease or congestive heart failure. Pairwise comparisons were significant between group B and the other groups at the Bonferroni-adjusted (0.05/3) level in both unadjusted and adjusted analyses.

We calculated mean O − E scores for the sites of care represented in the database. It was not possible to calculate a mean O − E score for seven of the 101 sites, because these sites had only one or zero patients who met our inclusion criteria. Among the other 94 sites, there was a wide range of practice (Appendix B): 17 sites had a mean O − E score below − 0.1, whereas eight sites had a mean O − E score above 0.1. Extremes of practice in this regard were not solely contributed by sites with few patients, suggesting that we observed true differences in practice rather than merely statistical variation. For example, the three sites with the highest average O − E scores had relatively large patient panels of 57, 40, and 124 patients each.

To explore the possible impact of this site-level variation in management, we divided sites into deciles based on their mean absolute value O − E scores and plotted the deciles against mean site TTR (Fig. 2). There was a strong negative correlation between mean absolute O − E score and mean TTR (r = − 0.77, P = 0.003), indicating that extremes of warfarin dose management were associated with worse INR control on the site level.

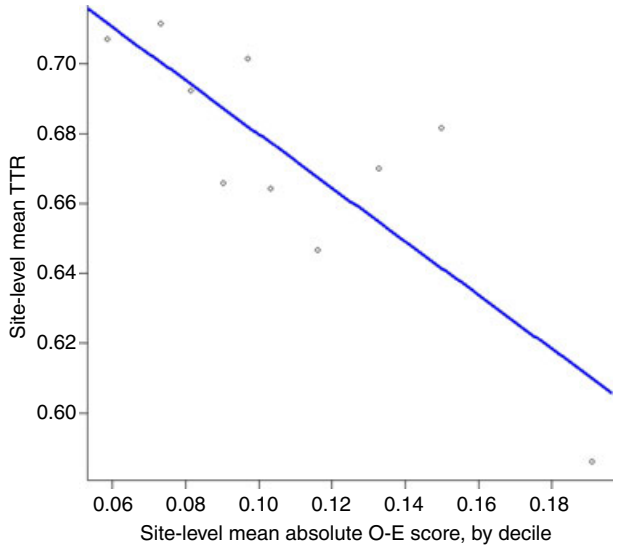


Fig. 2. Mean percentage time in the therapeutic International Normalized Ratio (INR) range (TTR), plotted against the mean absolute value of observed minus expected score (O − E score) by site ($n = 94$ sites). Each dot represents either nine or 10 sites and corresponds to a decile of the O − E score. For the line of correlation, $r = -0.77$, and $P = 0.003$.

Discussion

We used a large, nationally representative database of oral anticoagulation care to examine warfarin dose management styles. We modeled norms of behavior regarding dose changes in certain clinical situations. Our models were able to predict clinical behavior with great accuracy using relatively few variables: the INR value at the current visit, the number of days since the previous visit, and whether or not the dose was changed at the previous visit. We then used these models to describe variations in practice. There was considerable variation among patients and among sites of care regarding warfarin dose management styles in clinically similar situations. Both on the patient level and on the site level, a moderate tendency to change the warfarin dose was associated with higher TTR than either extreme of management. The apparent impact of O − E score upon TTR in our study was considerable. For example, TTR at the peak of the O − E curve (Fig. 1) was 74%. By contrast, TTR at an O − E score of 0.2 was 64% and TTR at an O − E score of − 0.2 was 62%. By comparison, Poller *et al.* found that mean TTR in a group randomized to computerized dose management was 63% vs. 53% among controls [16], a similar effect size.

Our simulation of four dosing strategies suggests that average management in this database was to change the dose when the INR was 1.8 or lower/3.2 or higher, whereas optimal management would be to change the dose when the INR was 1.7 or lower/3.3 or higher. This relatively minor change in management would be expected to increase TTR from 68% to 74%. It should not be surprising that this apparently small difference in management would make such a large difference in INR control, because 15% of the dose increases in our study occurred at an INR of 1.8, and 7% of the dose decreases occurred at an INR of 3.2. Our results suggest that such dose changes, which occurred when the INR was only slightly outside the target range, served merely to perturb the INR, setting up a cycle of adjustment and readjustment. It is likely that the patient would have done better if the INR were simply rechecked a week later, when many patients would have been found to be in-range once again, without any particular intervention. Our findings actually echo those of an audit of anticoagulation care by Rose, in which only 50% of patients were found to be within 0.5 of a target INR of 2.5 at any given time, whereas a full 80% of patients were within 0.75 of the target [25]. It appears to be incompatible with human physiology to expect patients to remain exclusively within an INR target range of 2.0–3.0; the target range that might be suggested by our results, namely 1.8–3.2, seems more attainable. Citing this article by Rose as evidence that target ranges are impractical, the British Society of Haematology Guidelines on Oral Anticoagulation [26] suggest target INR values for various indications (e.g. 2.5), rather than specifying target INR ranges (e.g. 2.0–3.0).

An important strength of this study is that we used the O − E method to describe warfarin dose management. By modeling norms of management, and then documenting deviation from those norms, we were able to ensure that dose management was being compared in similar situations. In particular, the variable regarding the follow-up interval ensured that our models captured not only the effect of the INR level, but also the clinician's *a priori* suspicion that the patient's warfarin dose would need to be adjusted at the visit. The interval between visits is likely to reflect many unmeasured factors (comorbid conditions, medication changes, bleeding risk, and history of INR stability) that would enter into clinician decision-making. These features of our predictive model strongly suggest that warfarin dose management was the cause of between-group differences in TTR rather than the reverse. In addition, the fact that TTR drops off on both sides of the optimal O − E value makes it difficult to imagine a plausible mechanism for reverse causation. Finally, controlling for important covariates did not alter between-group comparisons of TTR (Table 2), suggesting that differences in management rather than patient characteristics explained the effects that we observed.

However, our study does have limitations. Our results were derived from analysis of a single dataset, albeit a large and nationally representative dataset. Confirmation of these findings, using a different dataset, would provide an additional level of support. In addition, our study shares the limitations of any observational study, in that it cannot definitively establish cause and effect. Although our study may suggest an optimal dose management strategy in the absence of computerized decision support, definitive evidence would require prospective, randomized studies. We would urge that such a trial be undertaken as soon as possible, as our study suggests that improving the dosing of warfarin could meaningfully improve patient outcomes.

Second, we did not examine patients with target INR ranges other than 2–3, and nor did we examine warfarin dose management during the inception of therapy; our results cannot be generalized to these situations. Finally, although we did not consider the magnitude of dose changes, it should be noted that for the marginal situations that we modeled (e.g. INR of 1.8 or 3.2), a 5–10% adjustment of the weekly warfarin dose is likely to be the ideal choice. Larger dose changes would be reserved for more extreme deviations of INR from the target range, when the decision regarding whether to change the dose at all would be more obvious.

In summary, we found that warfarin dose management varied widely in similar clinical situations. This variation in practice had implications for INR control; extremes of management were associated with lower TTR than management closer to the mean. Our simulation suggests that, when the target range is 2.0–3.0, optimal management of warfarin would be to change the warfarin dose only when the INR is 1.7 or lower/3.3 or higher; a smaller tolerance for slightly out-of-range values seems to destabilize the INR through excessive dose adjustments. Finally, our study suggests that in addition to offering warfarin to as many optimal candidates as possible, we also need to optimize warfarin dose management to fully realize the benefits of anticoagulation.

Acknowledgements

The opinions expressed in this manuscript do not necessarily represent the views or policies of the Department of Veterans Affairs. A. J. Rose is supported by a career development award from the United States Department of Veterans Affairs, Health Services Research and Development Service.

Disclosure of Conflict of Interest

This study was funded by Bristol-Myers Squibb. Bristol-Myers Squibb had no role in the design and conduct of the study, the collection, management, analysis and interpretation of the data, or the preparation, review and approval of the manuscript.

Appendix A: Models to predict dose decrease and dose increase

Model to predict a dose decrease: predictor variables

1. INR at this visit
   a. 2.9 or below
   b. 3.0
   c. 3.1
   d. 3.2
   e. 3.3
   f. 3.4
   g. 3.5 or above
2. Time since last INR
   a. 0–7 days
   b. 8–14 days
   c. 15 days or more
3. Dose increase at the last visit? (yes/no)

DERIVATION SET
c-statistic 0.898
Hosmer–Lemeshow test $P < 0.001$

Group	Number in group	Observed dose decreases	Expected dose decreases
1	16 022	280	395
2	4403	134	144
3	3373	221	174
4	3713	530	463
5	3481	1027	1003
6	3843	2841	2854

VALIDATION SET
c-statistic = 0.896
Hosmer–Lemeshow test $P < 0.001$

Group	Number in group	Observed dose decreases	Expected dose decreases
1	10 737	186	263
2	2964	109	105
3	2351	156	125
4	2479	326	300
5	2338	742	703
6	2437	1800	1823

Model to predict a dose increase: predictor variables

1. INR at this visit
   a. 1.7 or below
   b. 1.8
   c. 1.9
   d. 2.0 or above
2. Days since last visit
   a. 0–7 days
   b. 7–14 days
   c. 15+ days
3. Dose decrease at last visit (yes/no)

DERIVATION SET
c-statistic = 0.883
Hosmer-Lemeshow test $P < 0.001$

Group	Number in group	Observed dose increases	Expected dose increases
1	16 349	337	445
2	4637	174	215
3	3825	405	345

Group	Number in group	Observed dose increases	Expected dose increases
4	3534	1031	918
5	3421	1666	1571
6	3079	2079	2198

VALIDATION SET
c-statistic = 0.879
Hosmer–Lemeshow test $P < 0.001$

Group	Number in group	Observed dose increases	Expected dose increases
1	10 941	254	332
2	3089	128	148
3	2576	279	247
4	2319	669	589
5	2322	1138	1063
6	2059	1350	1439

Appendix B: Means, 95% confidence intervals and *z*-scores for 94 sites of care regarding mean site O − E score. Sites of care are listed from lowest O − E score to highest.

Site	*n*	Mean O − E score	95% CI, lower bound	95% CI, upper bound	*z*-score
A*	8	− 0.18	− 0.247	− 0.113	− 2.093
B*	6	− 0.173	− 0.295	− 0.051	− 2.012
C	33	− 0.156	− 0.232	− 0.08	− 1.809
D	25	− 0.147	− 0.22	− 0.075	− 1.704
E	53	− 0.147	− 0.167	− 0.126	− 1.692
F	23	− 0.145	− 0.189	− 0.101	− 1.673
G	3	− 0.144	− 0.357	0.068	− 1.667
H	14	− 0.142	− 0.184	− 0.101	− 1.643
I	12	− 0.135	− 0.186	− 0.085	− 1.557
J	51	− 0.133	− 0.16	− 0.106	− 1.528
K	4	− 0.119	− 0.412	0.173	− 1.366
L	19	− 0.118	− 0.177	− 0.059	− 1.353
M	2	− 0.118	− 0.223	− 0.012	− 1.346
N	10	− 0.116	− 0.174	− 0.058	− 1.33
O	23	− 0.115	− 0.158	− 0.072	− 1.311
P	13	− 0.113	− 0.16	− 0.066	− 1.289
Q	24	− 0.111	− 0.154	− 0.069	− 1.267
R	40	− 0.098	− 0.117	− 0.078	− 1.103
S	147	− 0.093	− 0.109	− 0.078	− 1.053
T	19	− 0.087	− 0.139	− 0.036	− 0.981
U	5	− 0.087	− 0.25	0.077	− 0.972
V	68	− 0.079	− 0.105	− 0.054	− 0.883
W	268	− 0.079	− 0.091	− 0.067	− 0.876
X	8	− 0.078	− 0.119	− 0.036	− 0.865
Y	20	− 0.076	− 0.108	− 0.045	− 0.85
Z	153	− 0.075	− 0.091	− 0.059	− 0.833
AA	7	− 0.074	− 0.122	− 0.027	− 0.825
AB	115	− 0.072	− 0.088	− 0.056	− 0.793
AC	21	− 0.066	− 0.111	− 0.021	− 0.723
AD	23	− 0.065	− 0.103	− 0.026	− 0.709

Appendix B: (Continued)

Site	*n*	Mean O − E score	95% CI, lower bound	95% CI, upper bound	*z*-score
AE	98	− 0.064	− 0.083	− 0.046	− 0.702
AF	29	− 0.063	− 0.092	− 0.033	− 0.684
AG	93	− 0.061	− 0.088	− 0.035	− 0.667
AH	41	− 0.06	− 0.079	− 0.041	− 0.653
AI	25	− 0.06	− 0.123	0.004	− 0.651
AJ	24	− 0.058	− 0.101	− 0.015	− 0.628
AK	24	− 0.054	− 0.096	− 0.013	− 0.585
AL	44	− 0.045	− 0.101	0.01	− 0.476
AM	48	− 0.042	− 0.071	− 0.014	− 0.439
AN	48	− 0.04	− 0.072	− 0.008	− 0.414
AO	43	− 0.039	− 0.062	− 0.017	− 0.402
AP	36	− 0.035	− 0.062	− 0.008	− 0.353
AQ	8	− 0.034	− 0.071	0.004	− 0.335
AR	6	− 0.032	− 0.113	0.05	− 0.312
AS	32	− 0.027	− 0.07	0.015	− 0.258
AT	19	− 0.013	− 0.045	0.019	− 0.086
AU	44	− 0.012	− 0.065	0.041	− 0.074
AV	25	− 0.01	− 0.057	0.037	− 0.052
AW	21	− 0.007	− 0.052	0.038	− 0.015
AX	97	− 0.005	− 0.031	0.021	0.009
AY	34	− 0.003	− 0.036	0.029	0.029
AZ	97	− 0.003	− 0.022	0.016	0.031
BA	42	− 0.001	− 0.031	0.03	0.062
BB	78	0	− 0.026	0.026	0.072
BC	8	0	− 0.126	0.127	0.075
BD	36	0.002	− 0.032	0.035	0.09
BE	51	0.002	− 0.029	0.034	0.098
BF	44	0.003	− 0.061	0.066	0.105
BG	9	0.004	− 0.124	0.132	0.118
BH	25	0.005	− 0.027	0.037	0.128
BI	35	0.006	− 0.035	0.047	0.143
BJ	31	0.008	− 0.024	0.041	0.17
BK	39	0.011	− 0.018	0.039	0.2
BL	51	0.011	− 0.026	0.049	0.204
BM	40	0.014	− 0.018	0.045	0.233
BN	3	0.014	− 0.159	0.187	0.239
BO	103	0.014	− 0.007	0.035	0.24
BP	24	0.019	− 0.029	0.067	0.298
BQ	89	0.022	0.007	0.037	0.336
BR	23	0.03	− 0.025	0.085	0.431
BS	2	0.032	− 1.631	1.695	0.451
BU	51	0.034	− 0.007	0.074	0.474
BV	5	0.034	− 0.084	0.152	0.478
BW	48	0.038	0.004	0.072	0.521
BX	23	0.044	0.018	0.069	0.594
BY	68	0.049	0.021	0.077	0.658
BZ	14	0.049	0.017	0.081	0.66
CA	131	0.055	0.035	0.075	0.725
CB	121	0.055	0.029	0.082	0.732
CC	10	0.059	0.018	0.099	0.773
CD	5	0.062	− 0.277	0.4	0.808
CE	86	0.063	0.039	0.088	0.83
CF	13	0.071	0	0.142	0.92
CG	15	0.076	0.017	0.136	0.987
CH	33	0.084	0.057	0.111	1.081
CI	18	0.087	0.002	0.171	1.11
CJ	127	0.105	0.09	0.12	1.33
CK	6	0.114	− 0.057	0.285	1.439
CL	49	0.123	0.078	0.168	1.551

Appendix B: (Continued)

Site	*n*	Mean O − E score	95% CI, lower bound	95% CI, upper bound	*z*-score
CM	160	0.128	0.107	0.148	1.602
CN	4	0.129	− 0.328	0.587	1.625
CO*	57	0.165	0.137	0.194	2.057
CP*	40	0.167	0.141	0.193	2.073
CQ*	124	0.204	0.178	0.23	2.521

*Outlier site at the 0.05 level

References

1 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. *Circulation* 2006; **114**: e257–354.

2 Salem DN, O'Gara PT, Madias C, Pauker SG. Valvular and structural heart disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 2008;**133** (6 Suppl.):593S–629S.

3 Singer DE, Albers GW, Dalen JE, Fang MC, Go AS, Halperin JL, Lip GY, Manning WJ. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 2008;**133** (6 Suppl.):546S–92S.

4 Kearon C, Kahn SR, Agnelli G, Goldhaber S, Raskob GE, Comerota AJ. Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 2008;**133** (6 Suppl.):454S–545S.

5 Gurwitz JH, Monette J, Rochon PA, Eckler MA, Avorn J. Atrial fibrillation and stroke prevention with warfarin in the long-term care setting. *Arch Intern Med* 1997; **157**: 978–84.

6 McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003; **348**: 2635–45.

7 Smith NL, Psaty BM, Furberg CD, White R, Lima JA, Newman AB, Manolio TA. Temporal trends in the use of anticoagulants among older adults with atrial fibrillation. *Arch Intern Med* 1999; **159**: 1574–8.

8 Stafford RS, Singer DE. Recent national patterns of warfarin use in atrial fibrillation. *Circulation* 1998; **97**: 1231–3.

9 van Walraven C, Jennings A, Oake N, Fergusson D, Forster AJ. Effect of study setting on anticoagulation control: a systematic review and metaregression. *Chest* 2006; **129**: 1155–66.

10 van Walraven C, Oake N, Wells PS, Forster AJ. Burden of potentially avoidable anticoagulant-associated hemorrhagic and thromboembolic events in the elderly. *Chest* 2007; **131**: 1508–15.

11 Kimmel SE, Chen Z, Price M, Parker CS, Metlay JP, Christie JD, Brensinger CM, Newcomb CW, Samaha FF, Gross R. The influence of patient adherence on anticoagulation control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Arch Intern Med* 2007; **167**: 229–35.

12 Sconce E, Avery P, Wynne H, Kamali F. Vitamin K supplementation can improve stability of anticoagulation for patients with unexplained variability in response to warfarin. *Blood* 2007; **109**: 2419–23.

13 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; **6** (Suppl.): 160S–98S.

14 Schwarz UI, Ritchie MD, Bradford Y, Li C, Dudek SM, Frye-Anderson A, Kim RB, Roden DM, Stein CM. Genetic determinants of response to warfarin during initial anticoagulation. *N Engl J Med* 2008; **358**: 999–1008.

15 Manotti C, Moia M, Palareti G, Pengo V, Ria L, Dettori AG. Effect of computer-aided management on the quality of treatment in anticoagulated patients: a prospective, randomized, multicenter trial of APROAT (Automated PRogram for Oral Anticoagulant Treatment). *Haematologica* 2001; **86**: 1060–70.

16 Poller L, Shiach CR, MacCallum PK, Johansen AM, Munster AM, Magalhaes A, Jespersen J. Multicentre randomised study of computerised anticoagulant dosage. European Concerted Action on Anticoagulation. *Lancet* 1998; **352**: 1505–9.

17 Ryan PJ, Gilbert M, Rose PE. Computer control of anticoagulant dose for therapeutic management. *BMJ* 1989; **299**: 1207–9.

18 Garcia D, Regan S, Crowther M, Hughes RA, Hylek EM. Warfarin maintenance dosing patterns in clinical practice: implications for safer anticoagulation in the elderly population. *Chest* 2005; **127**: 2049–56.

19 Garcia DA, Regan S, Crowther M, Hylek EM. The risk of hemorrhage among patients with warfarin-associated coagulopathy. *J Am Coll Cardiol* 2006; **47**: 804–8.

20 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.

21 Rose AJ, Ozonoff A, Henault LE, Hylek EM. Warfarin for atrial fibrillation in community-based practice. *J Thromb Haemost* 2008; **6**: 1647–54.

22 Cherry DK, Burt CW, Woodwell DA. National Ambulatory Medical Care Survey: 2001 summary. *Adv Data* 2003; (**337**): 1–44.

23 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.

24 Hastie TJ, Tibshirani RJ. Generalized Additive Models. New York: Chapman and Hall, 1990.

25 Rose PE. Audit of anticoagulant therapy. *J Clin Pathol* 1996; **49**: 5–9.

26 British Society of Haematology. Guidelines on oral anticoagulation. *Br J Haematol* 1998; **101**: 374–87.

Epidemiology of Subtherapeutic Anticoagulation in the United States

Adam J. Rose, MD, MSc; Al Ozonoff, PhD; Richard W. Grant, MD, MPH; Lori E. Henault, MPH; Elaine M. Hylek, MD, MPH

Background—Low international normalized ratio (INR; ≤1.5) increases risk for thromboembolism. However, little is known about the epidemiology of low INR.

Methods and Results—We prospectively collected data from 47 community-based clinics located throughout the United States from 2000 to 2002. We examined risk factors for low INR (≤1.5), reasons given in the medical record for low INR, and proportion of thromboembolic events that occurred during periods of low INR. Of the 4489 patients in our database, 1540 (34%) had at least 1 low INR. Compared with men, women had an increased incidence of low INR (adjusted incidence rate ratio, 1.44; $P<0.001$). Compared with patients anticoagulated for atrial fibrillation, patients anticoagulated for venous thromboembolism had an increased incidence of low INR (adjusted incidence rate ratio, 1.48; $P<0.001$). The 5 most common reasons for low INR were nonadherence (17%), interruptions for procedures (16%), recent dose reductions (15%), no reason apparent after questioning (15%), and second or greater consecutive low INR (13%). A total of 21.8% of thromboembolic events (95% CI, 12.2 to 35.4%) occurred during periods of low INR; 58% of these events were related to an interruption of warfarin therapy.

Conclusions—In this cohort of patients receiving warfarin, more than 1 in 5 thromboembolic events occurred during a period of low INR. Women and patients anticoagulated for venous thromboembolism were particularly likely to experience low INR. Improving adherence, minimizing interruptions of therapy, and addressing low INR more promptly could reduce the risk of low INR. **(*Circ Cardiovasc Qual Outcomes*. 2009;2:591-597.)**

Key Words: warfarin ■ thromboembolism ■ anticoagulants ■ quality of health care ■ medication therapy management

Warfarin is a highly effective therapy to prevent thromboembolic complications of venous thromboembolism (VTE),[1] atrial fibrillation (AF),[2,3] and valvular heart disease.[4] However, warfarin is an extremely challenging therapy in clinical practice[5]; a recent meta-analysis found that patients spend an average of only 66% of time in the therapeutic range when managed in specialized anticoagulation clinics, and only 57% of time when managed in usual care.[6] The large amount of time spent outside the target range has important clinical consequences: patients with better control have fewer hemorrhagic and thromboembolic events.[7]

Many studies have focused on the risk factors for and effects of high international normalized ratio (INR), and therefore, we know a considerable amount about these topics.[8–21] For example, Hylek et al[12] showed that risk factors for INR >6.0 included high-dose acetaminophen, new medications known to potentiate warfarin, advanced malignancy, diarrhea, decreased oral intake, and taking more warfarin than prescribed. However, there has been considerably less research regarding low INR. Some studies have demonstrated that low INR is associated with attenuation of the protective effects of anticoagulation therapy, as would be expected.[7,22–24] Although we know that low INR increases the risk of thromboembolism, relatively little is known about the epidemiology of low INR.

We therefore used data from a large nationally representative anticoagulation cohort to describe the epidemiology of low INR. We examined patient-level risk factors for INR ≤1.5, the threshold below which the risk of thromboembolism rises most acutely.[22,23] We report clinician-documented explanations for low INR, recorded at the time the INR result was obtained. We examine predictors of time until the next INR and the next in-range INR. Finally, we estimate the proportion of thromboembolic events attributable to low INR. Low INR has been an understudied topic, and our study can serve as the beginning of an effort to understand, address, and reduce this phenomenon.

Received March 5, 2009; accepted August 3, 2009.

From the Center for Health Quality, Outcomes, and Economic Research (A.J.R., A.O.), Bedford VA Medical Center, Bedford, Mass; the Department of Medicine, Section of General Internal Medicine (A.J.R., L.E.H., E.M.H.), Boston University School of Medicine, Boston, Mass; the Department of Biostatistics (A.O.), Boston University School of Public Health, Boston, Mass; and the Clinical Epidemiology Unit, General Medicine Division, Department of Medicine (R.W.G.), Massachusetts General Hospital and Harvard Medical School, Boston, Mass.

Correspondence to Adam J. Rose, MD, MSc, Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, 200 Springs Rd, Building 70, Bedford, MA 01730. E-mail adamrose@bu.edu

Circ Cardiovasc Qual Outcomes **is available at http://circoutcomes.ahajournals.org** **DOI: 10.1161/CIRCOUTCOMES.109.862763**

WHAT IS KNOWN

- Low international normalized ratio (INR; ≤1.5) is a risk factor for thromboembolism in patients receiving warfarin. Reducing its occurrence in clinical practice could improve patient outcomes.
- However, little is known about the epidemiology of low INR, especially which patients are more likely to experience it and what factors may contribute to it.

WHAT THE STUDY ADDS

- In our study, the 5 most common reasons for low INR were nonadherence (17%), interruptions for procedures (16%), recent dose reductions (15%), no reason apparent after questioning (15%), and second or greater consecutive low INR (13%).
- Low INR was more common among women and patients anticoagulated for venous thromboembolism, even after controlling for covariates. These novel findings require confirmation and further investigation.
- Our study identifies groups of patients who may be at higher risk for low INR and factors that seem to cause it. This can serve as the beginning of a concerted effort to reduce the incidence of low INR in clinical practice.

Methods

Study Enrollment

Data collection for the Anticoagulation Consortium to Improve Outcomes Nationally (ACTION) study has been described elsewhere.[18,25–29] Physician practices that were registered users of CoumaCare software (Bristol-Myers Squibb) were invited to participate. CoumaCare was a freely available software package which provided a rudimentary electronic medical record for anticoagulation management. The software assisted with record keeping and patient tracking, but did not provide advanced functions such as dosing guidance. The uniformity of data structure provided by the software package allowed us to collect data from diverse community-based sites at a time when few such sites had any sort of electronic medical record.

In total, 174 practices registered online to participate, and 101 sites had the technological capability and the review board approval necessary to proceed. One of the functions of CoumaCare is to allow providers to input a note regarding each INR value and how it was addressed. Data for the current study are drawn from the 47 study sites which had text notes for at least 90% of INR values. Excluded study sites had some text notes, but often only when the INR was out of range, precluding a full investigation of the reasons for low INR values.

All sites had at least 1 dedicated provider managing warfarin, usually within the setting of a community-based physician group practice. Enrollment began in April 2000 and follow-up ended in March 2002. Missing data fields and data entry errors were resolved directly with the sites by the data coordinating center on a weekly basis before the data were transmitted to study investigators. The study protocol was approved by the Western Institutional Review Board of Olympia, Wash, and by local review boards where they existed.

Variables

We identified all INR values ≤1.5 and reviewed anticoagulation clinic notes to determine the stated reason for the low value. Investigators developed 10 categories to encompass the reasons for low INR, and then assigned a reason to each value using chart review. One of these 10 categories, "initiation phase," was defined as all consecutive low INR values before the first in-range or high INR was recorded. Another category, "continuing low," was used for all successive low INRs, regardless of the reason for the first in the series, until the next in-range or high INR value.

We characterized all patients regarding age, gender, race, indication for anticoagulation, and comorbid conditions. In addition, we recorded instances when warfarin was intentionally interrupted (a "hold") by reviewing all 84 915 anticoagulation notes. The great majority of these holds were for minor procedures such as biopsies or colonoscopies, whereas relatively few were for major surgery. For the purpose of this analysis, however, all holds were considered the same. We confirmed all thromboembolic events (ischemic stroke, systemic embolus, deep venous thrombosis, and pulmonary embolus) through chart review.

Statistical Analyses

We tabulated demographics (age, gender, and race) and comorbid conditions (hypertension, diabetes, prior stroke, coronary artery disease, congestive heart failure) and performed bivariate comparisons between patients with at least 1 low INR versus those without. Because of correlated data within sites of care, we computed probability values for bivariate comparisons using 10 000 Monte Carlo permutations (stratified on clinical site) of the indicator for at least 1 low INR value. We modeled the rate of low INR per person-year using patient-level risk factors as independent predictors. Covariates included demographics, indication for therapy, comorbid conditions, and warfarin holds. For indication for therapy, the reference category was atrial fibrillation without prior stroke, chosen because it was the most numerous. Patients with valvular heart disease comprised another category of indication for therapy; most such patients had mechanical replacement valves (90%), but some had other conditions (such as mitral stenosis). We further divided patients with valvular heart disease into those with target ranges of 2 to 3 versus 2.5 to 3.5, with the hypothesis that those with a high target range would be protected from low INR. We used a Poisson regression model, fit with generalized estimating equations (SAS PROC GENMOD, SAS version 9.1, SAS Institute, Inc), to account for intraclass correlation by site of care. For this analysis, we excluded low INR values attributable to the inception phase of warfarin therapy, when low INR is to be expected.

Among patients who had at least 1 low INR, we computed the relative frequencies of the 10 reasons for low INR, which had been assigned through chart review (see "Variables" above). We then used Cox regression to model the effect of these reasons on the time until the next INR and the next in-range INR, controlling for the same patient-level covariates described above. The patient's next in-range INR was identified using either a normal (2 to 3) or high (2.5 to 3.5) target INR range, corresponding to the target range stated in the clinical record. The reference category for these analyses was low INR because of a previous dose reduction. Such patients might be expected to have a relatively uniform risk for future INR instability, and to receive relatively uniform management. Our Cox models assumed separate baseline hazards by site of care, and our standard errors accounted for the fact that some patients contributed multiple low INR values (SAS PROC TPHREG).

We used linear interpolation[30] to divide all patient-time into 3 categories: days when the INR was ≤1.5, days when the INR was 1.5 to 2.0, and days when the INR was ≥2.0. For several patients, interpolation was not possible because the final INR determination occurred before a thromboembolic event; in such cases, we carried the last known INR value forward. We compared rates of thromboembolic events among these 3 categories of patient-time, using 1000 bootstrap resamples of the major event INR values to calculate 95% CIs. Finally, we calculated the proportion of thromboembolic events occurring during periods of low INR. Analyses were performed using the R statistical package version 2.8 (R Foundation) and SAS version 9.1. The authors had full access to the data and take

Table 1. Demographics, Indication for Anticoagulation, Comorbid Conditions, Profile of Care, and Number of Low INR Values Among Our Study Cohort (n=4489)

Variable	Patients Without Low INR (n=2934)	Patients With Low INR (n=1540)	*P* Value
Age, y, mean (SD)	70.9 (11.6)	70.9 (12.2)	0.70
Female gender, %	40.1	48.1	<0.001*
Race, %			0.31*
White	93.1	91.8	
Black	2.0	2.2	
All others	4.8	6.0	
Indication for anticoagulation, %			<0.001*
Atrial fibrillation without prior stroke	47.0	46.0	
Atrial fibrillation with prior stroke	5.4	6.4	
Venous thromboembolism	12.3	15.7	
Valvular heart disease: target INR 2.0–3.0	2.9	3.2	
Valvular heart disease: target INR 2.5–3.5	15.5	9.9	
Stroke/embolus	8.5	9.6	
All others	8.4	9.4	
Comorbid conditions, %			
Congestive heart failure	17.6	20.0	0.19*
Coronary artery disease	25.8	23.2	0.83*
Diabetes mellitus	14.8	16.6	0.15*
Hypertension	43.2	44.9	0.29*
Profile of care			
Months in database, mean (SD)	10.5 (4.5)	12.5 (4.2)	<0.001
No. of INR values, mean (SD)	14.6 (6.9)	22.4 (10.0)	<0.001
INR values/mo, mean (SD)	1.61 (1.07)	1.89 (0.75)	<0.001
Interruptions for procedures/y, mean (SE)	0.26 (0.57)	0.72 (0.99)	<0.001
INR values ≤1.5, mean (SD)	. . .	4.1 (4.5)	. . .
INR values ≤1.5, median (25–75)	. . .	3.0 (3.0)	. . .

P values account for intraclass correlation within clinical site.

**P* values for categorical variables calculated using 10 000 Monte Carlo simulations.

responsibility for its integrity. All authors have read and agree to the manuscript as written.

Results

Risk Factors for Low INR

Of the 4489 patients, 1540 (34%) had at least 1 low INR (Table 1). Most demographic and clinical parameters were similar between patients with and without low INR, with the exception of gender and holds. We examined risk factors for low INR using multivariable Poisson regression (Table 2).

Table 2. Patient-Level Risk Factors for INR of ≤1.5, Excluding Low INR Attributable to the Inception Phase of Warfarin Therapy, When Low INR Is to Be Expected (n=4489)

Predictor	Unadjusted IRR (95% CI)	Adjusted IRR (95% CI)
Age group, y		
≤64	. . .	. . .
65–74	0.80 (0.69, 0.92)*	0.84 (0.72, 0.98)*
≥75	0.94 (0.79, 1.13)	0.95 (0.81, 1.12)
Female sex	1.45 (1.29, 1.64)†	1.44 (1.28, 1.62)†
Race		
White	. . .	. . .
Black	1.51 (1.01, 2.26)*	1.25 (0.83, 1.88)
All others	1.16 (0.85, 1.58)	1.15 (0.87, 1.54)
Comorbid conditions		
Congestive heart failure	1.13 (0.94, 1.35)	1.17 (0.98, 1.39)
Coronary artery disease	0.96 (0.84, 1.08)	0.99 (0.87, 1.12)
Diabetes mellitus	1.12 (0.98, 1.28)	1.14 (1.01, 1.30)*
Hypertension	1.03 (0.93, 1.15)	0.98 (0.89, 1.07)
Interruptions for procedures (per interruption)	1.46 (1.34, 1.59)†	1.47 (1.37, 1.58)†
Indication for anticoagulation		
Atrial fibrillation without prior stroke	. . .	. . .
Atrial fibrillation with prior stroke	1.24 (1.03, 1.50)*	1.26 (1.03, 1.54)*
Venous thromboembolism	1.60 (1.32, 1.94)†	1.48 (1.24, 1.75)†
Valvular heart disease: target INR 2.0–3.0	1.40 (1.05, 1.88)*	1.37 (1.04, 1.82)*
Valvular heart disease: target INR 2.5–3.5	0.56 (0.46, 0.68)†	0.59 (0.49, 0.70)†
Stroke/embolus	1.12 (0.90, 1.40)	1.12 (0.92, 1.36)
All others	1.15 (0.91, 1.45)	1.22 (0.99, 1.50)

We used a Poisson regression model for count data; patients could have more than one low INR value. The adjusted analysis adjusts for all other predictors in the table, and confidence intervals account for intraclass correlation within the site of care using generalized estimating equations. Results are expressed as incidence rate ratios relative to the reference category.

**P*<0.05.

†*P*<0.001.

Women had an increased incidence of low INR (incidence rate ratio [IRR], 1.44; $P<0.001$). Compared with the reference category (AF without prior stroke), patients anticoagulated for valvular heart disease with a high INR target had less low INR (IRR, 0.59; $P<0.001$). Patients anticoagulated for VTE, in contrast, had more low INR than the reference category (IRR, 1.48; $P<0.001$), as did patients with valvular heart disease and a normal INR target (IRR, 1.37; $P<0.05$). Warfarin holds independently predicted low INR (IRR, 1.47 per hold; $P<0.001$).

Reasons for Low INR

Clinicians managing anticoagulation gave 10 reasons to explain the 3456 low INR values (Table 3). Five reasons

Table 3. Relative Frequencies of 10 Reasons Given by Clinicians for INR ≤1.5 (n=3456 INR Values)

Reason	Frequency, %
Adherence issues	17
Holds for procedures	16
No reason apparent after questioning	15
Dose reduced in reaction to high INR or bleeding	15
Continuing low value (ie, second or third consecutive low INR)	13
Initiation of therapy	8
Dietary intake of vitamin K	6
Interaction with other medications	4
All other reasons	3
No data/unknown	2

collectively accounted for three quarters of low INR values: nonadherence (17%), holds (16%), no reason apparent after questioning (15%), dose reductions attributable to high INR or bleeding (15%), and continuing low values (13%). The initiation phase of warfarin therapy (8%), dietary intake of vitamin K (6%), and interactions with other medications (4%) accounted for most remaining low INR values. The frequencies of these reasons were not meaningfully different when compared by age group, gender, race, and indication for therapy.

We further examined the category of "continuing low." The mean time between a continuing low value and the preceding INR value was 8.1 days (SD, 5.6); the median was 7 days (interquartile range, 4 to 10). The distribution of reasons for a low INR preceding a continuing low differed from the overall distribution of reasons for low INR values (Table 4). In particular, the most frequent reasons for the INR before a continuing low were another continuing low (22%), dose reductions (20%) and holds (20%). These 3 reasons were much more common preceding a continuing low than in the overall sample; other reasons were reduced accordingly.

Care Provided to Address Low INR

After a low INR, the median time until the next INR was 8 days (interquartile range, 7 to 14), and the median time until the next in-range INR was 16 days (interquartile range, 8 to 30). The next INR occurred sooner (Table 5) when the patient

Table 4. Reason for the Low INR Value Immediately Preceding a "Continuing Low" Value (n=442)

Reason	Frequency, %
Continuing low value	22
Dose reduced in reaction to high INR or bleeding	20
Holds for procedures	20
No reason apparent after questioning	12
Adherence issues	9
Interaction with other medications	5
Dietary intake of vitamin K	6
All other reasons	3
No data/unknown	2

Table 5. Hazard Ratios for Time to Next INR Value Among 3337 Low INR Values and for Time to Next In-Range INR Value Among 3165 Low INR Values

Parameter	Time to Next INR Test, HR (95% CI)	Time to Next In-Range INR, HR (95% CI)
Indication for anticoagulation		
Atrial fibrillation	. . .	. . .
Venous thromboembolism	1.05 (0.94, 1.16)	0.95 (0.85, 1.05)
Valvular heart disease/prosthetic valve	1.64 (1.45, 1.84)†	1.02 (0.90, 1.15)
Prior stroke/embolus	1.03 (0.91, 1.17)	1.10 (0.96, 1.25)
All others	0.98 (0.86, 1.12)	0.90 (0.78, 1.03)
Stated reason for low value		
Dose reduced due to high INR or bleeding	. . .	. . .
Adherence issues	0.78 (0.69, 0.89)†	0.98 (0.86, 1.11)
Continuing low value	1.16 (1.02, 1.33)*	0.97 (0.84, 1.11)
Dietary intake of vitamin K	0.77 (0.65, 0.91)*	0.87 (0.73, 1.03)
Holds for procedures	0.90 (0.80, 1.03)	1.13 (0.99, 1.29)
Initiation of therapy	1.93 (1.64, 2.26)†	1.33 (1.13, 1.56)†
Interaction with other medications	0.97 (0.80, 1.17)	1.07 (0.88, 1.29)
All other reasons	0.99 (0.80, 1.22)	1.01 (0.81, 1.26)
No reason apparent after questioning	0.80 (0.70, 0.91)†	1.00 (0.87, 1.14)
No data/unknown	0.70 (0.51, 0.96)*	0.77 (0.55, 1.07)

A hazard ratio above 1.0 indicates reduced time until the event. These analyses control for demographics (age, gender, and race). Our Cox models assumed separate baseline hazards by site of care, and our standard errors accounted for the fact that some patients contributed multiple low INR values.

*$P<0.05$.

†$P<0.001$.

was anticoagulated for valvular heart disease (hazard ratio [HR], 1.64; $P<0.001$), but these patients did not record an in-range INR sooner than other patients (HR, 1.01; $P=0.91$). The next INR also occurred sooner during the initiation phase of warfarin therapy (HR, 1.93; $P<0.001$); an in-range INR was also recorded sooner for this category (HR, 1.33; $P<0.001$). Conversely, patients whose low INR values were attributed to nonadherence, dietary vitamin K, and no apparent reason waited longer for a repeat INR than the reference category (HR, 0.78, 0.77, and 0.80; $P<0.05$ for all). In general, with the exception of initiation-phase patients, differences in time until next INR test did not translate into differences in time until the next in-range INR.

Impact of Low INR on Thromboembolic Events

There were 55 major thromboembolic events during the study (Table 6), 12 of which occurred when the INR was ≤1.5, 9 of which when the INR was 1.5 to 2.0, and 34 when the INR was ≥2.0. Of the 12 events that occurred when the INR was ≤1.5,

Table 6. Rates of Thromboembolic Events in Different INR Ranges

INR Value	Thromboembolic Events	Patient-Time, y	Rate, Events/100 Patient-y (95% CI)	Unadjusted IRR (95% CI)	Fraction of Events, % (95% CI)
INR ≤1.5	12	70.8	16.95 (8.5–26.8)	16.3 (8.1–25.7)	21.8 (12.2–35.4)
INR 1.5–2.0	9	545.1	1.65 (0.73–2.75)	1.59 (0.70–2.64)	16.4 (8.2–29.3)
INR ≥2.0	34	3262.3	1.04 (0.83–1.26)	1.0 (reference)	61.8 (47.7–74.3)
Total	55	3878.2	1.42 (1.11–1.95)	. . .	100

INR values determined by linear interpolation.

7 were associated with a hold (58%). The crude IRR for the lowest INR category, compared with INR ≥2.0, was 16.3 (95% CI, 8.1 to 25.7), whereas the IRR for mildly low INR was only 1.59 (95% CI, 0.70 to 2.64). The fraction of thromboembolic events occurring with an INR value ≤1.5 was 12/55, or 21.8% (95% CI, 12.2% to 35.4%). However, an additional 7 events occurred within 30 days of an INR ≤1.5, so the true fraction could be as high as 19/55, or 34.5% (95% CI, 22.6% to 48.7%).

Although the relative rate of thromboembolism during periods of low INR was high, the absolute rate of thromboembolism per episode of low INR was quite low. There were a total of 3375 separate episodes of INR ≤1.5 in our database, considering continuing low values as part of a single episode. A thromboembolic event occurred during or within 30 days after the end of only 19 of these episodes (0.6%; 95% CI, 0.3 to 0.9%), whereas the majority of episodes (99.4%) did not result in a thromboembolic event.

Discussion

We examined patient-level risk factors for low INR (≤1.5) using a large nationally-representative database of anticoagulation care. Low INR was not a rare event, occurring among 34% of our study population and accounting for 4.1% of INR values and 1.8% of patient-time. Low INR also had important consequences: there was an approximately 16-fold increase in the rate of thromboembolism during periods of low INR. We found that the fraction of thromboembolic events occurring during periods of low INR was at least 21.8%, considerably higher than the estimate by van Walraven et al (11%).[24] Our results suggest that reducing or eliminating low INR could improve patient outcomes considerably.

Despite the impressive relative risk of a thromboembolic event during periods of low INR, the absolute risk of thromboembolism related to an episode of low INR was small (0.6%). Therefore, careful consideration should be given to balancing expected risks and benefits of interventions such as "bridging" with low-molecular-weight heparin, which is associated with a significant risk of bleeding complications. Ongoing clinical trials may settle the issue of whether, and for which patients, the benefits of bridging outweighs the risks.[31] Until the results of such trials become available, our estimate of the absolute risk of thromboembolism related to an episode of low INR may help to guide clinical decision-making.

Regarding risk factors for low INR, warfarin holds predicted more low INR values, whereas a high-target INR range predicted fewer. More novel and unexpected findings included increased incidence of low INR among patients anticoagulated for VTE and among women. These findings, which persisted after controlling for age, comorbid conditions, and holds, are potentially important, but require confirmation and further investigation. One possible explanation might have been a difference in target INR ranges between groups; however, we compared target ranges by gender and by indication and did not find differences that could have explained our findings. If confirmed, our findings may be attributable to different physiological responses to warfarin in different groups of patients, or may reflect disparities in anticoagulation management. For example, it is possible that clinicians fear the consequences of high INR more in female patients, which may affect dosing decisions.

We also examined reasons given to explain low INR values in the clinical record. Although no single reason predominated, the 4 most common reasons bear comment. Nonadherence was the most commonly cited reason for low INR. The impact of adherence on anticoagulation control has been described by Kimmel et al.[32] In that study, 36% of patients missed more than 20% of bottle openings as measured by electronic bottle caps ("MEMS caps"); these patients had an odds ratio of 2.10 for INR below the target range. In our study, 17% of INR values ≤1.5 (a more serious deviation than merely below the target range) were attributed to adherence. We note that 15% of low INR values in our cohort could not be explained—it is possible that at least some of these were also attributable to inadequate adherence, which the patient did not recall or did not declare.

The next most common reason for low INR was intentional interruptions of warfarin for procedures ("holds"). Our group has previously demonstrated that holds are associated with the relatively poor anticoagulation control experienced by cancer patients.[28] Although some holds may be necessary, our results suggest that avoiding holds whenever possible will reduce low INR. As an example of a situation in which a hold may be avoided, dental procedures can often be performed without holds,[33,34] but this may be inconsistently applied in clinical practice.

The third most common reason for low INR was a recent dose reduction, most often in response to a previously recorded high INR value. This seesaw effect, where patients bounce between excessive and insufficient anticoagulation, may be partly attributed to the well-known fact that warfarin is a difficult drug to manage in clinical practice.[5] However, at least part of this erratic control may be associated with excessive "tinkering" with warfarin doses when the INR is close to the target range.[27] In another analysis of this database, our group has already shown that INR control could be improved by reserving dose changes for patients whose

INR deviates from the target range by 0.3 or more in either direction.[27]

The fourth most common reason for low INR was a "continuing low." Recall that we defined "low INR" as ≤1.5—a level of underanticoagulation clearly associated with patient harm.[22–24] Nevertheless, we found that the median time until a next INR test was 8 days, and the median time until a next in-range INR value was 16 days. These values themselves are not alarming, but the 75th percentile for each (14 and 30 days) indicate that for a considerable proportion of patients in our dataset, low INR was addressed without a particular sense of urgency. Indeed, we found that the most common reason for the INR preceding a continuing low was another continuing low.

Current guidelines may reinforce a lack of urgency in addressing low INR values, because they contain limited guidance about how to address low INR. For example, the 2008 ACCP guidelines[5] contain detailed instructions about how to deal with elevated INR, but limited information regarding how to address low INR, beyond the recommendation that "bridging" with low-molecular-weight heparin is not necessary for most patients. Although we agree that low-molecular-weight heparin is not warranted by the relatively low daily risk of thromboembolism in most patients,[35] it would be prudent to measure the INR weekly among patients with changes in clinical status or anticipated dose instability, to prevent prolonged or extreme deviations from the target INR range. We recognize, however, that some patients may not readily accept the burden of such frequent testing.

Our study has important strengths. This is the first systematic investigation of low INR in community practice. We used a large, nationally-representative database of community-based anticoagulation care in the United States, ensuring that our results are broadly generalizable. Finally, manual review of all 84 915 notes provided a level of clinical detail missing from many previous studies, which have used predominantly automated data. However, our study also has limitations. We were only able to ascertain the reason for low INR as stated by clinicians in the notes, but were unable to test the veracity of such claims. In addition, the reason for 15% of low values could not be determined by the clinician at the time of the visit, despite questioning the patient.

In conclusion, we used a nationally-representative database of community-based anticoagulation care to describe the epidemiology of low INR. In our study, low INR was associated with a 16-fold increase in the rate of thromboembolism. Despite this impressive relative risk, the absolute risk of thromboembolism per episode of low INR was only 0.6% per episode. Women and patients anticoagulated for VTE were at elevated risk for low INR. Nonadherence, interruptions for procedures, and insufficient urgency in addressing low INR all contribute to the incidence of low INR. The incidence of low INR could be reduced by interventions to improve adherence, minimize unnecessary interruptions of therapy, and encourage clinicians to address low INR with an appropriate (but not excessive) sense of urgency.

Sources of Funding

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; 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 Department of Veterans Affairs Health Services Research and Development Service. Dr Grant is supported by an NIDDK career development award (K23 DK067452). The opinions expressed in this manuscript do not necessarily represent the views or policies of the Department of Veterans Affairs.

Disclosures

Dr Rose has received research funding from Boehringer Ingelheim to study warfarin utilization in the Veterans Affairs Health System. Dr Hylek has served as an advisor to Boehringer Ingelheim, Bristol-Myers Squibb, Sanofi-Aventis, and The Medicines Company and participated in a clinical symposium sponsored by Bayer and Bristol-Myers Squibb.

References

1. Kearon C, Kahn SR, Agnelli G, Goldhaber S, Raskob GE, Comerota AJ. Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest.* 2008;133:454S–545S.
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, Hunt SA, Nishimura R, Ornato JP, Page RL, Riegel B, Priori SG, Blanc JJ, Budaj A, Camm AJ, Dean V, Deckers JW, Despres C, Dickstein K, Lekakis J, McGregor K, Metra M, Morais J, Osterspey A, Tamargo JL, Zamorano JL. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation. *Circulation*. 2006;114:e257–e354.
3. Singer DE, Albers GW, Dalen JE, Fang MC, Go AS, Halperin JL, Lip GY, Manning WJ. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest.* 2008;133:546S–592S.
4. Salem DN, O'Gara PT, Madias C, Pauker SG. Valvular and structural heart disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest.* 2008;133:593S–629S.
5. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest.* 2008;133:160S–198S.
6. van Walraven C, Jennings A, Oake N, Fergusson D, Forster AJ. Effect of study setting on anticoagulation control: a systematic review and metaregression. *Chest*. 2006;129:1155–1166.
7. White HD, Gruber M, Feyzi J, Kaatz S, Tse HF, Husted S, Albers GW. Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. *Arch Intern Med*. 2007;167:239–245.
8. Beyth RJ, Quinn L, Landefeld CS. A multicomponent intervention to prevent major bleeding complications in older patients receiving warfarin. A randomized, controlled trial. *Ann Intern Med*. 2000;133:687–695.
9. Beyth RJ, Quinn LM, Landefeld CS. Prospective evaluation of an index for predicting the risk of major bleeding in outpatients treated with warfarin. *Am J Med*. 1998;105:91–99.
10. Fihn SD, Callahan CM, Martin DC, McDonell MB, Henikoff JG, White RH. The risk for and severity of bleeding complications in elderly patients treated with warfarin. The National Consortium of Anticoagulation Clinics. *Ann Intern Med*. 1996;124:970–979.
11. Gage BF, Yan Y, Milligan PE, Waterman AD, Culverhouse R, Rich MW, Radford MJ. Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF). *Am Heart J*. 2006;151:713–719.
12. Hylek EM, Heiman H, Skates SJ, Sheehan MA, Singer DE. Acetaminophen and other risk factors for excessive warfarin anticoagulation. *JAMA*. 1998;279:657–662.
13. Hylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. *Ann Intern Med*. 1994;120:897–902.

14. Kakar P, Lane D, Lip GY. Bleeding risk stratification models in deciding on anticoagulation in patients with atrial fibrillation: a useful complement to stroke risk stratification schema. *Chest*. 2006;130:1296–1299.
15. Levine MN, Raskob G, Beyth RJ, Kearon C, Schulman S. Hemorrhagic complications of anticoagulant treatment: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;126: 287S–310S.
16. Palareti G, Leali N, Coccheri S, Poggi M, Manotti C, D'Angelo A, Pengo V, Erba N, Moia M, Ciavarella N, Devoto G, Berrettini M, Musolesi S. 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.
17. Hylek EM, Regan S, Go AS, Hughes RA, Singer DE, Skates SJ. Clinical predictors of prolonged delay in return of the international normalized ratio to within the therapeutic range after excessive anticoagulation with warfarin. *Ann Intern Med*. 2001;135:393–400.
18. Garcia DA, Regan S, Crowther M, Hylek EM. The risk of hemorrhage among patients with warfarin-associated coagulopathy. *J Am Coll Cardiol*. 2006;47:804–808.
19. Koo S, Kucher N, Nguyen PL, Fanikos J, Marks PW, Goldhaber SZ. The effect of excessive anticoagulation on mortality and morbidity in hospitalized patients with anticoagulant-related major hemorrhage. *Arch Intern Med*. 2004;164:1557–1560.
20. van der Meer FJ, Rosendaal FR, Vandenbroucke JP, Briet E. Bleeding complications in oral anticoagulant therapy. An analysis of risk factors. *Arch Intern Med*. 1993;153:1557–1562.
21. Visser LE, Bleumink GS, Trienekens PH, Vulto AG, Hofman A, Stricker BH. The risk of overanticoagulation in patients with heart failure on coumarin anticoagulants. *Br J Haematol*. 2004;127:85–89.
22. Hylek EM, Skates SJ, Sheehan MA, Singer DE. An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with non-rheumatic atrial fibrillation. *N Engl J Med*. 1996;335:540–546.
23. Palareti G, Legnani C, Cosmi B, Guazzaloca G, Cini M, Mattarozzi S. Poor anticoagulation quality in the first 3 months after unprovoked venous thromboembolism is a risk factor for long-term recurrence. *J Thromb Haemost*. 2005;3:955–961.
24. van Walraven C, Oake N, Wells PS, Forster AJ. Burden of potentially avoidable anticoagulant-associated hemorrhagic and thromboembolic events in the elderly. *Chest*. 2007;131:1508–1515.
25. Rose AJ, Ozonoff A, Henault LE, Hylek EM. Warfarin for atrial fibrillation in community-based practise. *J Thromb Haemost*. 2008;6: 1647–1654.
26. Garcia D, Regan S, Crowther M, Hughes RA, Hylek EM. Warfarin maintenance dosing patterns in clinical practice: implications for safer anticoagulation in the elderly population. *Chest*. 2005;127:2049–2056.
27. Rose AJ, Ozonoff A, Berlowitz DR, Henault LE, Hylek EM. Warfarin dose management affects INR control. *J Thromb Haemost*. 2009;7: 94–101.
28. 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.
29. Garcia DA, Regan S, Henault LE, Upadhyay A, Baker J, Othman M, Hylek EM. Risk of thromboembolism with short-term interruption of warfarin therapy. *Arch Intern Med*. 2008;168:63–69.
30. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost*. 1993;69:236–239.
31. National Institutes of Health: Effectiveness of Bridging Anticoagulation for Surgery (The BRIDGE Study). ClinicalTrials.gov identifier: NCT00786474. www.clinicaltrials.gov. Accessed March 2, 2009.
32. Kimmel SE, Chen Z, Price M, Parker CS, Metlay JP, Christie JD, Brensinger CM, Newcomb CW, Samaha FF, Gross R. The influence of patient adherence on anticoagulation control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Arch Intern Med*. 2007;167:229–235.
33. Sindet-Pedersen S, Ramstrom G, Bernvil S, Blomback M. Hemostatic effect of tranexamic acid mouthwash in anticoagulant-treated patients undergoing oral surgery. *N Engl J Med*. 1989;320:840–843.
34. Souto JC, Oliver A, Zuazu-Jausoro I, Vives A, Fontcuberta J. Oral surgery in anticoagulated patients without reducing the dose of oral anticoagulant: a prospective randomized study. *J Oral Maxillofac Surg*. 1996;54:27–32 discussion 323.
35. 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;28:960–967.

JOURNAL OF WOMEN'S HEALTH
Volume 18, Number 2, 2009

DOI: 10.1089/jwh.2007.0486

Beliefs of Women's Risk as Research Subjects: A Four-City Study Examining Differences by Sex and by Race/Ethnicity

Stefanie L. Russell, D.D.S.,[1] Ralph V. Katz, D.M.D., M.P.H., Ph.D.,[1] Nancy R Kressin, Ph.D.,[2] B. Lee Green, Ph.D.,[3] Min Qi Wang, Ph.D.,[4] Cristina Claudio, Ph.D.,[5] and Krassimira Tzvetkova, D.D.S., M.S.D., M.S. C.R.[6]

Abstract

Background: Given the history of vulnerability of women of childbearing age to medical treatments that have caused injury, for example, diethylstilbestrol (DES) and thalidomide, it is surprising that, to date, little research has directly examined attitudes of the general public regarding the vulnerability of women when they participate in biomedical research studies.
Methods: We asked three questions about beliefs of women as biomedical research subjects of 623 white, 353 black, and 157 Hispanic people in four U.S. cities: (1) Do you believe that women are more likely to be "taken advantage of" when they become subjects in a medical research project as compared to men? (2) Do you believe that women of childbearing age (15–45-year-olds) should become study participants in medical research projects? and, if the response was no or don't know/not sure, (3) Would you still say no or don't know/not sure to question 2 even if it meant that we would not know anything about the health and medical treatments for women aged 15–45 years?
Results: Overall, women were 60% more likely than men to state that women were more likely than men to be "taken advantage of," even when controlling for potential confounders, and both black and Hispanic participants were much more likely than white participants to state that this was the case. The majority of respondents (57.4%) said that women of childbearing age should not be research subjects; among women, both black and Hispanic people were less likely than white people to change their minds when prompted that this might mean that "nothing would be known about the health and medical treatments for women aged 15–45 years."
Conclusions: A substantial proportion of the participants reported knowledge of historical events, and this knowledge was related, particularly in black participants, to attitudes toward vulnerability of women as biomedical research subjects.

Introduction

In light of a history of unethical human experimentation in the United States (e.g., the Tuskegee Syphillis Study), it is not remarkable that much has been written about actual or perceived vulnerability of racial and ethnic minority research subjects.[1–6] Given the history of vulnerability of women, specifically women of childbearing age, to medical treatments in the past that have caused injury not only to them but to their children, for example, diethylstilbestrol (DES)[7] and thalidomide,[8] it is surprising that, to date, little research has directly examined attitudes of the general public about the vulnerability of women when they participate in biomedical research studies. Although investigations reporting on inclusion of women in studies have documented sex differences in clinical trial recruitment rates[9–12] or have reported on reasons women decline to participate in clinical trials,[13–16] there are no reports of beliefs and attitudes among the lay public regarding whether women, compared with men, are more likely to be "taken advantage of" when they become subjects in biomedical research studies, whether the benefits of including women of childbearing age outweigh

[1]New York University College of Dentistry, Epidemiology and Health Promotion, New York, New York.
[2]Department of Veterans Affairs and Center for Health Quality, Outcomes & Economic Research, Bedford, Massachusetts.
[3]H. Lee Moffitt Cancer Center & Research Institute, Office of Institutional Diversity, Tampa, Florida.
[4]University of Maryland Department of Public and Community Health, College Park, Maryland.
[5]School of Dentistry, University of Puerto Rico, Community Dentistry Section Medical Sciences Campus, San Juan, Puerto Rico.

the potential risks involved in their participation in these studies, and whether beliefs vary between the sexes or by race/ethnicity.

The purpose of this analysis was to examine (1) attitudes about women as subjects in biomedical research in a geographically, sociodemographically, ethnically, and racially diverse group of adults, (2) whether beliefs about perceived risks to women when they participate in biomedical research studies vary by sex or by race/ethnicity, and (3) the relationship of knowledge of two historical events (DES and thalidomide) to attitudes of whether women should participate in biomedical research studies.

Materials and Methods

The Tuskegee Legacy Project (TLP) questionnaire, a survey designed to examine reasons for lack of participation of racial and ethnic minorities, was administered via random-digit dial (RDD) telephone interviews to participants aged ≥18 years in four cities: Birmingham, Alabama; Tuskegee, Alabama; Hartford, Connecticut, and San Antonio, Texas. The TLP questionnaire, a 60-item instrument, primarily designed to address a range of issues related to the recruitment of minorities into biomedical studies, also contains questions related to inclusion of women in biomedical research. Details on the history and development of the TLP questionnaire and justifications of the methodological decisions both for the selection of the four cities and for the analysis of the TLP questionnaire have been published elsewhere.[1,17] This study was approved by the Institutional Review Boards of the University of Connecticut Health Center and of New York University.

The Survey Research Unit (SRU) of the University of Alabama at Birmingham (UAB) administered the TLP questionnaire via RDD telephone interview. The target population was noninstitutionalized persons aged ≥18 years living in households with working telephones in the four targeted cities. The goal was to complete the 25-minute telephone interview with 900 adults in the following racial/ethnic groups: (1) 300 black respondents (100 in Hartford, 100 in Birmingham, 100 in Tuskegee, AL), (2) 100 Puerto Rican Hispanics (Hartford), (3) 100 Mexican Americans (San Antonio), and (4) 400 white adults (100 in Hartford, 100 in Birmingham, 100 in Tuskegee, 100 in San Antonio). A simple random sample was drawn based on the three-digit telephone exchange used for local calling areas within each city. SRU at UAB screened for nonworking and business numbers; unresolved numbers were retired after 20 attempts. A total of 13 interviewers were trained for the survey using computer-assisted telephone interview (CATI) technology; these interviewers were supervised at all times and randomly electronically monitored.

Key variables from the TLP questionnaire

Figure 1 shows the three questions from the TLP (the Women as Research Subjects questions, or WARS domain) that specifically explored attitudes about women's participation in biomedical studies. These questions were around the midpoint of the questionnaire (questions 34–36 of 60) directly after questions that explored attitudes regarding racial and ethnic minorities' vulnerability as biomedical research subjects. The general concept of vulnerability of women as research subjects is introduced in the first question, and the

1. Do you believe that WOMEN in the US are more likely to be "taken advantage of" when they become subjects in a medical research project as compared to MEN? Would you say that was:
__ Always __ Most of the time __ Some of the time __ Rarely __ Never

2. Do you believe that women of childbearing age (15-45 year olds) should become Study participants in medical research projects?
__ Yes __ No __ Don't know/Not sure

IF the response to Q. #2 was "NO" or "Don't know/Not sure:

3. Would you still say . . .
A) "NO" that women of childbearing age should NOT be subjects in medical research studies
or
B) "Don't know/not sure" whether women of childbearing age should be subjects in medical research studies
Even if it meant that we would not know anything about the health and medical treatments for women aged 15-45 yrs.?
__ I would still say "NO" or "Don't know/Not sure"
__ That changes my mind, so YES they should be included.

FIG. 1. Questions that explored attitudes toward the vulnerability of women as participants in biomedical research; TLP questionnaire.

second question specifically asks if women of childbearing age should be included in research studies. The third question is a probe for those who responded No or Not sure about whether they think women of childbearing age should be included in research studies; that is, the interviewer informed the respondent about the potential negative consequences of a total lack of knowledge about women's health for women of childbearing age that would inevitably result and inquired if this consideration changed their mind.

In addition, because we were also interested in determining if knowledge of DES and thalidomide was related to responses on the WARS domain questions, people were asked: Which of the following medical research studies have you heard about? Respondents answered either Yes, No, or don't know/not sure to a series of eight studies, one described as the DES or Diethylstilbestrol Study and another as the Thalidomide Study. For purposes of these analyses, we grouped those who stated No/don't know with those who responded No. We hypothesized that knowledge of DES and thalidomide would vary by age, by sex, by racial/ethnic group, and by education and income. Age was calculated using date of birth, and we classified education as either less than high school graduate, high school graduate alone or with some college vs. college degree or greater. We categorized income as <$20,000 a year, ≥$20,000 but <$75,000 a year vs. ≥$75,000 a year.

Statistical analysis

For bivariate analyses, study participants were stratified by sex and by racial/ethnic category and compared regarding the WARS domain questions using Pearson chi-square analyses and ANOVA. The responses to the WARS domain questions were collapsed for the purpose of bivariate and multivariate analyses. We collapsed responses of question 1 into two categories: Yes (all, most, and some of the time) vs. No (rarely and never). Questions 2 and 3 were also collapsed into two categories for the purpose of bivariate and logistic regression analyses: Yes vs. No (no and don't know/not sure). We used logistic regression in order to control for potential confounders, including age, sex, race/ethnicity, education, income, and city/county. We report adjusted odds ratios (OR) and 95% confidence intervals (CI).

Results

Characteristics of study participants

The TLP questionnaire was administered to 1133 adults in Birmingham, Tuskegee, Hartford, and San Antonio, with response rates of 70%, 65%, 49%, and 50%, respectively. The final study sample comprised 353 black (31.2%), 157 Hispanic (15.9%), and 623 white respondents (55.0%), and 51.7% of respondents were women. Hispanics self-identified as either Puerto Rican (24.8%) or Mexican American (75.2%). The demographic characteristics of the participants are shown in Table 1. Hispanic participants were younger than black and white participants (ANOVA, $p \leq 0.001$) and were more likely to be women than were white and black respondents (chi-Square 7.0, *df* 2, $p = 0.03$). White participants were more educated (chi-square = 65.6, *df* 4, $p \leq 0.001$) and had higher incomes (chi-square 80.7, *df* 4, $p \leq 0.001$) compared to black or Hispanic participants.

Analysis of question 1: Do you believe that women in the United States are more likely to be "taken advantage of" when they become subjects in a medical research project as compared to men?

Overall, for the total respondent sample, a majority of persons (52.1%) stated that women, compared to men, were more likely to be "taken advantage of" as research subjects either always or most of the time (17.3%) or sometimes (34.8%) (Fig. 2). Overall, women were more likely to state that this was the case always, most of the time, or sometimes (19.1% of women vs. 15.5% of men; chi-square 13.0. *df* 2, $p = 0.002$). Blacks were nearly 5 times more likely as whites and Hispanics were nearly 3.5 times more likely as whites to state

Table 1. Demographic Characteristics of Study Participants, TLP

	Race/ethnicity			
	White (n = *623*)	*Black* (n = *353*)	*Hispanic* (n = *157*)	*Total* (n = *1133*)
Sex (% women)	51.7%	47.9%	60.5%	51.7%
Age (years, mean ± SD)	53.8 ± 17.0	49.1 ± 16.5	41.5 ± 16.1*	50.6 ± 17.2
Education				
< High school graduate	11.8%	21.6%	26.1%	16.8%
High school graduate	51.3%	60.5%	58.6%	55.2%
≥ college graduate	36.9%*	17.9%	15.3%	28%
Income				
< $20,000	21.3%	42.8%	41.7%	31.1%
$20,000–74,999	58.4%	52.1%	52.5%	55.5%
≥ $75,000	20.3%*	5.1%	5.8%	13.3%
Geographic location*				
Birmingham, AL	16.9%	29.5%	0.6%	18.5%
Tuskegee, AL	29.5%	30.0%	0.6%	25.7%
Hartford, CT	35.6%	30.6%	23.6%	34.1%
San Antonio, TX	18.0%	4.5%	75.2%	21.7%

*Significant at the $p \leq 0.05$ level.

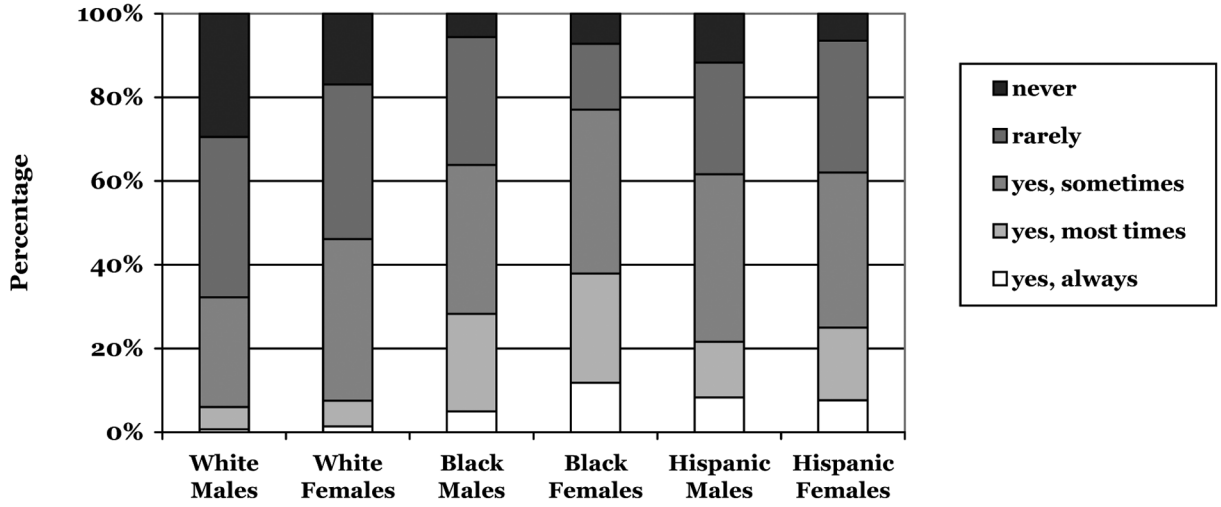


FIG. 2. Response to question 1: Percentage, by race/ethnicity and gender, who believe either that women, compared to men, are more likely to be taken advantage of when they become research subjects ($n = 1133$).

that this was the case always or most of the time (blacks 32.7%, Hispanics 23.7% vs. whites 6.8%; chi-square 133.1, *df* 4, $p \leq 0.001$). When we examined differences by sex within each racial/ethnic category we found that among both white and black respondents, women were more likely than men to state that women as research subjects are more likely "taken advantage of" (whites: 46.1% of women vs. 32.2% of men, chi-square 11.8, *df* 2, $p = 0.003$; blacks: 77.1% of women vs. 63.9% of men, chi-square 7.4, *df* 2, $p = 0.024$). Among Hispanics, however, men and women did not significantly differ in their responses (62.0% of women vs. 61.7% of men, chi-square 0.30, *df* 2, $p = 0.88$).

When controlling for differences in age, income, and race/ethnicity between the sexes, the odds of believing that women are more likely to be "taken advantage of" when they become subjects in a medical research study was 60% higher for women compared with men ($OR_{adjusted} = 1.6$, 95% CI 1.2, 2.1). When controlling for differences in age, income, education, and geographic location between racial/ethnic groups among women only, we found that the odds of believing that women compared to men are more likely to be "taken advantage of" when they become subjects in a medical research project were four times higher in black women ($OR_{adjusted} = 3.9$, 95% CI 2.5, 6.2) and two times higher in Hispanic women ($OR_{adjusted} = 1.9$, 95% CI 1.2, 3.1) compared to white women. For men, the odds of believing that women compared to men are more likely to be "taken advantage of" when they become subjects in a medical research project were also higher

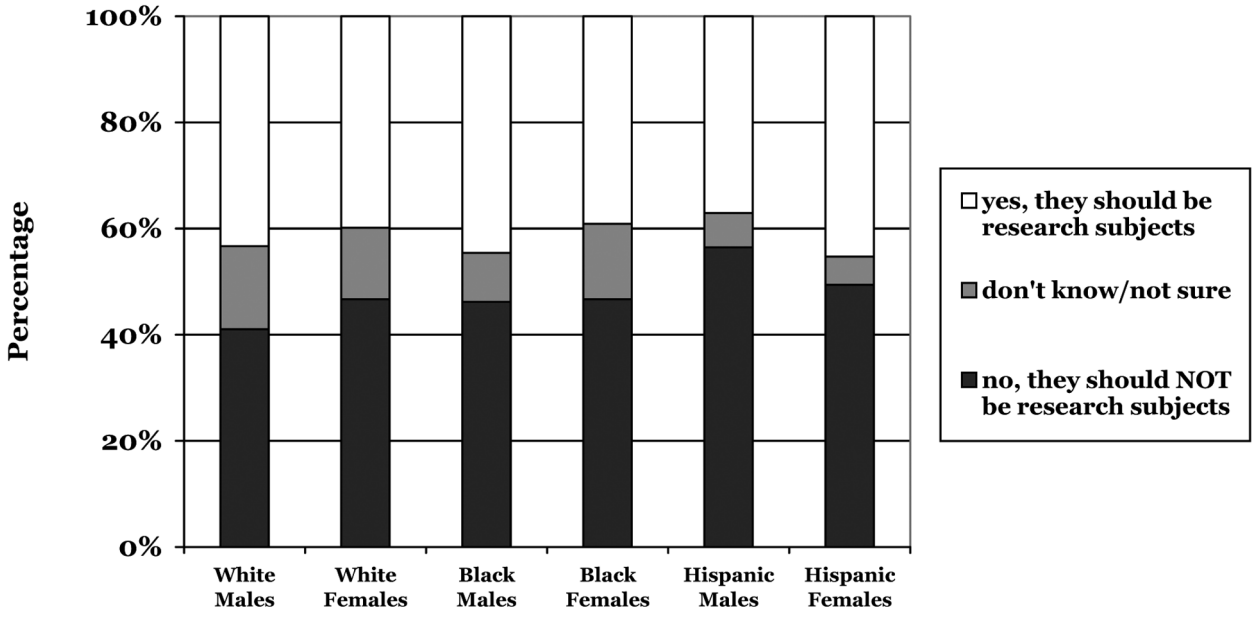


FIG. 3. Response to question 2: Percentage, by race/ethnicity and gender, who believe either that women of childbearing age should not become research subjects or are not sure they should become research subjects ($n = 1133$).

in both black men ($OR_{adjusted}$ = 3.6, 95% CI 2.3, 5.5) and Hispanic men ($OR_{adjusted}$ = 3.1, 95% CI 1.7, 6.1) compared to white men.

When we examined responses to question 1 by sex within each race/ethnic group, we found that for white and black participants, women were more likely to believe that women would be "taken advantage of" when they become subjects in a medical research project compared to men (controlling for differences in sex in income level for whites and in age for Blacks) whites: $OR_{adjusted}$ = 1.7, 95% CI 1.2, 2.4; blacks: $OR_{adjusted}$ = 1.8, 95% CI 1.1, 3.0). For Hispanics, the odds of believing that women were more likely compared to men to be "taken advantage of" when they become subjects in a medical research project did not vary by sex in logistic regression analyses.

Analysis of question 2: Do you believe that women of childbearing age (15–45-year-olds) should become study participants in medical research projects?

Overall, 45.0% of the respondents said that women of childbearing age should not be research subjects, and an additional 12.4% were unsure whether they should be research subjects (Fig. 3). In order to test for differences by sex and by race/ethnicity when controlling for differences between groups, we grouped those who stated Don't know/not sure with those who stated No. We found no differences in the distribution of responses to question 2 between men and women, or between white, black, and Hispanic respondents in the overall sample in either bivariate or multivariate analyses, controlling for differences in age, geographic location, income, and education between the sexes (for women, compared to men, the $OR_{adjusted}$ = 1.1, 95% CI 0.8, 1.4) and controlling for differences in sex, age, and income between the racial/ethnic groups (for blacks, compared to whites, the $OR_{adjusted}$ = 1.0, 95% CI 0.8, 1.4; for Hispanics, compared to whites, the $OR_{adjusted}$ = 1.2, 95% CI 0.8, 1.9). In addition, when we examined responses to question 2 within each sex group, we found that for both men and for women, there was no significant difference by race/ethnicity (black women vs. white women, $OR_{adjusted}$ = 1.1, 95% CI 0.7, 1.5; Hispanic women vs. white women, $OR_{adjusted}$ = 0.9, 95% CI 0.5, 1.7; black men vs. white men, $OR_{adjusted}$ = 1.2, 95% CI 0.7, 1.8; Hispanic men vs. white men, $OR_{adjusted}$ = 1.4, 95% CI 0.6, 3.1). We also failed to find differences in responses by sex within each racial/ethnic group (white men vs. white women, $OR_{adjusted}$ = 1.2, 95% CI 0.8, 1.8; black men vs. black women, $OR_{adjusted}$ = 0.8, 95% CI 0.7, 1.1; Hispanic men vs. Hispanic women, $OR_{adjusted}$ = 0.5, 95% CI 0.3, 1.1).

Analysis of question 3 (includes only those who stated No or don't know/not sure to question 2, n = 659): Would you still say No or Don't know/not sure that women of childbearing age should not be subjects in medical research studies, even if it meant that we would not know anything about the health and medical treatments for women aged 15–45 years?

Less than half of those who stated No or Don't know/not sure to question 2 (39.0%) changed their minds when reminded of the resulting negative consequences for knowledge about women's health that would result from barring women from clinical research (Fig. 4). The odds of changing one's mind did not vary by sex ($OR_{adjusted}$ = 1.4 for women vs. men, 95% CI 1.0, 2.1), although the difference between the sexes overall did approach statistical significance

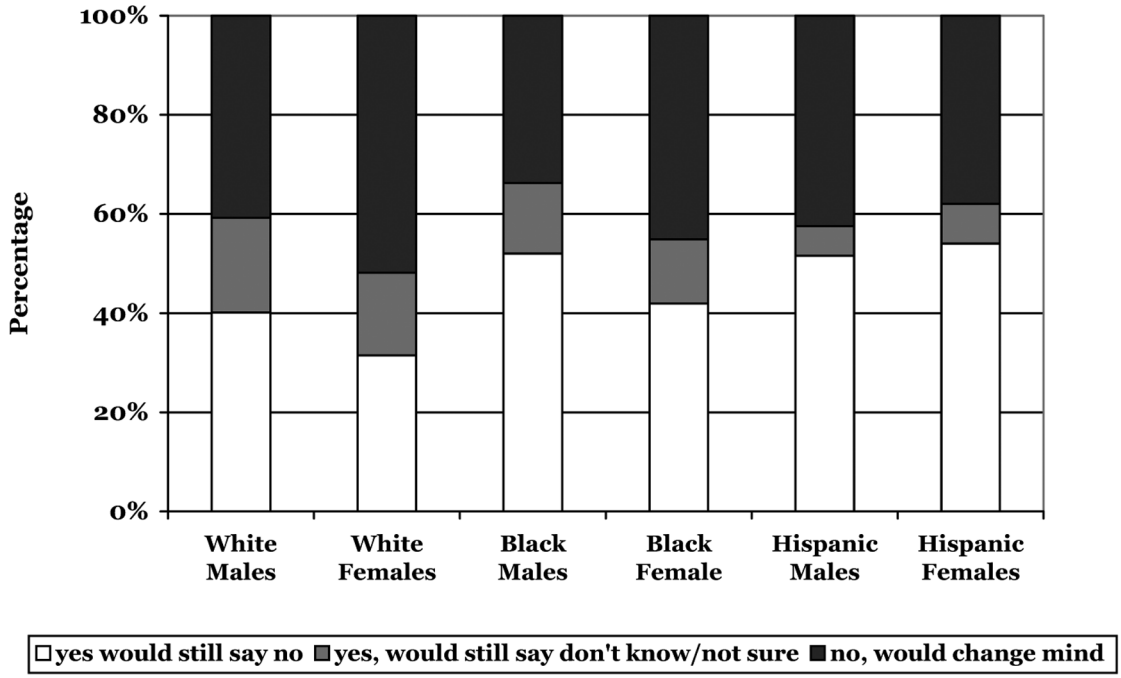


FIG. 4. Response to question 3: Percentage, by race/ethnicity and gender, who would still say No or Don't know/not sure whether women of childbearing age should become research subjects, even if it meant knowing nothing about the health and medical treatments for women aged 14–45 (n = 659).

($p = 0.06$). We found that both black women and Hispanic women were less likely than white women to change their minds when reminded of the negative consequences for women's health (blacks, $OR_{adjusted} = 0.5$, 95% CI 0.3, 1.0; Hispanics, $OR_{adjusted} = 0.5$, 95% CI 0.3, 1.0, both compared to whites), whereas among men, we found no difference in the proportion of those who changed their minds ($OR_{adjusted}$ for blacks 0.6, 95% CI 0.3, 1.1; $OR_{adjusted}$ for Hispanics, 1.0, 95% CI 0.4, 2.8).

Bivariate analysis of knowledge of historical events (knowledge of either thalidomide or DES) and demographic characteristics of participants

Of the total study sample, 12.0% of participants reported having heard of DES, and 22.7% reported having heard of thalidomide. Knowledge of neither thalidomide nor DES was related to sex. Black participants were less likely than white participants to report knowledge of thalidomide (33.4% of whites vs. 10.8% of blacks; chi-square 61.32, *df* 1, $p \leq 0.001$), but there was no difference between blacks and whites regarding knowledge of DES (14.7% of whites vs. 12.1% of blacks; chi-square 1.25, *df* 1, $p = 0.26$). Hispanics were less likely than whites to report knowledge of thalidomide (33.4% of whites vs. 7.7% of Hispanics; chi-square 61.32, *df* 1, $p \leq 0.001$) and of DES (12.1% of whites vs. 6.4% of Hispanics; chi-square 6.98, *df* 1, $p = 0.03$). Knowledge of thalidomide (but not DES) was related to age, with 30.7% of those age 50+ reporting knowledge of thalidomide vs. 14.0% of those age ≤49 (chi-square 43.57, *df* 1, $p \leq 0.001$). Knowledge of both thalidomide and DES was clearly related to both education and to income. Of those with "some college," 44.1% reported knowledge of thalidomide and 23.4% reported knowledge of DES. In contrast, of those with a high school education or less, only 14.7% knew of thalidomide (chi-square 51.23, *df* 1, $p \leq 0.001$) and 7.8% knew of DES (chi-square 110.56, *df* 1, $p \leq 0.001$). Of those with family incomes ≥$35,000 per year, 33.3% reported knowledge of thalidomide and 16.4% reported knowledge of DES vs. 14.4% and 9.2%, respectfully, of those with incomes ≤$34,000 per year (for thalidomide, chi-square 50.25, *df* 1, $p \leq 0.001$; for DES, chi-square 11.79, *df* 1, $p \leq 0.001$). We found geographic differences in knowledge of thalidomide, with more respondents in Hartford (49.8%) than in Birmingham (13.2%), Tuskegee (16.3%), and San Antonio (20.6%) reporting knowledge of thalidomide (chi-square 41.16, *df* 3, $p \leq 0.001$). Geographic location was not related to knowledge of DES (chi-square 2.15, *df* 1, $p = 0.54$).

Analysis of relationship between DES and thalidomide and responses to the WARS domain questions

Results of the adjusted analysis of the relationship between knowledge of DES and thalidomide and responses to the WARS domain questions are shown in Table 2. Overall, knowledge of DES, but not of thalidomide, was independently, albeit moderately, increased in those who believed that women were more likely, compared to men, to be "taken advantage of" when controlling for potential confounders (sex, race/ethnicity, education, income, and geographic location), with those who stated that women, compared to men, are likely Always, Most of the time, or Sometimes to be taken advantage 60% more likely to report knowledge of DES ($OR_{adjusted} = 1.58$, 95% CI 1.02, 1.93) compared to those who responded Never or Rarely.

Knowledge of DES, but not of thalidomide, was also independently related to whether one thought that women of childbearing age should participate in biomedical research as participants (question 2), with those who stated that women of childbearing age should not participate in biomedical research as participants less likely to report knowledge of DES ($OR_{adjusted} = 0.59$, 95% CI 0.41, 0.86). On further examination, we found that the relationship between stating that women of childbearing age should not participate in biomedical research as participants and being less likely to report knowledge of DES was evident only among black participants ($OR_{adjusted} = 0.59$, 95% CI 0.41, 0.86) and among men ($OR_{adjusted} = 0.51$, 95% CI 0.30, 0.87). In fact, we also found that blacks who stated that women of childbearing age should not participate in biomedical research as participants were also less likely to report knowledge of thalidomide ($OR_{adjusted} = 0.36$, 95% CI 0.17, 0.76).

Responses to question 3, whether one changed one's mind about women of childbearing age participating in biomedical research, given the consequences, were not related to knowledge of DES ($OR_{adjusted} = 0.91$, 95% CI 0.52, 1.63) or thalidomide ($OR_{adjusted} = 1.15$, 95% CI 0.77, 1.73) overall. However, this is likely due to the difference in directionality of response seen between men and women. Although the differences within the sexes were not statistically significant (perhaps because of reduced sample size), we found that among men, those who changed their minds from No or Not sure to Yes, women of childbearing age should be research subjects, given the negative consequences for women's health, were about two times more likely to report knowledge of DES and of thalidomide; among women, those who changed their minds were 60% less likely to report knowledge of DES and were 40% less likely to report knowledge of thalidomide.

Discussion

In the late 1970s, largely because of the recognition of the deleterious effects of two medications (thalidomide and DES) taken by thousands of pregnant women on both themselves and their fetuses, the Food and Drug Administration (FDA) prohibited women of childbearing age from participating in clinical research as research subjects.[18] Within 10 years from their expulsion from clinical research, however, it was recognized by the U.S. Public Health Service Task Force on Women's Health that the subsequent fallout from women's exclusion from research was that women's health had been compromised because of lack of knowledge about women's health and treatment of diseases in women.[19] It soon become apparent, for example, that certain diseases (e.g., cardiovascular disease) presented or behaved differently in women than in men and that drug metabolism might vary by sex, for example, with important side effects more prevalent in women than in men.[20] Despite this legislation, several reports have shown that in some cases, especially in the case of racial/ethnic minorities, women still may be underrepresented in biomedical research.[10,21,22]

We examined beliefs and attitudes regarding vulnerability of women when they participate in biomedical research studies. Specifically, we examined whether people believed

Table 2. Reported Knowledge of DES and Thalidomide: Relationship to Questions Exploring Vulnerability of Women as Biomedical Research Participants, TLP[a]

	Reported knowledge of:	*Men*	*Women*	*Whites*	*Blacks*	*Hispanics*	*Total*
Q1: Those who believe that women in the U.S. are more likely to be "taken advantage of" when they become subjects in a medical research project as compared to men always/most times/ sometimes vs. rarely/never	DES	1.60 (0.84, 2.93)	1.43 (0.77, 2.64)	**1.87**[b] **(1.04, 3.36)**	1.55 (0.74, 3.26)	1.73 (0.40, 7.38)	**1.58 (1.02, 1.93)**
	Thalidomide	1.56 (0.84, 2.93)	0.92 (0.56, 1.51)	0.94 (0.64, 1.38)	1.50 (0.62, 3.64)	1.81 (0.47, 7.05)	1.00 (0.70, 1.43)
Q2: Those who believe that women of childbearing age (15–45-year-olds) should not become study participants in medical research projects vs. those who say yes, they should participate	DES	**0.51 (0.30, 0.87)**	0.69 (0.41, 1.14)	0.98 (0.56, 1.70)	**0.30 (0.16, 0.59)**	1.09 (0.30, 4.04)	**0.59 (0.41, 0.86)**
	Thalidomide	0.67 (0.43, 1.06)	0.80 (0.50, 1.28)	0.88 (0.59, 1.31)	**0.36 (0.17, 0.76)**	0.72 (0.22, 2.35)	0.78 (0.57, 1.04)
Q3: Would you still say no, or don't know/not sure, that women of childbearing age should not be subjects in medical research studies *even if it meant that we would not know anything about the health and medical treatments for women aged 15–45 years?*	DES	2.37 (0.96, 5.85)	0.47 (0.21, 1.07)	1.43 (0.70, 2.92)	0.79 (0.14, 4.58)	0.96 (0.14, 6.57)	1.11 (0.65, 1.89)
	Thalidomide	1.91 (0.99, 3.65)	0.61 (0.34, 1.11)	1.45 (0.87, 2.40)	0.55 (0.14, 2.24)	0.29 (0.32, 2.58)	1.03 (0.67, 1.58)

[a]Odds ratios.
[b]Bold indicates statistically significant results at a level of $p \leq 0.05$.

the benefits of including women of childbearing age outweigh the potential risks involved in their participation in these studies and if beliefs varied by sex or by race/ethnicity in a geographically and sociodemographically diverse group of adults. We found that a substantial proportion—roughly half—of participants in this RDD survey thought that women, compared to men, were more likely to be "taken advantage of" as research subjects and that most respondents (67%) were either against or unsure if women of childbearing age should participate as research subjects. When the interviewer informed these respondents about the potential negative consequences of a total lack of knowledge about women's health for women of childbearing age that would inevitably result, fewer than half (39%) of respondents changed their minds about participation of women of childbearing age. Clearly, it is likely that this reluctance on the part of the general public to recognize the importance of inclusion of women in research studies could be one of the barriers in achieving the goals of the NIH mandate[23] and the legal obligation that all federally funded studies include women unless there is a clear reason to exclude them.[24]

Those who stated that women were more likely to be taken advantage of were more likely to be either black or Hispanic. It has been widely reported that African Americans and other minority groups are more reluctant than white people to participate in biomedical research because of a history of research abuses against black people in this country.[4,25,26] Although it is understandable that racial/ethnic minorities might be more suspicious of potential research abuses against women of childbearing age (a group whom many might view as especially vulnerable to exploitation or more likely to receive special protection due to their childbearing potential and the potential dangers to a developing fetus), it remains to be established whether knowledge of research abuses translates into hesitance or refusal to participate in biomedical research. Indeed, several investigations have found no difference between racial/ethnic minorities and white people in willingness to participate in biomedical research,[1,27] and other investigators have found that racial/ethnic differences in biomedical research participation can be accounted for by differences in socioeconomic status.[28,29] When we controlled for differences in socioeconomic status among the three racial/ethnic groups, however, the profound differences we found by race/ethnicity persisted: the odds of believing that women are more likely to be "taken advantage of" when they become subjects in a medical research project were almost 4 times as high for black respondents and 2.5 times as high for Hispanic respondents, compared to whites respondents.

Women were more likely than men to believe that women were more likely to be "taken advantage of" than men as participants in biomedical studies. Given the power imbalance by gender that exists in the United States, it is not surprising that most women would be more suspicious of research abuses on women than would men. We found this attitudinal difference by sex, however, only among white and black people, but not among Hispanics. Despite the fact that Hispanics are now the nation's largest ethnic minority group, comprising 14% of the nation's total population, and the fact that higher rates of certain prevalent diseases, for example, diabetes, are found in Hispanics, data on recruitment of Hispanic women in biomedical research are sparse.[9,10,30] Further research should determine if the NIH Revitalization Act, which, since 1993 has required federally funded studies to include women of childbearing potential unless there was clear justification for exclusion, has caused a substantial change in recruitment of women, especially those who are racial and ethnic minorities.

We found that a substantial proportion of the participants reported knowledge of historical events that have caused *in utero* or subsequent medical harm to the children of pregnant women who took thalidomide or DES during pregnancy and that knowledge of DES, in particular, was more common in those who stated that women were more likely to be "taken advantage of" and to beliefs of whether women of childbearing age should participate in biomedical research studies. Interestingly, those who stated that women are more likely than men to be "taken advantage of" were more likely to report knowledge of DES, although on further analysis, we found that this was the case only among white participants. Those who stated that women of childbearing age should not participate in biomedical research were actually less likely to report knowledge of DES. We found this relationship, however, only in men and only in black respondents, for whom, in particular, knowledge of either historical event was independently, inversely, and strongly related (thalidomide: $OR_{adjusted} = 0.36$, 95% CI 0.17, 0.74; DES: $OR_{adjusted} = 0.32$ 95% CI 0.16, 0.63); that is, those who stated that women of childbearing age should not participate in biomedical research as participants were, as was the case of the overall analysis, less likely to report knowledge of thalidomide or DES. One explanation for the directionality of the relationship between knowledge of DES and stating that women should be participants is that persons who knew of the DES study were aware of details, including the lack of testing performed on DES and the recognition that it was the lack of appropriate testing and use of this drug that caused subsequent harm. This finding supports the view that black people might be more sensitive than white people to issues surrounding clinical research in the United States.

To our knowledge, this is the first large-scale study to document high levels of wariness and differences in attitudes, by sex and by race/ethnicity, regarding women's participation in biomedical research in a large, racially/ethnically diverse group of American adults and to examine if knowledge of past historical events that caused in harm to offspring influences these beliefs. Although level of education has been found to be a predictor of participation in biomedical research,[15] further investigation into the relationship of race/ethnicity (and other demographic factors), beliefs, and attitudes toward biomedical research, and willingness to participate in clinical studies is clearly warranted. Also, studies that examine the relationship among sex, race/ethnicity, and participation should include measures of socioeconomic status in their analysis, as differences in socioeconomic status between the sexes and between the various racial and ethnic groups in the United States tend to be profound. Cultural differences and gender differences (i.e., beliefs and attitudes that vary between men and women as a result of the roles they play in society) should also be addressed.[13] Indeed, studies have found recruitment strategies to be more successful when they are culturally appropriate and directed toward a particular subgroup of the population, for example, high-risk persons.[31]

In 1994, the Institute of Medicine, at the request of the NIH, issued a report on Women in Health Research, in which they explained that their report was focused on "justice," that is, that "women and men should have the opportunity to par-

ticipate equally in the benefits and burdens of research."[32] Although one should bear in mind the principal arguments to support the NIH policy of increased representation of women (first, to increase generalizability and to allow for valid analyses of differences by sex, and, second, because participation in biomedical research is often seen as advantageous to participants), our findings are parallel to those of others, who have suggested that more needs to be understood about whether both men and women, perhaps especially those who are racial/ethnic minorities, find these arguments compelling.[33]

Disclosure Statement

The authors have no conflicts of interest to report.

References

1. Katz RV, Kegeles SS, Kressin NR, et al. The Tuskegee Legacy Project: Willingness of minorities to participate in biomedical research. J Health Care Poor Underserved 2006;17:698–715.
2. Kennedy BM, Burnett MF. Clinical research trials: A comparison of African Americans who have and have not participated. J Cult Diversity 2002;9:95–101.
3. Corbie-Smith G, Thomas SB, Williams MV, Moody-Ayers S. Attitudes and beliefs of African Americans toward participation in medical research. J Gen Intern Med 1999;14:537–546.
4. Shavers-Hornaday VL, Lynch CF, Burmeister LF, Torner JC. Why are African Americans under-represented in medical research studies? Impediments to participation. Ethnicity Health 1997;2:31–45.
5. Gamble VN. Under the shadow of Tuskegee: African Americans and health care. Am J Public Health 1997;87:1773–1778.
6. Harris Y, Gorelick PB, Samuels P, Bempong I. Why African Americans may not be participating in clinical trials. J Natl Med Assoc 1996;88:630–634.
7. Herbst AL. Diethylstilbestrol and adenocarcinoma of the vagina. Am J Obstet Gynecol 1999;81:1576–1578.
8. Knightley P, Evans H, Potter E, Wallace M. Suffer the children: The story of thalidomide. New York: Viking Press, 1979.
9. Geller SE, Goldstein AM, Carnes M. Adherence to federal guidelines for reporting of sex and race/ethnicity in clinical trials. J Womens Health 2006;15:1123–1231.
10. Murthy VH, Krumholz HM, Gross C. Participation in cancer clinical trials. Race-, sex-, and age-based disparities. JAMA 2004;291:2720–2726.
11. Ramasubbu K, Gurm H, Latiker D. Gender bias in clinical trials: Do double standards still apply? J Womens Health Gend Based Med 2001;10:757–764.
12. Harris DJ, Douglas PS. Enrollment of women in cardiovascular clinical trials funded by the National Heart, Lung and Blood Institute. N Engl J Med 2000;343:475–480.
13. BeLue R, Taylor-Richardson KD, Lin J, Rivera AT, Grandison D. African Americans and participation in clinical trials: Differences in beliefs and attitudes by gender. Contemp Clin Trials 2006;27:498–505.
14. Nguyen TT, Somkin CP, Ma Y, Fung LC, Nguyen T. Participation of Asian-American women in cancer treatment research: A pilot study. J Natl Cancer Inst 2005;35:102–105.
15. Brown BA, Long HL, Milliken N. What's to know about study recruitment? We asked recruiters. Women Health Issues 2002;12:116–121.
16. Rimer BK, Schildkraut JM, Lerman C, Lin TH, Audrain J. Participation in a women's breast cancer risk counseling trial. Who participates? Who declines? High Risk Breast Cancer Consortium. Cancer 1996;77:2348–2355.
17. Katz RV, Kegeles SS, Green LB, et al. The Tuskegee Legacy Project: History, preliminary scientific findings, and unanticipated societal benefits. Dental Clin North Am 1993;47: 1–19.
18. Food and Drug Administration. General considerations for the clinical evaluation of drugs. Publication No. HEW (FDA) 77-3040, 1977.
19. United States Public Health Service. Report of the Public Service Task Force on Women's Health Issues. Public Health Reports 1985;100:73–106.
20. Ebert SN, Liu XK, Woosley RL. Female gender as a risk factor for drug-induced cardiac arrythmias: Evaluation of clinical and experimental evidence. J Womens Health 1998;7: 547–557.
21. Simon V. Wanted: Women in clinical trials. Science 2005;308: 1517.
22. Heiat A, Gross CP, Krumholz HM. Representation of the elderly, women, and minorities in heart failure clinical trials. Arch Intern Med 2002;162:1682–1688.
23. NIH Guide for Grants and Contracts, 1990;19:Nos. 31, 35.
24. Federal Register. NIH Guidelines on the inclusion of women and minorities as subjects in clinical research, 1994;59:14508.
25. Kressin NR, Meterko M, Wilson NJ. Participation in biomedical research: Are there racial disparities within the VA? J Natl Med Assoc 2000;92:62–69.
26. Thomas SB, Quinn, SC. Public Health then and now, The Tuskegee Syphilis Study, 1932 to 1972: Implications for HIV education and AIDS risk education programs in the black community. Am J Public Health 1991;81:1498–1505.
27. Wendler D, Kington R, Madans J, et al. Are racial and ethnic minorities less willing to participate in health research? PLoS Med 2006;3:e19, Epub 2005 Dec. 6.
28. Gross CP, Filardo G, Mayne ST, Krumholz HM. The impact of socioeconomic status and race on trial participation for older women with breast cancer. Cancer 2005;103:483–491.
29. Unson CG, Ohannessian C, Kenyon L, Case A, Reisine S, Prestwood K. Barriers to eligibility and enrollment among older women in a clinical trial on osteoporosis: Effects of ethnicity and SES. J Aging Health 2004;16:426–443.
30. Crane PB, Letvak S, Lewallen L, Hu J, Jones E. Inclusion of women in nursing research: 1995–2001. Nurs Res 2004;53: 237–242.
31. Wilson KM, Orians CE. Considerations in recruiting underscreened women to focus groups for screening for cervical cancer. Health Promotion Pract 2005;6:379–384.
32. Fadin R. Introduction. In: Mastroianni AD, Faden R, Federman D, eds. Women and health research: Ethical and legal issues of including women in clinical studies. Institute of Medicine: National Academies Press, Washington, DC, 1994:Vol 1.
33. Garber M, Arnold RM. Promoting the participation of minorities in research. Am J Bioeth 2006;6:W14–20.

Address reprint requests to:
Dr. Stefanie Luise Russell
New York University College of Dentistry
Epidemiology and Health Promotion
345 East 24th Street
New York, New York 10010

E-mail: slr6@nyu.edu

Some Medical Inpatients With Unhealthy Alcohol Use May Benefit From Brief Intervention*

RICHARD SAITZ, M.D., M.P.H.,† TIBOR P. PALFAI, PH.D.,† DEBBIE M. CHENG, SC.D.,† NICHOLAS J. HORTON, SC.D.,† KIM DUKES, PH.D.,† KEVIN L. KRAEMER, M.D., M.SC.,† MARK S. ROBERTS, M.D., M.P.P.,† ROSANNE T. GUERRIERO, M.P.H.,† AND JEFFREY H. SAMET, M.D., M.A., M.P.H.†

Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, and Youth Alcohol Prevention Center and Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts

ABSTRACT. Objective: Studies of alcohol brief intervention for medical inpatients have mixed results. We explored potential moderators of the effectiveness of brief intervention for unhealthy alcohol use among medical inpatients. **Method:** This is a secondary analysis of a randomized controlled trial of brief motivational counseling among 341 urban-hospital medical inpatients (99 women) with unhealthy alcohol use. Self-reported main outcomes were receipt of alcohol treatment by 3 months in subjects with dependence and change in the mean number of drinks per day 3 and 12 months after enrollment in all subjects. **Results:** Among subjects with dependence, the effect of brief intervention on receipt of alcohol treatment differed significantly by gender and age (p = .02 for each interaction). In stratified analyses, brief intervention was associated with receipt of alcohol treatment in women (adjusted odds ratio [AOR] = 3.9, 95% confidence interval [CI]: 1.2-12.7), and younger (<44 years) subjects (AOR = 3.6, 95% CI: 1.3-10.1). Among subjects with nondependent, unhealthy alcohol use, brief intervention was significantly associated with fewer drinks per day and better physical health-related quality of life at 3 months. However, among those with dependence, intervention was associated with worse physical health-related quality of life and more hospital use, and no changes in drinking. In adjusted analyses among those with and without dependence, brief intervention was not associated with mental health-related quality of life, alcohol problems, or readiness to change. Effects of brief intervention on consumption outcomes were not consistently moderated by demographic characteristics, comorbidity/health, or readiness to change. **Conclusions:** Some medical inpatients with unhealthy alcohol use, particularly women, younger adults, and patients without dependence may benefit from brief intervention. Few factors that were expected to moderate brief intervention effects did so. Additional research should assess which medical inpatients, if any, can benefit from brief intervention. (*J. Stud. Alcohol Drugs* **70:** 426-435, 2009)

PROFESSIONAL ORGANIZATIONS RECOMMEND screening and brief intervention for all adults with unhealthy alcohol use (i.e., the spectrum from drinking risky amounts through dependence; Institute of Medicine, 1990; U.S. Preventive Services Task Force, 2004). Brief intervention, however, has proven efficacy in decreasing alcohol consumption and related consequences only in outpatients with unhealthy, but not dependent, alcohol use (Wilk et al., 1997). Further, the results from studies of the efficacy of brief intervention among other populations, such as hospitalized patients, are unclear and sometimes negative, as reported in several large randomized trials (Emmen et al., 2004; Freyer-Adam et al., 2008; Saitz et al., 2007).

Many factors may moderate the efficacy of brief intervention. Younger adult women and patients with an alcohol-attributable diagnosis (e.g., alcoholic hepatitis) might benefit more than others from intervention (Blow et al., 2006; Weisner et al., 2001). Race and ethnicity may affect receipt of alcohol-treatment services (Schmidt et al., 2007), and cognitive impairment may lower adherence to alcohol treatment (Bates et al., 2006). In an emergency department brief intervention, stage of change and self-efficacy did not appear to moderate brief intervention effects on consumption, whereas the patient's attribution of injury to alcohol did moderate these effects (Walton et al., 2008). Similarly and perhaps surprisingly, in another study (with a pre-/post-

Received: January 24, 2008. Revision: September 29, 2008.

*This study was supported by National Institute on Alcohol Abuse and Alcoholism grant RO1 12617 and National Center for Research Resources, General Clinical Research Center grant M01 RR00533.

†Correspondence may be sent to Richard Saitz, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118 or via email at: rsaitz@bu.edu. Tibor Palfai is with the Youth Alcohol Prevention Center, Boston University School of Public Health and the Department of Psychology, College of Arts and Sciences, Boston University, Boston, MA. Debbie Cheng is with the Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine and the Department of Biostatistics, Boston University School of Public Health, Boston, MA. Nicholas Horton is with the Department of Mathematics and Statistics, Smith College, Northampton, MA. Kim Dukes is with DM-STAT, Inc., Malden, MA. Kevin Kraemer is with the Division of General Internal Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA. Mark Roberts is with the Division of General Internal Medicine, Section of Decision Sciences and Clinical Systems Modeling, University of Pittsburgh School of Medicine, Pittsburgh, PA. Rosanne Guerriero is with the Lahey Clinic, Burlington, MA. Jeffrey Samet is with the Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine and the Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA.

design) the presence or absence of dependence did not appear to affect changes in drinking after brief intervention (Guth et al., 2008). However, many hospital studies that support the use of brief intervention exclude patients with characteristics that may decrease the intervention's effectiveness (e.g., psychiatric comorbidity, other drug use), despite the fact that these patients represent the population identified by alcohol screening (Chick et al., 1985; Heather et al., 1996; McManus et al., 2003). In the few inpatient studies on brief intervention that have included the broad spectrum of patients with unhealthy alcohol use, results generally have been negative (Saitz et al., 2007; Watson, 1999). Therefore, it is likely that brief intervention has efficacy only in certain people and settings.

Despite this limited efficacy, large and well-funded federal efforts are underway to implement brief intervention for all patients with unhealthy alcohol (and drug) use (Center for Substance Abuse Treatment, 2007). In addition, few studies have examined moderators of brief intervention efficacy, and none (to our knowledge) have done so in hospitalized patients. Therefore, to help clarify the effects of brief intervention among medical inpatients and to study moderators of these effects, we explored data collected as part of a randomized controlled trial involving a broad spectrum of patients with unhealthy alcohol use that did not support the efficacy of brief intervention in the study group overall. We examined whether demographic factors, alcohol use severity (dependence), health/comorbidity, and readiness to change moderated the intervention's effects on receipt of treatment, alcohol consumption, alcohol problems, readiness to change, health-related quality of life, and health care use.

Method

Subjects

As described previously (Saitz et al., 2007), we enrolled 341 adult subjects (99 women) from the medicine service of a large, urban teaching hospital. Eligibility criteria included current (past-month) drinking of risky amounts (defined for eligibility as >14 standard drinks per week or ≥5 drinks per occasion for men; >11 drinks per week or ≥4 drinks per occasion for women and people age ≥ 66 years); 2 contacts to assist with follow-up; no plans of moving from the area in the next year; and a Mini-Mental State Examination score of 21 or more (Folstein et al., 1975; Smith et al., 2006). Eligible subjects who enrolled in the clinical trial provided written informed consent. The Institutional Review Board of the Boston University Medical Center approved this study.

Assessments

Research associates interviewed subjects before randomization to assess the characteristics listed in Table 1. One coauthor (R.S.) reviewed medical records to determine current primary and alcohol-attributable medical diagnoses (Adams et al., 1993). At 3 and 12 months, research associates reassessed, via interview, most domains covered at enrollment.

Randomization and intervention

Subjects were randomized to the control or intervention group. Control subjects received usual care (i.e., they were told their screening results and advised they could discuss their alcohol use with their physicians). Intervention subjects were assigned to 30 minutes of brief motivational counseling that was based on the principles of motivational interviewing (Miller and Rollnick, 1991, 2002; Miller et al., 1995a). Sessions were conducted by counseling and clinical psychology doctoral students whom we trained and included feedback, an open discussion (lasting about 20 minutes), and construction of a change plan (Saitz et al., 2007).

Outcomes and measurements

The primary outcomes in this study were self-reported receipt of alcohol treatment in the past 3 months among subjects with alcohol dependence and change in the mean number of drinks per day from enrollment to 3 and 12 months in subjects with and without dependence. We measured receipt of treatment with a standardized interview based on the Treatment Services Review (McLellan et al., 1992a) and Form 90 (Miller, 1996). Treatment included residential treatment, outpatient treatment (e.g., counseling or therapy), medications, employee-assistance programs, or mutual-help groups (e.g., Alcoholics Anonymous). We measured past-30-day consumption with the Timeline Followback method (Sobell and Sobell, 1992). From this, we determined mean drinks per day, days abstinent, and heavy drinking episodes (≥5 drinks per occasion for men and ≥4 for women and people age ≥ 66 years). We also assessed the proportions of subjects who abstained for all 30 days, had at least one heavy drinking episode, and drank risky amounts (>14 drinks per week or ≥5 drinks per occasion for men; >7 drinks per week or ≥4 drinks per occasion for women and people age ≥ 66 years). Additional secondary outcomes included alcohol problems (total score on the Short Inventory of Problems; Miller et al., 1995b), readiness to change (Taking Steps scale of the Stages of Change Readiness and Treatment Eagerness Scale [SOCRATES]; Miller and Tonigan, 1996), physical and mental health-related quality of life (physical and mental component summary scale scores on the Short-Form Health Survey [SF-12]; Ware et al., 1998), and emergency-department visits and days of medical hospitalization (both determined by a standardized interview based on the Treatment Services Review and Form 90; McLellan et al., 1992a; Miller, 1996).

TABLE 1. Characteristics at enrollment of subjects with unhealthy alcohol use, by dependence status and randomized group (n = 341)

Variable	Subjects with nondependent, unhealthy alcohol use		Subjects with dependent, unhealthy alcohol use	
	Control (n = 40)	Interv. (n = 40)	Control (n = 129)	Interv. (n = 132)
Demographics				
Women, no. (%)	14 (35%)	9 (23%)	45 (35%)	31 (23%)
Age, mean (SD)	45 (13)	46 (13)	44 (10)	44 (10)
Race/ethnicity				
Black, no. (%)	16 (40%)	15 (38%)	64 (50%)	60 (45%)
White, no. (%)	20 (50%)	18 (45%)	46 (36%)	49 (37%)
Hispanic, no. (%)	2 (5%)	4 (10%)	11 (9%)	13 (10%)
Unemployed, past 3 months, no. (%)	21 (53%)	21 (53%)	83 (64%)	91 (69%)
Homeless, ≥1 night, past 3 months, no. (%)	5 (13%)	3 (8%)	34 (26%)	44 (33%)
Medical diagnoses				
Principal diagnosis, most common at current admission[a]				
Rule out myocardial infarction, no. (%)	4 (10%)	9 (23%)	26 (20%)	22 (17%)
Asthma, bronchitis, and COPD, no. (%)	11 (28%)	3 (8%)	10 (8%)	12 (9%)
Pancreatitis, no. (%)	0 (0%)	1 (3%)	13 (10%)	19 (14%)
Cellulitis, no. (%)	5 (13%)	2 (5%)	9 (7%)	6 (5%)
Diabetes, no. (%)	0 (0%)	4 (10%)	5 (4%)	5 (4%)
Alcohol-attributable diagnosis,[b] no. (%)	3 (8%)	3 (8%)	17 (13%)	28 (21%)
Any alcohol-attributable diagnosis,[b] current admission), no. (%)	9 (23%)	10 (25%)	57 (44%)	80 (61%)
Comorbidity[c] lifetime, median score (Q1-Q3)	1 (0-1)	1 (0-2)	1 (0-2)	1 (0-2)
DSM-IV Alcohol Diagnoses,[a] past year				
Alcohol abuse, no. (%)	8 (20%)	7 (18%)	–	–
Alcohol dependence, no. (%)	–	–	129 (100%)	132 (100%)
No alcohol diagnosis, no. (%)	32 (80%)	33 (83%)	–	–
Alcohol consumption,[a] past 30 days				
Drinks/day, median (Q1-Q3)	1.5 (0.8-2)	1.4 (0.7-2)	5 (2-12)	5 (2-10)
Drinks/drinking day, median (Q1-Q3)	6 (4-10)	5 (3-6)	9 (6-16)	12 (7-16)
Maximum no. of drinks/occasion, median (Q1-Q3)	10 (6-13)	8 (6-12)	17 (12-24)	18 (12-24)
Alcohol-related characteristics				
Readiness to change, taking steps,[a] median score (Q1-Q3)	24 (16-30)	24 (16-28)	31 (26-34)	31 (27-35)
Family history of alcoholism,[d] no. (%)	31 (82%)	29 (73%)	105 (83%)	119 (93%)
Alcohol problems,[a] past 3 months, median score (Q1-Q3)	1 (0-3)	1 (0-4)	16 (9-27)	21 (9-34)
Drug use, past 30 days				
Cigarettes,[e] no. (%)	24 (60%)	26 (65%)	105 (81%)	102 (77%)
Heroin/cocaine use,[f] no. (%)	8 (20%)	4 (10%)	43 (33%)	33 (25%)
Any drug use,[g] no. (%)	20 (50%)	14 (35%)	86 (67%)	75 (57%)
Psychiatric/violence history				
Panic disorder,[h] (current), no. (%)	1 (3%)	0 (0%)	24 (19%)	31 (23%)
Generalized anxiety disorder,[h] current, no. (%)	17 (43%)	17 (43%)	109 (85%)	104 (79%)
Substantial depressive symptoms,[i] current, no. (%)	20 (50%)	12 (30%)	101 (79%)	110 (83%)
Substantial PTSD symptoms,[j] current, no. (%)	7 (18%)	3 (8%)	54 (42%)	75 (57%)
Victim of interpersonal violence (e.g., physical, sexual),[k] lifetime, no. (%)	22 (55%)	19 (48%)	102 (79%)	96 (73%)
Health-related quality of life (HRQL)[a]				
Physical HRQL, mean score (SD)	38 (10)	40 (10)	38 (9)	38 (9)
Mental HRQL, mean score (SD)	46 (12)	51 (10)	38 (11)	37 (13)
Health care use, past 3 months				
Alcohol treatment,[a] no. (%)	1 (3%)	0 (0%)	33 (26%)	52 (40%)
Expanded alcohol treatment,[l] no. (%)	3 (8%)	0 (0%)	42 (33%)	60 (46%)
Any psychiatric treatment, no. (%)	5 (13%)	2 (5%)	34 (26%)	42 (32%)
Psychiatric hospitalization, no. (%)	0 (0%)	0 (0%)	6 (5%)	6 (5%)
Medical hospitalization, no. (%)	10 (25%)	4 (10%)	39 (30%)	45 (34%)
Days hospitalized, median, Q1-Q3	0 (0-5)	0 (0-0)	0 (0-2)	0 (0-2)
Emergency-department use, no. (%)	14 (35%)	9 (23%)	65 (50%)	67 (51%)
Emergency-department visits, median (Q1-Q3)	0 (0-1)	0 (0-0)	1 (0-2)	1 (0-2)

Notes: Interv. = intervention; COPD = chronic obstructive pulmonary disease; Q1 = quartile 1 (or 25th percentile); Q3 = quartile 3 (or 75th percentile); DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; PTSD = posttraumatic stress disorder. [a]See the Method section for a description of how this characteristic was measured; [b]includes any of the following: acute alcoholic cirrhosis of the liver, alcoholic cardiomyopathy, alcoholic gastritis, alcoholic hepatitis, alcohol intoxication, alcoholic liver damage, alcoholic fatty liver, alcoholic pellagra, alcoholic polyneuropathy, alcohol withdrawal, alcohol withdrawal

TABLE 1 *Notes Continued on following page*

Statistical analyses

We conducted all analyses based on the intention-to-treat principle. Reported *p* values are two-tailed, and a *p* value less than .05 was considered statistically significant. We analyzed data with SAS/STAT software Versions 8.2 and 9.1.3 (SAS Institute Inc., Cary, NC). To describe the study sample and compare groups, we used the chi-square test, Fisher's exact test, two-sample *t* test, and Wilcoxon rank sum test, as appropriate. We analyzed dichotomous outcomes using the chi-square test and logistic regression models, continuous outcomes using the two-sample *t* test and linear regression models, and counts (i.e., drinking and health care use) using Poisson models that accounted for overdispersion. We planned, a priori, to assess possible moderators of the intervention, including demographic factors (gender, race, age [<44 and ≥44 years], homelessness [≥1 night in the past 3 months]), comorbidity/health (mental health-related quality of life [<45 and ≥45 on the mental component summary scale of the SF-12], any alcohol-attributable medical diagnosis at current admission, any past-30-day heroin or cocaine use), and readiness to change (<30 or ≥30 on the SOCRATES Taking Steps scale). Post hoc, we decided to include cognitive functioning (<27 and ≥27 on the Mini-Mental State Examination) among the comorbidity/health factors. Within dependence strata, we tested for these possible moderators. When moderators were significant, we reported stratified analyses. Regression analyses stratified only by dependence controlled for clinically important baseline imbalances between the treatment arms; analyses with further stratifications adjusted for the baseline value of the outcome only.

Results

Screening and enrollment data have been described elsewhere (Saitz et al., 2006, 2007). In summary, 341 patients of 986 who screened positive for drinking risky amounts enrolled in the randomized trial (Figure 1). Of enrolled subjects, 172 were randomized to the intervention group and 169 to the control group. Over 12 months, 11 subjects died; 90% (308) of all enrolled subjects completed at least one follow-up interview. Subjects who completed any follow-up were generally similar to those lost to follow-up.

Subjects with nondependent, unhealthy alcohol use in the control group were similar to those in the intervention group with the exception of mental health-related quality of life, which was significantly higher in the intervention group (Table 1). Among subjects with dependence, those in the control group were significantly more likely than those in the intervention group to be women; controls were less likely to have received alcohol treatment, detoxification, or halfway house services in the past 3 months and to have any alcohol-attributable diagnosis, a family history of alcoholism, and substantial depressive symptoms.

Alcohol treatment (subjects with dependence only)

In the main study findings previously reported by Saitz et al. (2007), brief intervention was associated with receipt of alcohol treatment, but the association was not significant and it was attenuated in adjusted analyses (Table 2). In analyses adjusted for potential confounders and simultaneously for interactions that were significant ($p \leq .10$) in unadjusted analyses, interactions between the intervention and gender and age remained significant (both $p = .02$). In subsequent stratified analyses, brief intervention was associated with increased receipt of alcohol treatment by women and younger men (<44 years) in unadjusted models (Table 2). In adjusted analyses, the results remained statistically significant for women but not for younger men. Of note, older men with dependence were significantly more likely than younger men with dependence to be white (46% vs 30%), live alone (47% vs 27%), be unemployed (69% vs 53%), and have worse physical health-related quality of life (mean SF-12 physical component summary score, 36 vs 40) but were significantly less likely to have substantial symptoms of posttraumatic stress disorder (39% vs 58%).

Because women, regardless of age, appeared to benefit most from the brief intervention, we explored other possible gender effects. We detected an interaction between the intervention and alcohol-attributable medical diagnosis among women ($p = .006$) but not men. Of women with any alcohol-attributable medical diagnosis, those in the intervention group were more likely than those in the control group to receive treatment (Table 2). Among women with higher cognitive function, receipt of treatment was more likely among subjects in the intervention group than in the control

TABLE 1 *Notes Continued*
convulsion, alcohol withdrawal delirium, alcohol withdrawal hallucinosis, other alcoholic psychosis, alcoholic amnestic syndrome, other alcoholic dementia, alcoholic pancreatitis, or other diagnoses thought to be alcohol-attributable by the investigator (e.g., holiday heart, alcoholic ketoacidosis, alcohol-related rhabdomyolysis) (Adams et al., 1993); [c]determined by a validated questionnaire (Katz et al., 1996); [d]determined by the Family History-Research Diagnostic Criteria (Andreasen et al., 1977); $n = 38$ for controls with nondependent, unhealthy alcohol use; [e]based on a response of "yes, every day in the past 30 days" to the question: "Do you currently smoke?" (Patrick et al., 1994); [f]determined by the Addiction Severity Index (McLellan et al., 1992b); [g]determined by the Addiction Severity Index and includes use of heroin, methadone, other opiates/analgesics, barbiturates, sedatives/hypnotics/tranquilizers, cocaine, amphetamines, marijuana/cannabis, or hallucinogens; [h]determined by the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (Spitzer et al., 1994); [i]≥16 on the Center for Epidemiologic Studies Depression scale (Boyd et al., 1982; Radloff, 1977); [j]≥44 on the Post Traumatic Stress Disorder Checklist (Blanchard et al., 1996); [k]determined by adapted items from the Traumatic Life Events Questionnaire-Revised (Kubany et al., 2000); [l]includes alcohol treatment, except for medications, plus hospitalization for detoxification (any type); participation in any detoxification program; or halfway house services.

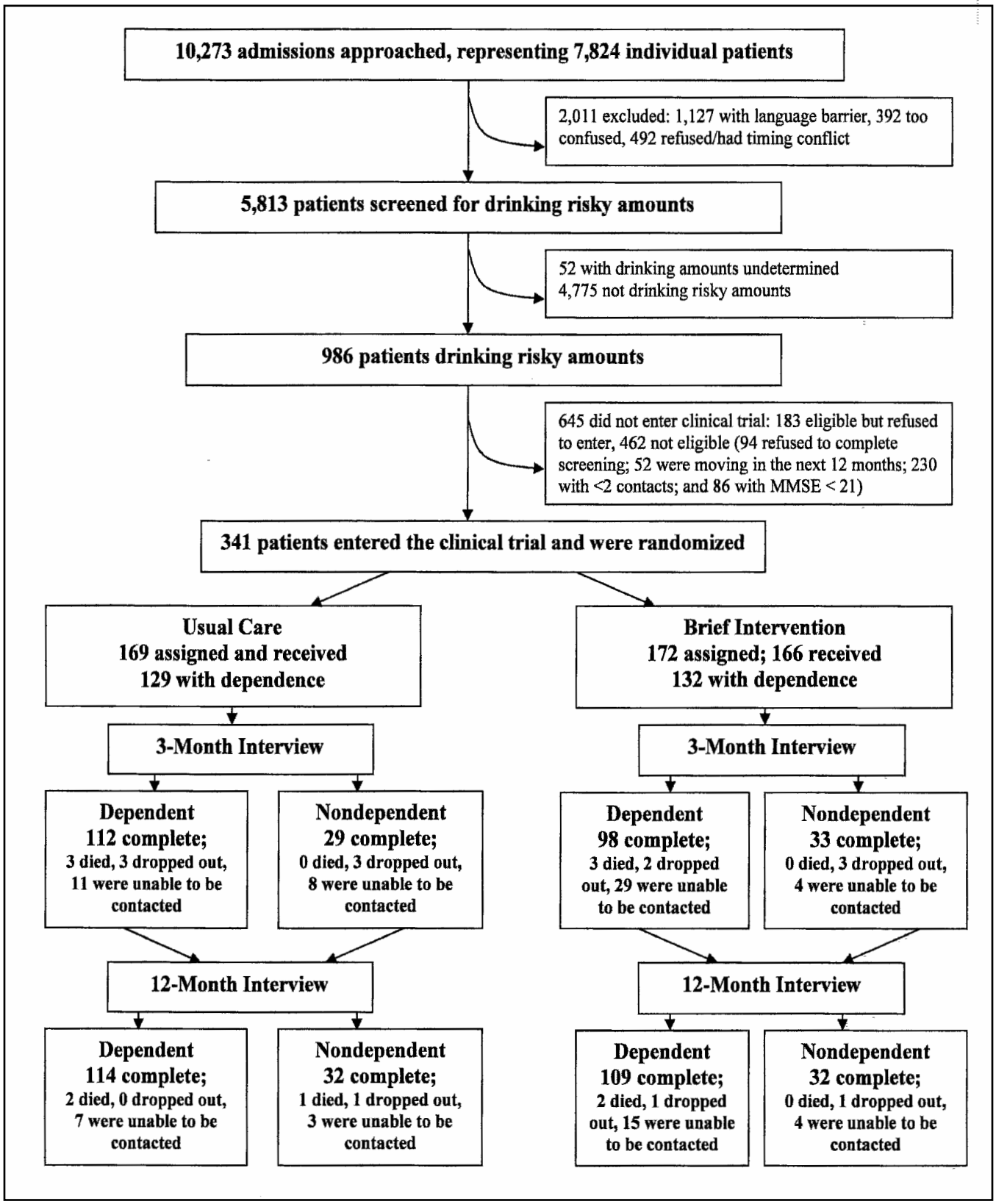


FIGURE 1. Screening and enrollment. Dependence/dependent refers to a diagnosis of alcohol dependence. MMSE = Mini-Mental State Examination. Subjects who dropped out at 3 months were permanently lost to follow-up. Subjects who could not be contacted at 3 months may have been contacted at 12 months. Analyses for treatment at 3 months included only subjects with dependence. All other analyses included all randomized subjects with available data.

Table 2. Alcohol treatment by 3 months in subjects with alcohol dependence

Variable	Numbers and proportions		OR (95% CI) Intervention vs control	Adjusted ORs (95% CI) Intervention vs control
	Control	Intervention		
Overall (n = 209)[a]	44/112 (39%)	50/97 (52%)	1.6 (0.9-2.8)	1.2 (0.6-2.5)[d]
Stratified by gender (interaction p = .02)*				
Men (n = 145)	29/71 (41%)	34/74 (46%)	1.2 (0.6-2.4)	0.71 (0.32-1.6)
Women (n = 64)	15/41 (37%)	16/23 (70%)	4.0 (1.3-11.8)	3.9 (1.2-12.7)
Stratified by age (interaction p = .02)*				
Age < 44 (n = 93)	18/51 (35%)	27/42 (64%)	3.3 (1.4-7.7)	3.6 (1.3-10.1)
Age ≥ 44 (n = 116)	26/61 (43%)	23/55 (42%)	1.0 (0.5-2.0)	0.56 (0.23-1.4)
Stratified by gender and age (interaction p = .02 for men, p = .86 for women)				
Men age < 44 (n = 58)	9/27 (33%)	19/31 (61%)	3.2 (1.1-9.3)	3.0 (0.74-12.2)
Men age ≥ 44 (n = 87)	20/44 (45%)	15/43 (35%)	0.6 (0.3-1.5)	0.26 (0.08-0.86)
Women age < 44 (n = 35)	9/24 (38%)	8/11 (73%)	4.4 (0.9-21.2)	4.7 (0.84-26.7)
Women age ≥ 44 (n = 29)	6/17 (35%)	8/12 (67%)	3.7 (0.8-17.4)	3.2 (0.60-17.1)
Stratified by gender and alcohol-attributable medical diagnosis[c] (interaction p = .97 for men, p = .006 for women)				
Men with a diagnosis (n = 82)	16/33 (48%)	25/49 (51%)	1.1 (0.5-2.7)	0.70 (0.25-2.0)
Men without a diagnosis (n = 63)	13/38 (34%)	9/25 (36%)	1.1 (0.4-3.1)	0.48 (0.12-2.0)
Women with a diagnosis (n = 29)	4/15 (27%)	13/14 (93%)	35.7 (3.5-368.8)	63.5 (3.7-1083.5)
Women without a diagnosis (n = 35)	11/26 (42%)	3/9 (33%)	0.7 (0.1-3.3)	0.55 (0.09-3.2)
Stratified by gender and cognitive functioning (interaction p = .87 for men, p = .21 for women)				
Men with high cognitive functioning[b] (n = 76)	16/40 (40%)	18/36 (50%)	1.5 (0.6-3.7)	1.1 (0.40-3.1)
Men with low cognitive functioning (n = 69)	13/31 (42%)	16/38 (42%)	1.0 (0.4-2.6)	0.35 (0.08-1.4)
Women with high cognitive functioning (n = 32)	6/21 (29%)	8/11 (73%)	6.7 (1.3-34.0)	7.8 (1.3-46.3)
Women with low cognitive functioning (n = 32)	9/20 (45%)	8/12 (67%)	2.4 (0.6-10.8)	1.9 (0.36-10.1)

Notes: OR = odds ratio; CI = confidence interval. *p value from analysis adjusting simultaneously for interactions between the intervention and age, gender, cognitive functioning, and alcohol-attributable medical diagnosis (interactions identified in initial unadjusted analyses). [a]210 subjects with alcohol dependence were interviewed at 3 months; however, 1 did not answer questions about alcohol treatment; [b]high cognitive functioning is ≥27 on the Mini-Mental State Examination; [c]includes any of the following: acute alcoholic cirrhosis of the liver, alcoholic cardiomyopathy, alcoholic gastritis, alcoholic hepatitis, alcohol intoxication, alcoholic liver damage, alcoholic fatty liver, alcoholic pellagra, alcoholic polyneuropathy, alcohol withdrawal, alcohol withdrawal convulsion, alcohol withdrawal delirium, alcohol withdrawal hallucinosis, other alcoholic psychosis, alcoholic amnestic syndrome, other alcoholic dementia, alcoholic pancreatitis, or other diagnoses thought to be alcohol-attributable by the investigator (e.g., holiday heart, alcoholic ketoacidosis, alcohol-related rhabdomyolysis) (Adams et al., 1993); [d]overall analysis adjusted for gender, alcohol treatment in the 3 months before enrollment, family history of alcoholism, any drug use, alcohol problem score, and alcohol-attributable medical diagnoses; remainder of adjusted odds ratios in the table adjusted for alcohol treatment in the 3 months before enrollment only.

group. We found no significant interactions between the intervention and race, mental health-related quality of life, homelessness, heroin or cocaine use, or readiness to change among women or men. Findings of the analyses adjusted for prior alcohol treatment were similar (Table 2).

Alcohol consumption

At 3 and 12 months in both adjusted and unadjusted analyses, across the six outcome measures (Table 3) for those with and without dependence, only one comparison was significant: In adjusted analyses, among subjects with nondependent, unhealthy alcohol use at 3 months, the intervention group drank less than controls (adjusted means 1.5 vs 3.8 drinks per day, incidence rate ratio [IRR] = 0.38, p = .02). At 12 months among subjects with nondependent, unhealthy alcohol use, 19% were abstinent, 67% had at least one heavy drinking episode, and 67% were drinking risky amounts. For subjects with dependence, the proportions were 31%, 61%, and 61%, respectively.

TABLE 3. Alcohol-consumption outcomes at 3 and 12 months in subjects with unhealthy alcohol use

	3 months[a]						12 months[f]					
	Without dependence			With dependence			Without dependence			With dependence		
Consumption measures	Control	Interv.	*p*	Control	Interv.	*p*	Control	Interv.	*p*	Control	Interv.	*p*
Past 30 days, mean												
No. drinks per day	3.63	1.72	.16	5.33	5.26	.95	2.28	2.10	.81	4.97	5.72	.49
No. drinks per day, adj.[b]	3.84	1.46	.02*	3.97	4.12	.87	2.04	1.35	.22	3.29	4.14	.25
No. heavy drinking episodes[c]	6.66	5.18	.47	10.07	10.30	.88	6.00	5.63	.85	10.25	10.34	.95
No. heavy drinking episodes, adj.	4.91	4.06	.52	7.68	9.24	.21	4.75	4.39	.81	7.30	8.13	.47
No. days abstinent	19.00	19.91	.72	16.91	17.40	.76	19.72	19.75	.99	17.47	17.19	.86
No. days abstinent, adj.	19.24	20.15	.69	17.35	15.57	.27	19.87	19.78	.97	18.40	16.45	.24
Past 30 days, with the control group as the reference, OR (95% CI)												
Drinking risky amounts[d]	0.9 (0.3-2.6)			0.9 (0.5-1.5)			0.9 (0.3-2.5)			0.9 (0.5-1.6)		
Drinking risky amounts, adj.[e]	0.9 (0.2-3.6)			1.1 (0.6-2.1)			1.1 (0.3-3.7)			1.2 (0.6-2.1)		
Heavy drinking episodes	0.8 (0.3-2.3)			0.8 (0.5-1.5)			0.9 (0.3-2.5)			1.0 (0.6-1.7)		
Heavy drinking episodes, adj.	0.7 (0.2-2.7)			1.0 (0.5-1.9)			1.1 (0.3-3.7)			1.2 (0.7-2.2)		
Abstinence	1.1 (0.3-3.9)			1.1 (0.6-2.2)			1.0 (0.3-3.5)			1.2 (0.7-2.0)		
Abstinence, adj.	0.7 (0.1-3.6)			0.9 (0.4-1.8)			0.8 (0.2-3.3)			1.0 (0.5-1.8)		

Notes: Interv. = intervention; adj. = adjusted; OR = odds ratio; CI = confidence interval. *Incidence rate ratio (for intervention effect) = 0.38. [a]At 3 months among subjects without dependence, 62 were included in the unadjusted analyses, and 60 were included in the adjusted analyses; at 3 months among subjects with dependence, 209 were included in the unadjusted analyses and 204 were included in the adjusted analyses; [b]adjusted for gender, alcohol treatment in the 3 months before enrollment, family history of alcoholism, any drug use, alcohol problem score, and alcohol-attributable medical diagnoses; [c]≥5 drinks per occasion for men or ≥4 drinks per occasion for women and people age ≥ 66 years; [d]>14 standard drinks per week or ≥5 drinks per occasion for men; >7 drinks per week or ≥4 drinks per occasion for women and people age ≥ 66 years; [e]adjusted for mean drinks per day at enrollment, gender, alcohol treatment in the 3 months before enrollment, family history of alcoholism, any drug use, alcohol problem score, and alcohol-attributable medical diagnoses; [f]at 12 months among subjects without dependence, 64 were included in the unadjusted analyses, and 63 were included in the adjusted analyses; at 12 months among subjects with dependence, 223 were included in the unadjusted analyses, and 217 were included in the adjusted analyses.

To identify possible reasons why we failed to find the hypothesized intervention effects on consumption outcomes, we examined 12-month data and found no significant interactions between the intervention and possible moderators (demographics, comorbidity/health, or readiness to change) on abstinent days or heavy episodic drinking. For drinks per day, we identified only two interactions, age ($p = .01$) and heroin or cocaine use ($p = .03$), among subjects with nondependent, unhealthy alcohol use. Among younger (<44 years) but not older subjects with nondependent, unhealthy alcohol use, the intervention decreased drinks per day (adjusted mean drinks 0.6 vs 2.7, IRR = 0.24, $p = .004$). In stratified analyses, intervention effects appeared larger among drug users compared with nonusers (adjusted mean drinks 0.8 vs 4.9, IRR = 0.17 for drug users [$p = .17$]; and adjusted mean drinks 1.6 vs 1.9, IRR = 0.85 for nonusers [$p = .64$]). However, effects were not significant in either subgroup. The intervention effect appeared to be stronger for those who used heroin or cocaine.

Readiness to change, quality of life, health care use, and alcohol problems

At 3 months among subjects with nondependent, unhealthy alcohol use, intervention subjects had better physical health-related quality of life (adjusted mean SF-12 physical component summary scores 43 vs 38, adjusted mean difference 5, $p = .03$), but among subjects with dependence, intervention was associated with worse physical health-related quality of life (adjusted physical component summary scores 36 vs 40, adjusted mean difference -4, $p = .02$). Physical component summary scores did not differ by intervention group at 12 months. Aside from an improvement in mental health-related quality of life in unadjusted analyses among nondependent intervention subjects at 12 months (mean SF-12 mental component summary scores 51 vs 43, unadjusted mean difference 8, $p = .02$), mental component summary scores did not differ by intervention group in unadjusted or adjusted analyses at 3 or 12 months. In adjusted and unadjusted analyses stratified by the presence or absence of dependence, intervention was not significantly associated with alcohol problems or readiness to change. Aside from a greater number of days hospitalized in adjusted analyses (12.3 vs 5.0 days, IRR = 2.47, $p = .01$) and greater emergency-department visits in unadjusted analyses (1.5 vs 1.0 visits, IRR = 1.55, $p = .03$) at 3 months, health care use among dependent intervention subjects did not differ by intervention group in unadjusted or adjusted analyses at 3 or 12 months.

Discussion

We assessed brief intervention effects and the possible moderating effects of various demographic and health characteristics among patients with the spectrum of unhealthy alcohol use who were identified by screening on a hospital medicine service. As such, our study is unique. In this study,

the brief motivational intervention had both hypothesized and unanticipated effects. The intervention increased receipt of treatment in women and may be beneficial in younger men but not older men with alcohol dependence. An alcohol-attributable medical diagnosis and higher cognitive functioning moderated the effects of the intervention in women; these women in the intervention group were more likely than those in the control group to receive treatment. Of note, poorer mental health, homelessness, other drug use, and readiness to change did not impact the effects of brief intervention on receipt of treatment.

Contrary to our hypothesis, brief intervention had little effect on alcohol consumption. Both intervention and control groups had lower consumption at follow-up than at study entry. Factors other than, or in addition to, the brief intervention may have played a role in decreasing consumption, including subjects' medical illnesses, hospitalization and related services, natural history, regression to the mean, and a detailed research assessment of alcohol use that may have motivated change. Among subjects with nondependent, unhealthy alcohol use, one adjusted comparison was significant—brief intervention was associated with less consumption (drinks/day) at 3 months. In subgroup analyses, the effect of brief intervention on consumption was limited to younger people, and appeared to be larger among those who used drugs (an unexpected finding). Homelessness, mental health-related quality of life, cognitive impairment, and readiness to change did not appear to moderate the effects of brief intervention on drinking. However, the sample sizes of the subgroups were small and thus were likely underpowered to detect interactions.

Brief intervention may also have led to improved physical and perhaps mental health-related quality of life among those with nondependent, unhealthy alcohol use. But we also found unexpected possible detrimental effects of brief intervention among dependent subjects on physical health-related quality of life and hospital and emergency-department use. Although unexpected, these potential adverse effects of brief intervention should not be ignored in future studies because we should not expect all brief intervention effects to be favorable. Intervention did not significantly affect the other nonconsumption outcomes we assessed.

Studies of the efficacy of brief intervention for unhealthy alcohol use in hospitalized patients have produced mixed results. In a systematic review of controlled studies of inpatients on hospital services, brief intervention was associated with decreased alcohol-related problems but not changes in consumption (Emmen et al., 2004). Among studies specific to medical inpatients, results are also conflicting—some support brief intervention, and others do not (Chick et al., 1985; Freyer-Adam et al., 2008; Kuchipudi et al., 1990; Saitz et al., 2007). Studies of inpatients who are hospitalized on nonmedical services (e.g., trauma, orthopedic surgery), however, have more consistently demonstrated the efficacy of brief intervention for decreasing alcohol consumption (Antti-Poika et al., 1988; Blondell et al., 2001; Elvy et al., 1988; Gentilello et al., 1999) and increasing alcohol-treatment entry and mutual-help group use (Antti-Poika et al., 1988; Blondell et al., 2001; Dunn and Ries, 1997; Elvy et al., 1988).

Several limitations should be considered when interpreting these results. First, this is a secondary data analysis assessing multiple associations; the results should be viewed only as exploratory and hypothesis-generating rather than confirmatory (Lagakos, 2006). Second, the sample size precluded adjustment for all confounding factors in subgroup analyses and also limited the detection of both moderators and intervention effects in subgroups, particularly in women and subjects without dependence. Despite these limitations, this study suggests that although universal screening and brief intervention on a medicine service may not be effective across a wide spectrum of patients, certain groups may benefit from such efforts. Brief intervention shows promise for (1) increasing receipt of treatment among alcohol-dependent women (particularly those with higher cognitive functioning or an alcohol-attributable diagnosis) and younger men, and (2) decreasing consumption among those with nondependent, unhealthy alcohol use. But in this setting, brief intervention's effects on a wide range of clinically important outcomes were not robust (regardless of dependence status). Still, there is some reason for optimism given that factors hypothesized to impede the success of brief intervention (e.g., poorer mental health, drug use, homelessness) do not appear to be responsible for the lack of overall effects on consumption.

The evidence from controlled trials in primary care settings is clear: Brief intervention for patients with nondependent, unhealthy alcohol use identified by screening has modest efficacy for decreasing consumption. Evidence from studies in other populations, however, remains conflicting. Regardless, screening and brief intervention programs for unhealthy alcohol and other drug use are being disseminated worldwide (e.g., by large federal efforts in the United States and by the World Health Organization) in diverse medical settings in which patients may differ greatly in circumstance and severity from those deemed most likely, according to the best evidence, to benefit from brief intervention. As such, attention should return to issues of efficacy as well as to effectiveness. Further, it should be clear that the targets of screening and brief intervention—unhealthy alcohol and other drug use—are not monolithic and amenable to single, simple solutions. Therefore, research should begin to address when, for whom, and under what circumstances these procedures are and are not effective (O'Connor, 2007), and clinical implementation efforts should consider these complexities as dissemination of screening and brief intervention programs proceeds.

Acknowledgment

We are grateful for the data management efforts of Alison Pedley at DM-STAT, Malden, MA.

References

Adams, W.L., Yuan, Z., Barboriak, J.J., and Rimm, A.A. Alcohol-related hospitalizations of elderly people: Prevalence and geographic variation in the United States. JAMA **270:** 1222-1225, 1993.

Andreasen, N.C., Endicott, J., Spitzer, R.L., and Winokur, G. The family history method using diagnostic criteria: Reliability and validity. Arch. Gen. Psychiat. **34:** 1229-1235, 1977.

Antti-Poika, I., Karaharju, E., Roine, R., and Salaspuro, M. Intervention of heavy drinking: A prospective and controlled study of 438 consecutive injured male patients. Alcohol Alcsm **23:** 115-121, 1988.

Bates, M.E., Pawlak, A.P., Tonigan, J.S., and Buckman, J.F. Cognitive impairment influences drinking outcome by altering therapeutic mechanisms of change. Psychol. Addict. Behav. **20:** 241-253, 2006.

Blanchard, E.B., Jones-Alexander, J., Buckley, T.C., and Forneris, C.A. Psychometric properties of the PTSD Checklist (PCL). Behav. Res. Ther. **34:** 669-673, 1996.

Blondell, R.D., Looney, S.W., Northington, A.P., Lasch, M.E., Rhodes, S.B., and McDaniels, R.L. Can recovering alcoholics help hospitalized patients with alcohol problems? J. Fam. Pract. **50:** 447, 2001.

Blow, F.C., Barry, K.L., Walton, M.A., Maio, R.F., Chermack, S.T., Bingham, C.R., Ignacio, R.V., and Strecher, V.J. The efficacy of two brief intervention strategies among injured, at-risk drinkers in the emergency department: Impact of tailored messaging and brief advice. J. Stud. Alcohol **67:** 568-578, 2006.

Boyd, J.H., Weissman, M.M., Thompson, W.D., and Myers, J.K. Screening for depression in a community sample: Understanding the discrepancies between depression symptom and diagnostic scales. Arch. Gen. Psychiat. **39:** 1195-1200, 1982.

Center for Substance Abuse Treatment. Screening, Brief Intervention, and Referral to Treatment, Rockville, MD: Substance Abuse and Mental Health Services Administration (available at: http://sbirt.samhsa.gov/) accessed 7/9/2007.

Chick, J., Lloyd, G., and Crombie, E. Counselling problem drinkers in medical wards: A controlled study. Brit. Med. J. Clin. Res. Ed. **290:** 965-967, 1985.

Dunn, C.W. and Ries, R. Linking substance abuse services with general medical care: Integrated, brief interventions with hospitalized patients. Amer. J. Drug Alcohol Abuse **23:** 1-13, 1997.

Elvy, G.A., Wells, J.E., and Baird, K.A. Attempted referral as intervention for problem drinking in the general hospital. Brit. J. Addict. **83:** 83-89, 1988.

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, 2004.

Folstein, M.F., Folstein, S.E., and McHugh, P.R. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. J. Psychiat. Res. **12:** 189-198, 1975.

Freyer-Adam, J., Coder, B., Baumeister, S.E., Bischof, G., Riedel, J., Paatsch, K., Wedler, B., Rumpf, H.-J., John, U., and Hapke, U. Brief alcohol intervention for general hospital inpatients: A randomized controlled trial. Drug Alcohol Depend. **93:** 233-243, 2008.

Gentilello, L.M., Rivara, F.P., Donovan, D.M., Jurkovich, G.J., Daranciang, E., Dunn, C.W., Villaveces, A., Copass, M., and Ries, R.R. Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. Ann. Surg. **230:** 473-483, 1999.

Guth, S., Lindberg, S.A., Badger, G.J., Thomas, C.S., Rose, G.L., and Helzer, J.E. Brief intervention in alcohol-dependent versus nondependent individuals. J. Stud. Alcohol Drugs **69:** 243-250, 2008.

Heather, N., Rollnick, S., Bell, A., and Richmond, R. Effects of brief counselling among male heavy drinkers identified on general hospital wards. Drug Alcohol Rev. **15:** 29-38, 1996.

Institute of Medicine. Broadening the Base of Treatment for Alcohol Problems, Washington, DC: National Academies Press, 1990.

Katz, J.N., Chang, L.C., Sangha, O., Fossel, A.H., and Bates, D.W. Can comorbidity be measured by questionnaire rather than medical record review? Med. Care **34:** 73-84, 1996.

Kubany, E.S., Haynes, S.N., Leisen, M.B., Owens, J.A., Kaplan, A.S., Watson, S.B., and Burns, K. Development and preliminary validation of a brief broad-spectrum measure of trauma exposure: The Traumatic Life Events Questionnaire. Psychol. Assess. **12:** 210-224, 2000.

Kuchipudi, V., Hobein, K., Flickinger, A., and Iber, F.L. Failure of a 2-hour motivational intervention to alter recurrent drinking behavior in alcoholics with gastrointestinal disease. J. Stud. Alcohol **51:** 356-360, 1990.

Lagakos, S.W. The challenge of subgroup analyses—reporting without distorting. New Eng. J. Med. **354:** 1667-1669, 2006.

McLellan, A.T., Alterman, A.I., Cacciola, J., Metzger, D., and O'Brien, C.P. A new measure of substance abuse treatment: Initial studies of the treatment services review. J. Nerv. Ment. Dis. **180:** 101-110, 1992a.

McLellan, A.T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., and Argeriou, M. The Fifth Edition of the Addiction Severity Index. J. Subst. Abuse Treat. **9:** 199-213, 1992b.

McManus, S., Hipkins, J., Haddad, P., Guthrie, E., and Creed, F. Implementing an effective intervention for problem drinkers on medical wards. Gen. Hosp. Psychiat. **25:** 332-337, 2003.

Miller, W.R. Form 90: A Structured Assessment Interview for Drinking and Related Behaviors: Test Manual. NIAAA Project Match Monograph Series, Vol. 5, NIH Publication No. 96-4004, Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism, 1996.

Miller, W.R. and Rollnick, S. Motivational Interviewing: Preparing People to Change Addictive Behavior, New York: Guilford Press, 1991.

Miller, W.R. and Rollnick, S. Motivational Interviewing: Preparing People for Change, 2nd Edition, New York: Guilford Press, 2002.

Miller, W.R. and Tonigan, J.S. Assessing drinkers' motivation for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). Psychol. Addict. Behav. **10:** 81-89, 1996.

Miller, W.R., Tonigan, J.S., and Longabaugh, R. The Drinker Inventory of Consequences (DrInC): An Instrument for Assessing Adverse Consequences of Alcohol Abuse: Test Manual. NIAAA Project MATCH Monograph Series, Vol. 4, NIH Publication No. 95-3911, Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism, 1995b.

Miller, W.R., Zweben, A., DiClemente, C.C., and Rychtarik, R.G. Motivational Enhancement Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals with Alcohol Abuse and Dependence. NIAAA Project MATCH Monograph Series, Vol. 2, DHHS Publication No. (ADM) 92-1894, Washington: Government Printing Office, 1995a.

O'Connor, P. Brief interventions for problem drinking: Another piece of the puzzle (editorial). Ann. Intern. Med. **146:** 223-225, 2007.

Patrick, D.L., Cheadle, A., Thompson, D.C., Diehr, P., Koepsell, T., and Kinne, S. The validity of self-reported smoking: A review and meta-analysis. Amer. J. Publ. Hlth **84:** 1086-1093, 1994.

Radloff, L. The CES-D scale: A self-report depression scale for research in the general population. Appl. Psychol. Meas. **1:** 385-401, 1977.

Saitz, R., Freedner, N., Palfai, T.P., Horton, N.J., and Samet, J.H. The severity of unhealthy alcohol use in hospitalized medical patients: The spectrum is narrow. J. Gen. Intern. Med. **21:** 381-385, 2006.

Saitz, R., Palfai, T.P., Cheng, D.M., Horton, N.J., Freedner, N., Dukes, K., Kraemer, K.L., Roberts, M.S., Guerriero, R.T., and Samet, J.H. Brief intervention for medical inpatients with unhealthy alcohol use: A randomized, controlled trial. Ann. Intern. Med. **146:** 167-176, 2007.

Schmidt, L.A., Ye, Y., Greenfield, T.K., and Bond, J. Ethnic disparities in clinical severity and services for alcohol problems: Results from the National Alcohol Survey. Alcsm Clin. Exp. Res. **31:** 48-56, 2007.

Smith, K.L., Horton, N.J., Saitz, R., and Samet, J.H. The use of the mini-mental state examination in recruitment for substance abuse research studies. Drug Alcohol Depend. **82:** 231-237, 2006.

Sobell, L.C. and Sobell, M.B. Timeline follow-back: A technique for assessing self-reported alcohol consumption. In: Litten, R.Z. and Allen, J.P. (Eds). Measuring Alcohol Consumption: Psychosocial and Biochemical Methods, Totowa, NJ: Humana Press, 1992, pp. 41-72.

Spitzer, R.L., Williams, J.B., Kroenke, K., Linzer, M., deGruy, F.V., III, Hahn, S.R., Brody, D., and Johnson, J.G. Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD 1000 study. JAMA **272:** 1749-1756, 1994.

U.S. Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: Recommendation statement. Ann. Intern. Med. **140:** 554-556, 2004.

Walton, M.A., Goldstein, A.L., Chermack, S.T., McCammon, R.J., Cunningham, R.M., Barry, K.L., and Blow, F.C. Brief alcohol intervention in the emergency department: Moderators of effectiveness. J. Stud. Alcohol Drugs **69:** 550-560, 2008.

Ware, J.E., Kosinkski, M., and Keller, S.D. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales, 3rd Edition, Lincoln, RI: QualityMetric, 1998.

Watson, H.E. A study of minimal interventions for problem drinkers in acute care settings. Int. J. Nurs. Stud. **36:** 425-434, 1999.

Weisner, C., Mertens, J., Parthasarathy, S., Moore, C., and Lu, Y. Integrating primary medical care with addiction treatment: A randomized controlled trial. JAMA **286:** 1715-1723, 2001.

Wilk, A.I., Jensen, N.M., and Havighurst, T.C. Meta-analysis of randomized control trials addressing brief interventions in heavy alcohol drinkers. J. Gen. Intern. Med. **12:** 274-283, 1997.

JACC: CARDIOVASCULAR IMAGING VOL. 2, NO. 11, 2009

ISSN 1936-878X/09/$36.00
PUBLISHED BY ELSEVIER INC. DOI:10.1016/j.jcmg.2009.09.007

Integrated Assessment of Diastolic and Systolic Ventricular Function Using Diagnostic Cardiac Magnetic Resonance Catheterization

Validation in Pigs and Application in a Clinical Pilot Study

Boris Schmitt, MD,* Paul Steendijk, PHD,† Karsten Lunze, MD,* Stanislav Ovroutski, MD,* Jan Falkenberg, MD,* Pedram Rahmanzadeh, MD,* Nizar Maarouf, MD,* Peter Ewert, MD, PHD,* Felix Berger, MD, PHD,* Titus Kuehne, MD, PHD*

Berlin, Germany; and Leiden, the Netherlands

OBJECTIVES This study sought to develop and validate a method for the integrated analysis of systolic and diastolic ventricular function.

BACKGROUND An integrated approach to assess ventricular pump function, myocontractility (end-systolic pressure–volume relationship [ESPVR]), and diastolic compliance (end-diastolic pressure–volume relation [EDPVR]) is of high clinical value. Cardiac magnetic resonance (CMR) is well established for measuring global pump function, and catheterization-combined CMR was previously shown to accurately measure ESPVR, but not yet the EDPVR.

METHODS In 8 pigs, the CMR technique was compared with conductance catheter methods (gold standard) for measuring the EDPVR in the left and right ventricle. Measurements were performed at rest and during dobutamine administration. For CMR, the ESPVR was estimated with a single-beat approach by synchronizing invasive ventricular pressures with cine CMR–derived ventricular volumes. The EDPVR was determined during pre-load reduction from additional volume data that were obtained from real-time velocity-encoded CMR pulmonary/aortic blood flow measurements. Pre-load reduction was achieved by transient balloon occlusion of the inferior vena cava. The stiffness coefficient β was calculated by an exponential fit from the EDPVR. After validation in the animal experiments, the EDPVR was assessed in a pilot study of 3 patients with a single ventricle using identical CMR and conductance catheter techniques.

RESULTS Bland-Altman tests showed good agreement between conductance catheter–derived and CMR-derived EDPVR. In both ventricles of the pigs, dobutamine enhanced myocontractility ($p < 0.01$), increased stroke volume ($p < 0.01$), and improved diastolic function. The latter was evidenced by shorter early relaxation ($p < 0.05$), a downward shift of the EDPVR, and a decreased stiffness coefficient β ($p < 0.05$). In contrast, in the patients, early relaxation was inconspicuous but the EDPVR shifted left-upward and the stiffness constant remained unchanged. The observed changes in diastolic function were not significantly different when measured with conductance catheter and CMR.

CONCLUSIONS This novel CMR method provides differential information about diastolic function in conjunction with parameters of systolic contractility and global pump function. (J Am Coll Cardiol Img 2009;2:1271–81) © 2009 by the American College of Cardiology Foundation

From the *Department of Congenital Heart Disease and Pediatric Cardiology, Deutsches Herzzentrum Berlin, Berlin, Germany; and the †Departments of Cardiology and Cardiothoracic Surgery, Leiden University Medical Center, Leiden, the Netherlands. 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.

Manuscript received May 11, 2009; revised manuscript received September 16, 2009, accepted September 17, 2009.

JACC: CARDIOVASCULAR IMAGING, VOL. 2, NO. 11, 2009
NOVEMBER 2009:1271–81

Heart failure is a common cause of mortality and death. In pathophysiology, different forms of systolic and diastolic heart failure can be defined (1–3). In many patients, combined forms of systolic and diastolic dysfunction coexist and are difficult to differentiate. Substantiated knowledge about the predominant form of heart failure is essential for optimizing treatments. Thus, an integrated approach for evaluating systolic contractility in conjunction with diastolic relaxation and compliance would be of high clinical value.

See page 1282

Recent innovations in imaging provided a variety of new methods for the noninvasive assessment of cardiac function. These methods allow determination of stress–strain relation or inflow profiles of the ventricles, for example. However, many of these parameters are pre-load–dependent and/or afterload-dependent or only reflect regional myocardial function. Therefore, analysis of pressure–volume relations is still regarded as the most reliable way to obtain load independent parameters of contractility and diastolic compliance (1,2).

ABBREVIATIONS AND ACRONYMS

β = stiffness constant

CMR = cardiac magnetic resonance

EDPVR = end-diastolic pressure–volume relation

Emax = slope of the end-systolic pressure–volume relation

Emax,i = slope of the end-systolic pressure–volume relation indexed to 100 mg myocardial muscle mass

ESPVR = end systolic pressure–volume relation

τ = parameter of early diastolic relaxation

VEC = velocity-encoded cine

Advances in fast imaging techniques made diagnostic cardiac magnetic resonance (CMR) catheterization a realistic option (4–8). It was shown to provide accurate estimates of the end-systolic pressure–volume relationship (ESPVR) by combining invasive ventricular pressures with cine CMR derived ventricular volumes. The ESPVR is widely considered the optimal quantification of systolic contractile function (6,8).

In the current study, we propose a novel CMR approach that combines real-time–derived ventricular volumes with invasively measured ventricular pressures for assessment of the end-diastolic pressure–volume relationship (EDPVR). The EDPVR characterizes ventricular chamber stiffness in a relatively load-independent fashion and showed practical importance for the assessment of patients with diastolic dysfunction (1,2,9,10).

The aim of this experimental study was to develop a CMR method for assessment of the EDPVR and to validate this method in the pig right and left ventricle using conductance catheter techniques as a gold standard reference. Subsequently, the applicability of this method in a clinical context was evaluated in a pilot study involving patients with a single ventricle after Fontan operation. In these patients, diastolic dysfunction has been reported by several investigators (11,12).

METHODS AND STUDY DESIGN

Animal experiments. The animal experiments were authorized by the responsible animal care authorities. The validation study was conducted in 8 pigs (31 ± 5 kg). The animals were pre-medicated with 5 mg/kg azaperone and 10 mg/kg ketamine intramuscularly. Anesthesia was maintained with 1.5% isoflurane inhalation. All CMR and conductance catheter measurements were performed at end expiratory breath-hold and during muscle relaxation with 0.01 mg/kg vecuronium bromide intravenously. After completion of measurements, animals were euthanized.

The timeline and brief description of the protocol are shown in Figure 1. Right and left ventricular pressure–volume relations were first assessed by conductance catheter (gold standard). Thereafter, the animals were transferred to the neighboring CMR laboratory. All measurements were performed at rest and repeated during continuous infusion of dobutamine at 10 μg/kg/min with at least a 10-min interval between inotropic stimulation and repeated measurements at rest.

Measured parameters. Parameters of ventricular global, myocontractile, and diastolic function were obtained using conductance catheter and/or CMR techniques.

GLOBAL PUMP FUNCTION (BY CINE CMR). This parameter is composed of ventricular end-diastolic, end-systolic, and stroke volume assessed by cine CMR.

MYOCONTRACTILE FUNCTION (BY CONDUCTANCE AND CMR CATHETERIZATION). The slope of the ESPVR (Emax) was defined as a measure of contractility and was derived from the pressure–volume loops as determined with conductance catheter and CMR techniques. The Emax was indexed to 100 mg myocardial muscle mass (Emax,i).

DIASTOLIC FUNCTION (BY CONDUCTANCE AND CMR CATHETERIZATION). From pressure measurements, we derived τ as a parameter of early diastolic relaxation. The stiffness constant (β) was determined from a set of EDPVR and was defined as a measure of diastolic compliance. The β value was

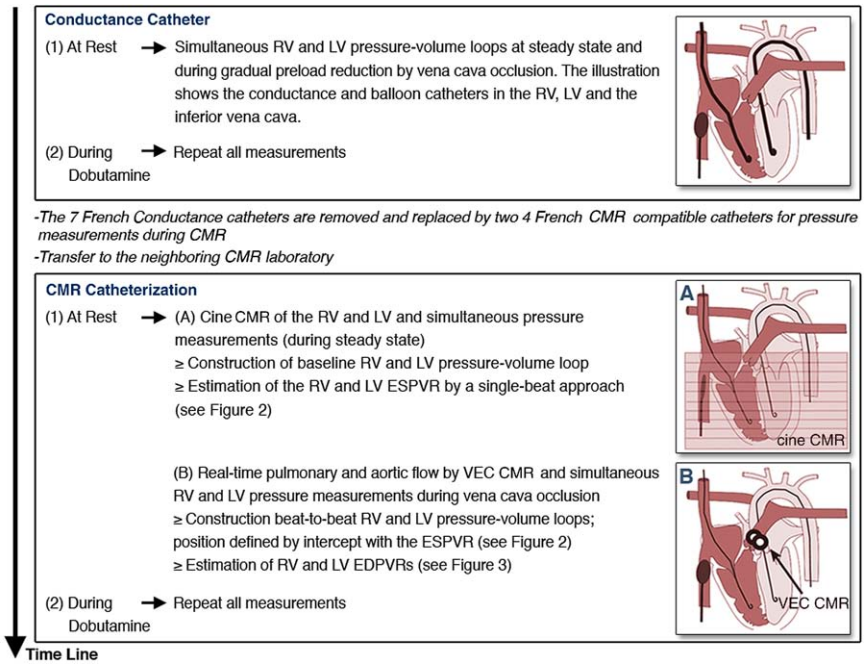


Figure 1. Time Line of the Study

The animals were first assessed at rest and during dobutamine administration using a conductance catheter technique. Then they were moved to the neighboring CMR laboratory and investigated again at rest and during dobutamine application. The figure shows the position of the catheters in the left and right ventricle and the inferior vena cava. The CMR was performed using a cine CMR method in the axial plane (**shaded box in A**) and by VEC CMR measurements in the aorta and the pulmonary truncus (**arrow in B**). CMR = cardiac magnetic resonance; EDPVR = end-diastolic pressure–volume relation; ESPVR = end-systolic pressure–volume relation; LV = left ventricle; RV = right ventricle; VEC = velocity-encoded cine.

calculated from the EDPVR with an exponential regression: $EDP = Ae^{\beta \cdot EDV}$ where EDP = end-diastolic pressure, EDV = end-diastolic volume, and A = curve-fitting constant. The β value was indexed to ventricular volumes for creating a dimensionless index where appropriate.

Concept of CMR-derived EDPVR. Because of its inherent nonlinearity, the assessment of the EDPVR ideally requires pressure–volume data acquired at multiple loading conditions. Therefore, ventricular loading was gradually altered by vena cava balloon occlusion. Beat-to-beat alteration in volume load and pressure were measured simultaneously with real-time CMR and liquid-filled catheters. The concept for measuring EDPVR using CMR-catheterization is based on working steps that are shown in Figures 2 to 4.

STEP A: CINE CMR. Biventricular phasic absolute volumes were acquired over several cardiac cycles with multislice-multiphase cine CMR. During CMR, ventricular pressures were measured continuously, averaged, and synchronized with the cine CMR derived volumes to construct a baseline pressure–volume loop under steady-state conditions. Synchronization of pressures and volumes was achieved by a trigger signal (Fig. 2, left). The ESPVR was estimated from the baseline loop using a single-beat approach as previously described (7,13).

STEP B: REAL-TIME CMR. Instantaneous blood flows were measured using real-time velocity-encoded cine (VEC) CMR in the pulmonary trunk and ascending aorta. Recording of ventricular pressures was started with the beginning of CMR and synchronized with the flow using the trigger signal. At steady state and in the absence of atrioventricular valve insufficiency, effective right or left ventricular stroke volumes are considered being equivalent to the effective antegrade pulmonary or aortic blood flow volumes. Therefore, ventricular chamber volumes can be computed by subtracting the effective stroke volumes from the end-diastolic volume of the cine CMR measurements (marked baseline EDPV in Fig. 3). After 3 to 4 heartbeats, pre-load was lowered by transient balloon occlusion of the vena cava. Importantly, for these unloaded beats

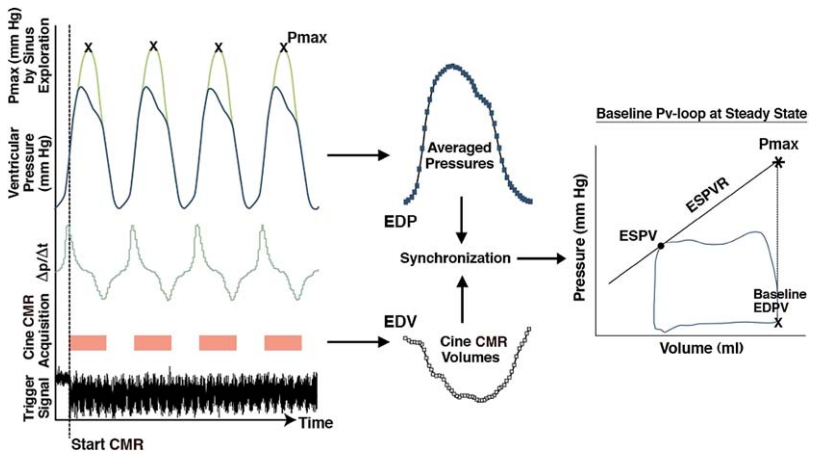


Figure 2. Baseline Pressure–Volume Loop at Steady-State Condition

The baseline loop **(right)** was constructed by synchronizing ventricular pressures with cine CMR–derived absolute volume data. The ESPVR was estimated from a single-beat approach. Pmax was calculated by a sinus wave extrapolation of the ventricular pressure curve. Further details are provided in the text. EDP = end-diastolic pressure; EDPV = end-diastolic pressure–volume point; EDV = end-diastolic volume; ESPV = end-systolic pressure–volume point; Pmax = maximum isovolumic ventricular pressure; Pv = pressure-volume; other abbreviations as in Figure 1.

ventricular filling is unknown, thus the absolute volume (horizontal position of the pressure–volume loop) is undetermined. Initially, end-diastolic volumes of all unloaded loops were arbitrarily arranged to show diminishment in stroke volume.

STEP C: POST-PROCESSING. To calibrate the absolute volume of the flow-derived real-time pressure–volume loops, we matched the end-systolic volume of each unloaded beat with the ESPVR volume intercept at the measured corresponding end-systolic pressure. The resulting end-diastolic pressure–volume points were used to determine the EDPVR and to calculate β.

Conductance catheter. The conductance catheter study was performed using a Leycom signal processor (CD-Leycom, Zoetermeer, the Netherlands) and 6-F to 7-F dual-field catheters (Millar, Houston, Texas). Catheter tips were positioned in the left or right ventricular apex, and measurements were performed during transient pre-load reduction by vena cava balloon occlusion using a 36-mm sizing balloon inflated with isotonic saline solution (AGA Medical, Plymouth, Minnesota). Parallel conductance was determined by the saline dilution method (14). The calibration factor alpha was calculated from CMR-derived stroke volumes.

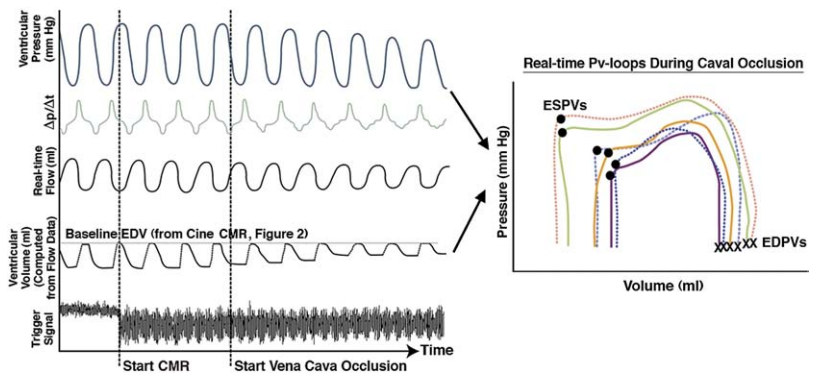


Figure 3. Composition of the Real-Time Pressure–Volume Loops From Simultaneously Measured Ventricular Pressure and Real-Time Velocity-Encoded Cine CMR–Derived Volume Data

Note that initially, end-diastolic volumes of all loops are unknown and were arbitrarily drawn equally spaced to better show the gradual decrease in end-diastolic pressure during unloading. Abbreviations as in Figures 1 and 2.

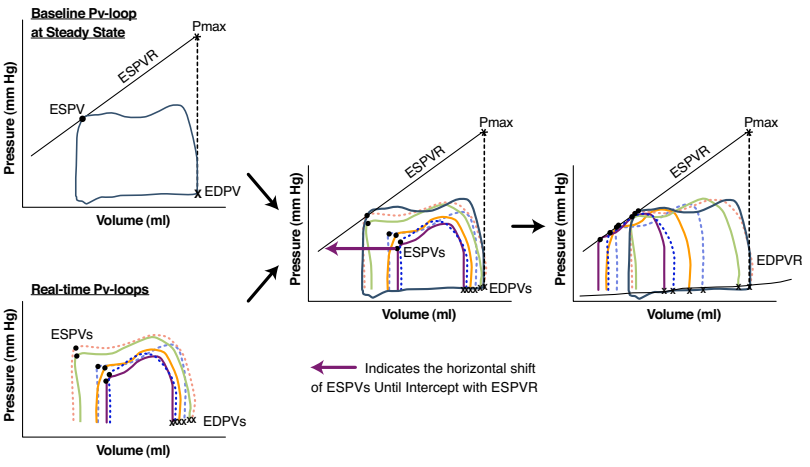


Figure 4. Stepwise Arrangement of Baseline and Real-Time Pressure-Volume Loops for Assessment of EDPVR

The horizontal position of the real-time pressure–volume loops (**lower left** and Fig. 3) was calibrated by matching the end-systolic pressure–volume points with the ESPVR determined from the baseline pressure–volume loop (**upper left** and Fig. 2). The ESPVs of each beat were shifted to match to the intercept of the ESPVR at the corresponding pressure **(middle)**. The EDPVR was determined by an exponential fit through the resulting EDPVs **(right)**. Further details are given in the text. Abbreviations as in Figures 1 and 2.

Data were post-processed using Conduct NT-V2.0.1 (CD-Leycom) and a MATLAB platform (The MathWorks, Natick, Massachusetts).

CMR catheterization. CMR ACQUISITION OF VENTRICULAR PRESSURES AND VISUALIZATION OF CATHETERS. After completion of conductance catheter measurements, catheters were replaced under X-ray angiography by 4-F fluid-filled pigtail diagnostic catheters (Cordis, Warren, New Jersey). The balloon catheter for vena cava occlusions was kept unchanged in the inferior vena cava. The pressure catheters were connected to a pressure transducer (Becton-Dickinson, Franklin Lakes, New Jersey) amplified, recorded, and analyzed with Ponemah software (all DSI, St. Paul, Minnesota). Initial pressure recordings were obtained in the catheter laboratory. Thereafter the animals were moved to the CMR laboratory, where steady state hemodynamic conditions were confirmed. In the CMR laboratory, an additional pressure transducer was placed together with the animal within the bore. Radiofrequency pulses induced pressure signal artifacts on this transducer at the beginning of each VEC CMR measurement. These artifacts were used as a trigger signal for synchronizing measured pressures with acquired volume and flow signals (Figs. 2 and 3). Correct catheter position during CMR was confirmed on interactive real-time CMR as previously described (15).

Ventricular volumes and myocardial mass. All CMR studies were performed on a 1.5-T scanner (release 2.6.1, Philips Intera, Best, the Netherlands). Ventricular chamber volumes and myocardial mass were determined from a stack of multislice-multiphase steady-state free-precession cine CMRs covering the entire heart (16). Sequence parameters were: repetition time/echo time 3.4/1.7 ms, slice thickness 6 mm, no gap, in-plane resolution 1.9 × 1.3 mm, 45 phases per cardiac cycle, number of averages 1, sensitivity encoding reduction factor 2. Analysis was performed using View Forum software (Release 6.1, Philips). Biventricular endocardial and epicardial borders were manually traced for computing ventricular volumes and myocardial mass where the septum was accounted left ventricular. 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.

Pulmonary and aortic blood flow. Quantitative blood flow was measured using real-time VEC CMR (17) orthogonal to the dominating flow direction in the pulmonary trunk and the ascending aorta. For the measurements, we corrected for potential phase errors arising from the concomitant magnetic field. Sequence parameters were: repetition time/echo time 23/6.5 ms, 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, echo planar imaging factor 41. This resulted in the dynamic scan time of 31 ms for the acquisi-

Table 1. Animal Experiments, Parameters of Cardiac Function by Cine CMR and VEC CMR (Where Indicated)

	At Rest	During Dobutamine Administration
General characteristics		
Body weight (kg)	30.3 ± 6.8	NA
RV muscle mass (g)	27.8 ± 6.3	NA
LV muscle mass (g)	83.9 ± 7.2	NA
Global ventricular pump function		
RV end-diastolic volume (ml)	45.6 ± 9.3	43.6 ± 8.4
RV end-systolic volume (ml)	15.1 ± 4.1	8.6 ± 3.2*
RV stroke volume (ml)	30.6 ± 4.5	35.3 ± 6.3*
RV stroke volume by VEC CMR (ml)	29.6 ± 3.1	34.9 ± 6.5*
LV end-diastolic volume (ml)	46.5 ± 9.6	45.8 ± 7.4
LV end-systolic volume (ml)	16.7 ± 6.3	10.1 ± 4.9*
LV stroke volume (ml)	29.7 ± 4.1	35.7 ± 5.7*
LV stroke volume by VEC CMR (ml)	29.1 ± 4.4	36.4 ± 5.3*
Cardiac output (l/min)	4.0 ± 1.2	6.4 ± 1.4*

*Significant differences ($p < 0.05$) between measurements at rest versus during dobutamine administration. LV = left ventricular; NA = not assessed; RV = right ventricular; VEC CMR = velocity-encoded cine cardiac magnetic resonance.

tion of 1 phase-contrast image, thus an average of 15 acquisitions per heartbeat depending on heart rate. Data analysis was performed with View Forum software. Antegrade and retrograde flows were measured as described elsewhere (18). For validation purposes, we compared left and right ventricular stroke volumes at baseline (without pre-load reduction) as measured with cine CMR and real-time VEC CMR each at rest and during dobutamine administration.

Clinical pilot study. The clinical study was performed in 3 patients with total cavopulmonary connection (Fontan). The patients were referred to our institution for cardiac catheterization and CMR because of decreasing exercise capacity and therefore to determine ventricular systolic and diastolic function and cardiovascular anatomy. After catheterization, the conscious patients were transferred to the CMR laboratory. Measurements were performed during breath-hold at end expiration. All gave informed consent for the study, which was approved by the responsible institutional review committee (reference number 47/04). Except for minor variation in size of the catheters, the conductance and CMR procedures were performed exactly as described in the animal experiments.

Statistical analysis. Agreements between conductance catheter- and CMR-derived β were determined using Bland-Altman tests. Differences between conductance catheter- and CMR-derived parameters as well as measurements at rest and during dobutamine administration were analyzed with a paired Student t test and Bonferroni correction for multiple comparisons where appropriate. Data are expressed as mean ± SD.

RESULTS

Validation study. COMPARISON OF CINE CMR AND VEC CMR STROKE VOLUMES. The data are shown in Table 1. At rest and during dobutamine administration, there was no significant difference between these methods.

CMR versus conductance catheter–derived diastolic compliance. Sequential conductance catheter and CMR measurements were realized at similar hemodynamic conditions, evidenced by the fact that right and left ventricular pressures as well as heart rates were not significantly different between the 2 experimental stages (Table 2). The hemodynamic responses to dobutamine also were similar (Table 2).

The Bland-Altman test showed good agreement between β values determined with the 2 methods at rest and during dobutamine administration (Fig. 5). Importantly, conductance catheter–derived and CMR-derived pressure–volume loops showed parallel changes for measurements of β at rest and during stress (Table 2, Fig. 6). The relative changes of Emax,i were also at similar levels (Table 2).

Response to dobutamine. GLOBAL PUMP FUNCTION (BASED ON CINE CMR). As expected, in response to dobutamine, the right and left ventricle showed a significant increase in stroke volume and cardiac output ($p < 0.01$) (Table 1).

MYOCONTRACTILITY (BASED ON CONDUCTANCE AND CMR CATHETERIZATION). Inotropic stimulation with dobutamine increased Emax,i significantly in both ventricles ($p < 0.01$). The response to dobutamine was more pronounced in the left compared with the right ventricle (Table 2). The noted changes of Emax,i were similar for the conductance catheter and CMR measurements (Table 2).

DIASTOLIC FUNCTION (BASED ON CONDUCTANCE AND CMR CATHETERIZATION). Active early relaxation, as indicated by smaller τ, improved significantly during dobutamine administration in both ventricles ($p < 0.01$) (Table 1). In addition, in all animals and both ventricles, the EDPVR shifted toward the bottom right of the pressure–volume diagram (Fig. 6). There was also a slight but significant decrease of β ($p < 0.05$) (Table 2, Fig. 5). The noted changes were again similar for the conductance catheter and CMR measurements.

Table 2. Animal Experiments, Parameters of Cardiac Function by CMR and Conductance Catheter

	CMR		Conductance Catheter	
	At Rest	During Dobutamine Administration	At Rest	During Dobutamine Administration
Hemodynamic data				
Heart rate	93 ± 16	133 ± 12*	89 ± 14	141 ± 14†
RV ESP/EDP pressures (mm Hg)	25.2 ± 7.7/5.1 ± 2.4	39.5 ± 5.7*/4.6 ± 1.4	24.2 ± 6.8/4.8 ± 2.8	41.2 ± 6.7†/4.4 ± 1.7
LV ESP/EDP pressures (mm Hg)	66.5 ± 10.1/4.6 ± 2.2	109.2 ± 11.9*/4.2 ± 1.5	63.4 ± 10.1/4.4 ± 2.1	103.2 ± 13.1†/4.1 ± 2.2
Myocardial contractility				
RV Emax,i (mm Hg/ml/100 g MM)	3.1 ± 1.9	5.2 ± 3.1*	3.6 ± 0.6	5.7 ± 1.5†
LV Emax,i (mm Hg/ml/100 g MM)	1.6 ± 0.6	3.3 ± 1.6*	1.8 ± 0.7	3.6 ± 1.1†
Diastolic relaxation				
RV τ (ms)	36.1 ± 9.2	27.2 ± 8.1*	39.4 ± 6.1	31.6 ± 5.4†
LV τ (ms)	31.8 ± 8.2	25.5 ± 7.3*	33.5 ± 6.6	22.9 ± 5.6†
Diastolic compliance				
RV β (1/ml)	0.024 ± 0.007	0.011 ± 0.008*	0.021 ± 0.005	0.010 ± 0.008†
RV β,i (1/ml/100 ml EDV)	0.011 ± 0.004	0.004 ± 0.002*	0.010 ± 0.002	0.007 ± 0.003†
LV β (1/ml)	0.027 ± 0.004	0.016 ± 0.006*	0.023 ± 0.009	0.015 ± 0.007†
LV β,i (1/ml/100 ml EDV)	0.012 ± 0.002	0.007 ± 0.003*	0.010 ± 0.003	0.004 ± 0.003†

*Significant differences ($p < 0.05$) between measurements at rest versus dobutamine for magnetic resonance. †Significant differences ($p < 0.05$) between measurements at rest versus dobutamine for conductance catheter measurements.
β = stiffness constant; β,i = stiffness constant indexed to 100 ml end-diastolic volume; EDP = end-diastolic pressure; EDV = end-diastolic volume; Emax,i = slope of the end-systolic pressure volume relation indexed to 100 g muscle mass; ESP = end-systolic pressure; MM = muscle mass; τ = parameter of early diastolic relaxation; other abbreviations as in Table 1.

Clinical experiments. Oxygen saturation was above 95%, also unaltered during dobutamine administration. All Fontan circuits were free from obstructions. Concordantly during dobutamine administration there was a decrease of end-diastolic and -systolic volumes (Table 3). Contractility, stroke

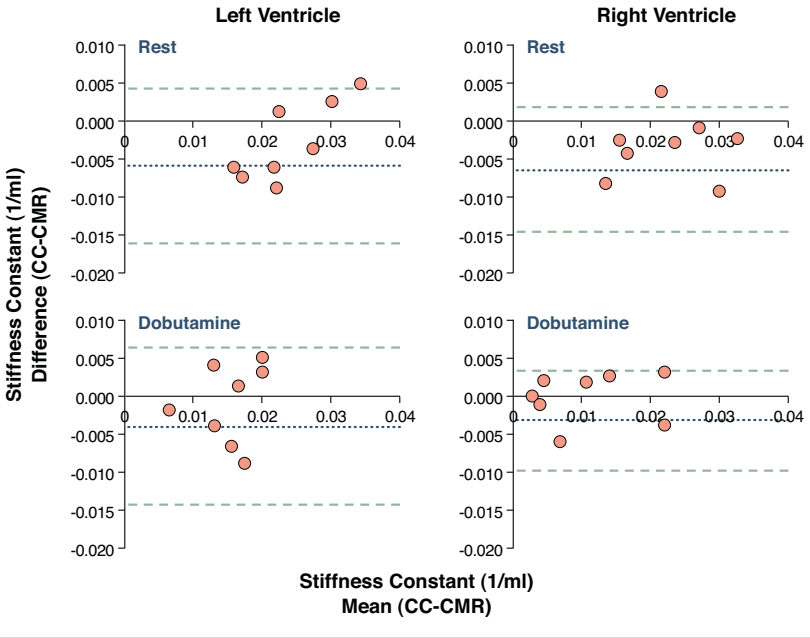


Figure 5. Comparison of Ventricular Stiffness Constant Measured by CC and by CMR

Bland-Altman plots show the difference between the results from conductance catheter (CC) and CMR measurements in 8 pigs for left and right ventricular stiffness constant **(left, right)**. Measurements were performed at rest and during dobutamine and are shown in separate plots. In all plots there is a negative bias, which indicates a slight but consistent overestimation by CMR in both the left and the right ventricle. Abbreviations as in Figure 1.

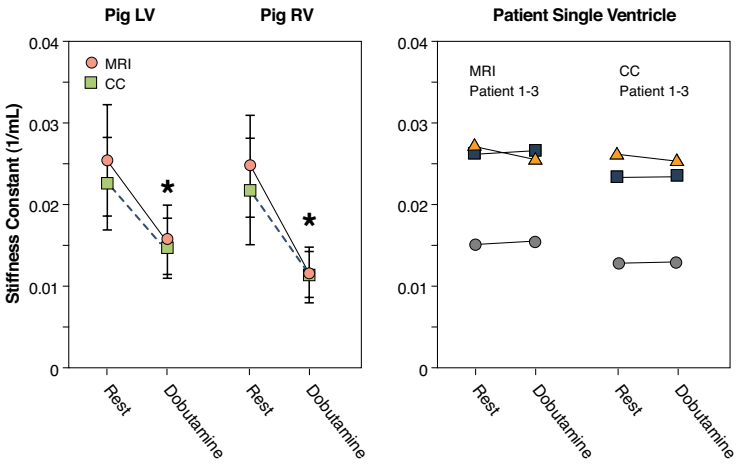


Figure 6. Changes in Stiffness Constant During Dobutamine Administration

Ventricular stiffness constants as measured with CMR and CC techniques in the pig left and right ventricle and in 3 patients with a single ventricle. There are no statistically significant differences between the CC and CMR measurements. In addition, this figure shows parallel decreases in the stiffness constant in CMR and CC measurements during dobutamine administration in pigs but no significant changes in Fontan patients. A value of $p < 0.05$ was considered significant. Abbreviations as in Figure 5.

volumes, and heart rate increased slightly, resulting in elevated cardiac outputs (Table 3). Active early relaxation was slightly enhanced. In contrast, the EDPVR shifted toward the left in the pressure–volume diagram and β remained nearly unchanged (Table 3, Figs. 6 and 7). Similar observations were made with conductance catheter and CMR techniques (Fig. 6).

Table 3. Clinical Study Data for Patients #1 to #3

	At Rest	During Dobutamine Administration
Global parameters		
Age (yrs)	19 ± 0.8	NA
Age at Fontan operation (yrs)	3 ± 0	NA
Body surface index (kg/m^2)	1.3 ± 0.1	NA
Heart rate (beats/min)	81 ± 3.3	113 ± 5
Ventricular volumes		
End-diastolic volume (ml/BSA)	97 ± 6.5	92.3 ± 7
End-systolic volume (ml/BSA)	53.7 ± 5	46.7 ± 4
Stroke volume (ml/BSA)	43.3 ± 1.9	45.7 ± 3.3
Ejection fraction (%)	44.8 ± 1.7	49.5 ± 1.2
Blood flow volumes		
Aorta (l/min)	3.7 ± 0.5	5.1 ± 0.5
Ventricular pressures		
End-diastolic pressure (mm Hg)	5.1 ± 0.4	5.8 ± 0.4
End-systolic pressure (mm Hg)	90.3 ± 4.9	103.7 ± 5.4
Contractility		
Emax,i (mm Hg/ml/100 g muscle mass)	3.5 ± 0.4	4.4 ± 0.4
Diastolic compliance and relaxation		
β (1/ml) × 100	2.3 ± 0.6	2.3 ± 0.8
β,i (1/ml/100 ml EDV) × 100	1.4 ± 0.3	1.4 ± 0.3
τ (ms)	36 ± 0.8	29.3 ± 1.2

Values are mean ± SD. All data are based on measurements during CMR. Individual data from each patient for conductance catheter-derived and CMR-derived stiffness constant are shown in Figure 7.
BSA = body surface area; other abbreviations as in Table 2.

DISCUSSION

We developed and validated a novel method for estimating the EDPVR, an index of ventricular chamber stiffness, by combining invasive pressure measurements with noninvasive real-time CMR volume and flow data. Our results indicate that the proposed method is suited to determine left and right ventricular chamber stiffness in conjunction with parameters of contractility and global pump function.

Technical consideration. The ESPVR and EDPVR are widely used in physiological studies because they reflect intrinsic systolic and diastolic myocardial function in a relatively load-independent fashion. Clinical application is still limited, mainly because of technical difficulties in performing the required measurements. The proposed CMR method might potentially expand clinical application because it is technically straightforward, and once the small and user-friendly catheters are in place, a complete set of right and left ventricular pressure–volume acquisitions can be obtained in <10 min.

Ideally, ESPVR and the EDPVR are determined from a family of pressure–volume loops during transient pre-load reduction (9,10). Previously, methods were introduced to estimate the ESPVR by single-

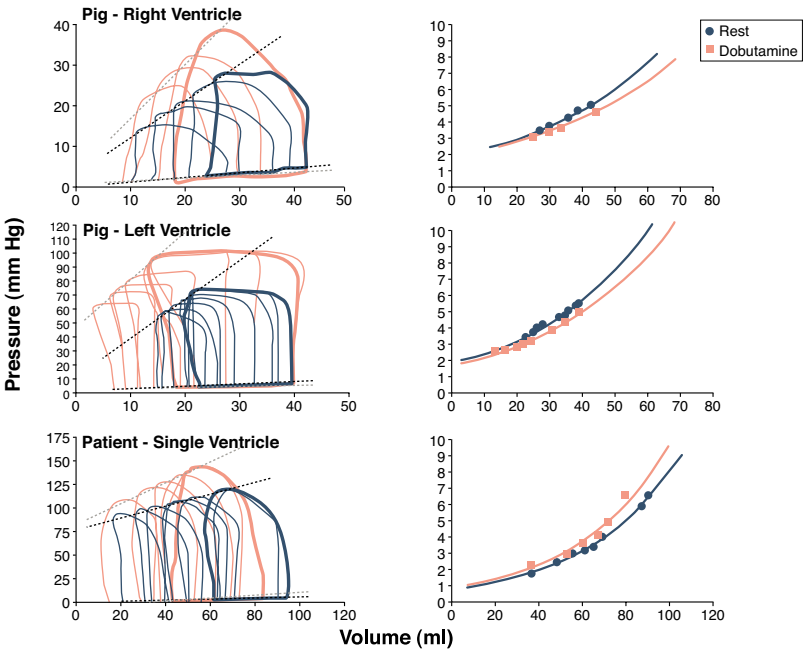


Figure 7. Ventricular Stiffness in the Healthy Animal Heart and in Fontan Patients

Representative CMR-derived pressure–volume loops of a pig right and left ventricle and a patient with a single ventricle. The end-systolic pressure–volume relation (ESPVR) and end-diastolic pressure–volume relation (EDPVR) at rest and during dobutamine administration are shown **(left, dashed lines)**. The EDPVRs are shown more detailed **(right)**. Note the left shift of the relation in the patient as opposed to the right-bottom shift in the pig's right and left ventricle during dobutamine administration. Abbreviations as in Figure 1.

beat approaches (7,8,13,19). However, similar simplified approaches to estimate the EDPVR are complicated by its inherent nonlinearity. Therefore, load interventions, which are commonly performed by balloon occlusion of the vena cava, are needed for determining profiles of ventricular chamber stiffness.

The conductance catheterization is an established method for measuring ventricular pressure–volume relations in animal experiments and human studies. However, by theory it requires a symmetrically shaped ventricle, and derived volumes need calibration with volumes obtained by a valid reference method, such as CMR (20). Consequently, conductance measurements in the nonsymmetrically shaped right ventricle or in a variety of congenitally malformed hearts can be problematic. Particularly these patients often need a differential analysis of ventricular systolic and diastolic function to determine optimal treatments.

To date, CMR is considered the gold standard in terms of accuracy and reproducibility for quantification of ventricular volumes (4,5). As mentioned, this measurement accuracy is crucial for the assessment of pressure–volume relations, particularly in asymmetrically shaped ventricles. Measurement errors by calibrating issues or by geometric assumptions are avoided (14,21). Similar to the conductance catheter, the CMR method proposed in this study requires invasive measurement of pressures. The fluid-filled CMR catheters used were, however, substantially smaller in size (4-F vs. 7-F) and easily positionable. To avoid susceptibility artifacts or radiofrequency pulse-induced heating during CMR, the application of metallic high-fidelity pressure-tipped catheters was excluded, and consequently materials without metallic components were used (15). A potential down side of fluid-filled catheters is the attenuation of pressure amplitudes. To minimize this effect, we limited the total length of the catheter and the connecting pressure line to <1 m.

In the current study, we determined the ESPVR using a single-beat approach as previously reported (7,8,13,19). This approach implies that ventricular volumes are measured by cine CMR over several beats and are synchronized with pressures averaged for the same time period. To minimize beat-to-beat variability, keep the hemodynamic condition and state of ventilation at identical levels during data acquisition.

Physiologic aspects. In the animal study, during dobutamine administration, diastolic function improved by faster early relaxation in conjunction with a slight bottom-right shift of the EDPVR. Similar to observations in previous studies (20,22), we also

noted a decrease of β. In contrast, in the patients, the EDPVR shifted during dobutamine administration toward the upper left in the pressure–volume diagram. Early relaxation improved slightly, and β did not change substantially.

Several reasons can account for the noted abnormal diastolic function in Fontan patients. The single ventricle, no matter of left/right type, is mostly of abnormal geometric shape with a direct impact on the mechanical properties of the ventricle and thus on systolic and diastolic function (12). Moreover, a recent pathohistological study showed abnormal myoarchitecture of the connective tissue matrix (23). During infancy these ventricles are exposed to prolonged cyanosis and volume load, which may induce fibrosis and thus have an impact on diastolic stiffness (24).

Accordance exists that the EDPVR reflects chamber capacity and compliance. Similar to the ESPVR, this relation is influenced by ventricular configuration, including size, and heart rate. Therefore, the impact of volume and heart rate changes during dobutamine administration must be considered when interpreting these data. The tendency toward a smaller β in the animal study and an invariable β in the patients was, however, also noted when β was indexed to ventricular chamber size (Tables 2 and 3).

In our study, parameters of contractility and chamber stiffness were measured with conductance catheter and CMR techniques nonsimultaneously. As mentioned earlier, alteration in volume load and heart rate, which can be induced among other factors by prolonged sedation, may have an impact on the ESPVR and EDPVR. We were able to keep baseline hemodynamic parameters, including heart rate, stable with changes of $<$10% between the sequential measurements (Table 2). We recommend measuring the EDPVR with the proposed CMR catheterization technique at heart rates of no more than 150 beats/min. This is for purely physiological reasons and for allowing reasonable temporal resolution in the real-time CMR flow measurements. The development of faster real-time CMR applications should be subject to future research.

Normalization of contractile and stiffness indexes may be required when comparing different study populations. Concerning the diastolic β, Burkhoff et al. (9) suggested the use of a dimensionless index by multiplying β by the myocardial wall volume, which can be directly obtained from the acquired cine CMR scans. The variability of wall volume measurements, particularly for the right ventricle, is known to be substantial (25). As an alternative, β might be normalized to chamber volume. This important issue must be systematically investigated in future research.

Study limitations. The CMR and conductance catheter measurements cannot be performed simultaneously. For comparison of measured data, we aimed for keeping the animals' physiological conditions, such as heart rate and blood pressure constant. Care must be taken when comparing physical exercise with dobutamine stress (26). Therefore, it would be inappropriate to directly translate our findings regarding chamber stiffness during dobutamine administration to exercise conditions. Some technical limitations should be mentioned. In the CMR setting, we measured ventricular pressures with fluid-filled catheters to avoid metallic components, in contrast with the high-fidelity, solid-state sensors incorporated in the conductance catheter. A fluid-filled catheter manometer system acts as a low-pass filter, thus high-frequency components are attenuated. However, by using relatively stiff and short catheters and carefully removing trapped air bubbles, adequate recordings of cardiac and arterial pressure signals can be obtained and the impact on derived indexes is expected to be very limited (27). For indexes that require high-frequency components, such as relaxation time constants, CMR-compatible pressure catheters or sophisticated signal processing may need to be considered (28,29). VEC CMR measurements are susceptible to several potential sources of error, which include turbulent flow and moving valve planes. The CMR method was introduced to the clinical scenario in a small, well-controlled pilot study. Future studies must include a much larger number of patients.

CONCLUSIONS

This work presents an CMR catheterization method for the assessment of diastolic and systolic pressure–volume relation in the left and right ventricle. The applied method combines cine CMR ventricular volumes, real-time VEC CMR blood flow, and invasive ventricular pressure measurements. Our results indicate that the proposed CMR method provides, in addition to parameters of systolic contractile and global pump function, accurate load-independent indexes of biventricular diastolic function. The proposed CMR method might potentially expand the application of pressure–volume relation in the clinical con-

text. Because ventricular volumes are measured by CMR without geometric assumptions, the technique might also be suitable for assessment of right ventricular disorders or congenitally malformed hearts.

Reprint requests and correspondence: Dr. Boris Schmitt, German Heart Institute Berlin, Congenital Heart Disease and Pediatric Cardiology, Augustenburger Platz 1, Berlin 13353, Germany. *E-mail: schmitt@dhzb.de.*

REFERENCES

1. Westermann D, Kasner M, Steendijk P, et al. Role of left ventricular stiffness in heart failure with normal ejection fraction. Circulation 2008;117:2051–60.
2. 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;350:1953–9.
3. Jessup M, Brozena S. Heart failure. N Engl J Med 2003;348:2007–18.
4. Bellenger NG, Davies LC, Francis JM, Coats AJ, Pennell DJ. Reduction in sample size for studies of remodeling in heart failure by the use of cardiovascular magnetic resonance. J Cardiovasc Magn Reson 2000;2:271–8.
5. Danilouchkine MG, Westenberg JJ, de Roos A, Reiber JH, Lelieveldt BP. Operator induced variability in cardiovascular MR: left ventricular measurements and their reproducibility. J Cardiovasc Magn Reson 2005;7:447–57.
6. Muthurangu V, Razavi RS. The value of magnetic resonance guided cardiac catheterisation. Heart 2005;91:995–6.
7. Kuehne T, Yilmaz S, Steendijk P, et al. Magnetic resonance imaging analysis of right ventricular pressure-volume loops: in vivo validation and clinical application in patients with pulmonary hypertension. Circulation 2004;110:2010–6.
8. Kuehne T, Yilmaz S, Schulze-Neick I, et al. Magnetic resonance imaging guided catheterisation for assessment of pulmonary vascular resistance: in vivo validation and clinical application in patients with pulmonary hypertension. Heart 2005;91:1064–9.
9. Burkhoff D, Mirsky I, Suga H. Assessment of systolic and diastolic ventricular properties via pressure-volume analysis: a guide for clinical, translational, and basic researchers. Am J Physiol Heart Circ Physiol 2005;289:H501–12.
10. Mandinov L, Eberli FR, Seiler C, Hess OM. Diastolic heart failure. Cardiovasc Res 2000;45:813–25.
11. Cheung YF, Penny DJ, Redington AN. Serial assessment of left ventricular diastolic function after Fontan procedure. Heart 2000;83:420–4.
12. Akagi T, Benson LN, Gilday DL, et al. Influence of ventricular morphology on diastolic filling performance in double-inlet ventricle after the Fontan procedure. J Am Coll Cardiol 1993;22:1948–52.
13. Brimioulle S, Wauthy P, Ewalenko P, et al. Single-beat estimation of right ventricular end-systolic pressure-volume relationship. Am J Physiol Heart Circ Physiol 2003;284:H1625–30.
14. Steendijk P, Staal E, Jukema JW, Baan J. Hypertonic saline method accurately determines parallel conductance for dual-field conductance catheter. Am J Physiol Heart Circ Physiol 2001;281:H755–63.
15. Krueger JJ, Ewert P, Yilmaz S, et al. Magnetic resonance imaging-guided balloon angioplasty of coarctation of the aorta: a pilot study. Circulation 2006;113:1093–100.
16. Bodhey NK, Beerbaum P, Sarikouch S, et al. Functional analysis of the components of the right ventricle in the setting of tetralogy of Fallot. Circ Cardiovasc Imaging 2008;1:141–7.
17. Korperich H, Gieseke J, Barth P, et al. Flow volume and shunt quantification in pediatric congenital heart disease by real-time magnetic resonance velocity mapping: a validation study. Circulation 2004;109:1987–93.
18. Kuehne T, Saeed M, Reddy G, et al. Sequential magnetic resonance monitoring of pulmonary flow with endovascular stents placed across the pulmonary valve in growing Swine. Circulation 2001;104:2363–8.
19. Steendijk P, Meliga E, Valgimigli M, Ten Cate FJ, Serruys PW. Acute effects of alcohol septal ablation on systolic and diastolic left ventricular function in patients with hypertrophic obstructive cardiomyopathy. Heart 2008;94:1318–22.
20. 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;64:510–5.
21. Karamanoglu M, Bennett TD. A right ventricular pressure waveform based pulse contour cardiac output algorithm in canines. Cardiovasc Eng 2006;6:83–92.
22. 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;29:265–72.
23. Sanchez-Quintana D, Climent V, Ho SY, Anderson RH. Myoarchitecture and connective tissue in hearts with tricuspid atresia. Heart 1999;81:182–91.
24. Gewillig M. The Fontan circulation. Heart 2005;91:839–46.
25. Mooij CF, de Wit CJ, Graham DA, Powell AJ, Geva T. Reproducibility of MRI measurements of right ventricular size and function in patients with normal and dilated ventricles. J Magn Reson Imaging 2008;28:67–73.
26. Oosterhof T, Tulevski, II, Roest AA, et al. 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;7:383–9.
27. Shinozaki T, Deane RS, Mazuzan JE. The dynamic responses of liquid-filled catheter systems for direct measurements of blood pressure. Anesthesiology 1980;53:498–504.
28. Konings MK, Bartels LW, Bakker CJ. Development of an MR-compatible, rotation-insensitive, annular pressure sensor. Phys Med Biol 2001;46:245–59.
29. Lambermont B, Gerard P, Detry O, et al. Correction of pressure waveforms recorded by fluid-filled catheter recording systems: a new method using a transfer equation. Acta Anaesthesiol Scand 1998;42:717–20.

Key Words: cardiac magnetic resonance ■ diagnostic catheterization ■ pressure–volume loops ■ diastolic function.

CLINICAL TRIALS *RECRUITMENT AND RETENTION* *Clinical Trials* 2009; **6**: 597–609

A cost-effectiveness analysis of subject recruitment strategies in the HIPAA era: results from a colorectal cancer screening adherence trial

Paul C Schroy, III[a], Julie T Glick[a], Patricia Robinson[a], Maria A Lydotes[a], Timothy C Heeren[b], Marianne Prout[c], Peter Davidson[a] and John B Wong[d]

Background Changes in regulatory standards that restrict use of identifiable health information can reduce patient recruitment to clinical trials and increase recruitment costs.

Purpose To compare subject accrual rates and costs of three recruitment strategies that comply with new regulatory standards within the context of a clinical trial evaluating the impact of shared decision-making on colorectal cancer screening adherence.

Methods Sequential cohorts of English-speaking, average-risk patients due for colorectal cancer screening were allocated to one of three recruitment strategies: (1) a provider-initiated electronic 'opt-in' referral (*Click*) method; (2) a provider-mediated 'opt-in' referral letter (*Letter*) method; and (3) an investigator-initiated direct contact 'opt-out' (*Call*) method.

Results During distinct 6-month recruitment periods between March 2005 and April 2006, 100 potential subjects were identified using the *Click* method, 847 by the *Letter* method, and 758 by the *Call* method. After excluding ineligible prescreened patients, accrual rates were higher for the *Call* method (188 of 531 [35.4%]) than either the *Click* (12 of 72 [16.7%]; $p = 0.002$) or *Letter* (17 of 816 [2.1%]; $p < 0.001$) methods. The average cost per patient enrolled for the *Call* ($156) method was competitive with the *Click* ($129) and substantially lower than the *Letter* ($1967) methods; the *Call* method was least expensive if combined with automated patient identification ($99). Data extrapolation suggest it would take 2.4 years at an overall cost of $138,518 to recruit a target sample of 900 patients by the *Call* method, 40.5 years at $62,419 for the *Click* method and 27.9 years at $1,737,757 for the *Letter* method.

Limitations The study was nonrandomized and findings may not be generalizable to other research settings.

Conclusion The investigator-initiated direct contact 'opt-out' strategy is significantly more cost-effective and feasible than provider-initiated and provider-mediated 'opt-in' strategies for patient recruitment to clinical trials. *Clinical Trials* 2009; **6**: 597–609. http://ctj.sagepub.com

[a]Department of Medicine, Boston University School of Medicine, Boston, MA, USA, [b]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA, [c]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA, [d]Department of Medicine, Tufts Medical Center, Tufts University School of Medicine, Boston, MA, USA

Author for correspondence: Paul C Schroy III, Boston Medical Center, 85 E. Concord St., Suite 7715, Boston, MA 02118, USA. E-mail: paul.schroy@bmc.org

10.1177/1740774509346703

Introduction

Randomized clinical trials (RCTs) serve as the gold standard for evaluating new screening, diagnostic, and therapeutic interventions. Regardless of the outcome of interest, successful completion of a RCT is contingent upon its ability to recruit a predetermined number of eligible patients within a planned time frame. Failure to enroll the projected sample in a timely manner not only results in costly delays but also compromises a trial's statistical power and scientific validity. Despite its importance, low patient accrual is one of the main reasons for the RCT failure [1–4].

Recent changes in regulatory standards under the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule [5] have been identified as a major barrier to patient recruitment in clinical research. Although the Privacy Rule does not directly regulate human studies research, it limits the ability of 'covered entities' (e.g., healthcare providers and hospitals) to disclose 'protected health information' (e.g., names and contact information) to investigators without a patient's written permission ('authorization') or, in select circumstances, a waiver that permits access without written authorization [6,7]. Consequently, common recruitment strategies employed in the past, such as medical record reviews and use of clinical databases or registries, are no longer permitted in the absence of written authorization, a waiver or use of de-identified data; instead, investigators are now required to identify research patients from within their own clinical practices, obtain referrals from other physicians or recruitment centers, or locate potential patients through advertising [8], all of which pose serious logistical constraints to recruitment and human studies research in general [9–14].

To further protect the rights and welfare of research patients under the Common Rule (U.S. Code of Federal Regulations 45CFR Part 46), many institutional review boards (IRBs) have imposed additional restrictions on recruitment due to concerns about intrusion and coercion. A prevailing sentiment is that the 'opt-out' strategy commonly used in the past is intrusive and possibly unethical, since a nonresponse infers that potential patients can still be contacted regardless of their interest in the study. Instead, many IRBs now mandate an 'opt-in' strategy, whereby potential patients must actively grant permission to be contacted by the research team about a particular study and a nonresponse prohibits further communication. Like the Privacy Rule, use of the opt-in recruitment strategy has been shown to reduce patient enrollment in clinical research and increase the potential for selection bias [15–17].

Researchers outside the United States have encountered similar obstacles to subject recruitment. In the United Kingdom, for example, legislation in the form of the Data Protection Act 1998 and Health and Social Care Act 2001 has imposed strict regulations regarding the use of identifiable health information for the purpose of human studies research [18–21]. Moreover, because of similar concerns about intrusion and coercion, most ethics committees in the United Kingdom have also mandated opt-in recruitment methods, despite evidence suggesting that public concern about an opt-out approach is minimal [22].

Besides compromising patient recruitment and potentially introducing threats to study validity, compliance with these new regulatory standards has been shown to increase personnel costs and resource expenditures [9,11,12,21]. Investigators of federally sponsored RCTs are ill-prepared to compensate for these increases in light of recent declines in National Institutes of Health (NIH) funding. Although a flat NIH budget ($29.5 billion) has again been proposed for fiscal year 2009 [23], actual funding will continue to decrease after adjustment for inflation. Consequently, investigators are placed in the untenable position of trying to optimize patient recruitment without the necessary funds to insure success. Insight into cost-effectiveness of different strategies for overcoming this dilemma would be invaluable. Hence, the primary objective of this study was to evaluate the cost-effectiveness of three different recruitment strategies that comply with new regulatory standards within the context of a clinical trial aimed at assessing the impact of shared decision-making on colorectal cancer screening (CRC) adherence.

Methods

Study objective and design

The overall objective of our RCT is to evaluate the impact of a computer-based decision aid on shared decision-making and patient adherence to CRC screening recommendations. Eligible patients are instructed to arrive one hour before a prearranged office visit with their primary care provider (PCP). After obtaining informed consent and completing a brief pretest, patients are randomized to one of two intervention arms (decision aid plus personalized risk assessment tool with feedback or decision aid alone) and a control arm, each of which involves an interactive computer session. Patients then meet with their providers to discuss screening and identify a preferred screening strategy. Before leaving the clinic, patients are asked to complete a posttest,

which in part assesses screening intentions, knowledge, and satisfaction with the decision-making process. The entire process takes between 30 and 60 min to complete. The target sample size was 900 patients.

Study sites

The study commenced in March 2005 at two urban ambulatory care sites. The first, Boston Medical Center (BMC), is a private, nonprofit academic medical center affiliated with the Boston University School of Medicine, which serves a mostly minority patient population (28% White, non-Hispanic). The second, the South Boston Community Health Center (SBCHC), is a community health center affiliated with BMC, which serves a mostly White, non-Hispanic patient population. Both institutions utilize the same electronic medical record (*Centricity*^{TM} (formerly Logician)). The study protocol and recruitment process were approved by the Boston University Medical Campus Institutional Review Board, which is responsible for overseeing human studies research at both participating institutions.

Patients

The study sample consists of asymptomatic, average-risk patients attending one of the primary care clinics at BMC or SBCHC. Patients are deemed eligible if they are 50–75 years of age, due for CRC screening, and under the care of one the PCPs (physicians or nurse practitioners) who had expressed interest in participating at one of the target sites. Potential subjects meeting any of the following criteria are excluded: (1) prior CRC screening by any method other than fecal occult blood testing; (2) fecal occult blood testing within the past year; (3) high-risk condition (personal history of colorectal cancer or polyps, family history of colorectal cancer or polyps involving one or more first-degree relatives, or chronic inflammatory bowel disease); (4) lack of fluency in written and spoken English; or (5) comorbidities that preclude CRC screening by any recommended method.

Patient identification

A CRC screening module was created within the electronic medical record (EMR) used by primary care clinicians at both participating sites. The module appears in the 'Impression' window of the EMR, which providers generally review in real-time during the clinic visit to prescribe medications, order diagnostic studies, and arrange follow-up. The module itself contains data entry fields that display the dates and results of completed screening tests (fecal occult blood testing, flexible sigmoidoscopy, colonoscopy, or barium enema), which are either downloaded directly from an electronic clinical data repository if performed at BMC beginning in 2002 or entered by the PCP. The module also generates a highly visible 'DUE' flag for patient in need of first-time or follow-up screening.

Patient recruitment

Sequential patient cohorts were allocated to one of three recruitment methods between March 2005 and April 2006: (1) a provider-initiated electronic referral opt-in (*Click*) method; (2) a provider-mediated referral letter opt-in (*Letter*) method; and (3) an investigator-initiated coordinator-mediated direct contact opt-out (*Call*) method.

Click method (March through September 2005)

Potential patients due for screening were initially identified using the EMR's 'DUE' flag at the time of a clinic visit by their provider. To comply with an opt-in recruitment strategy, the CRC screening module was built to include a check-box reminder for providers prompting them to briefly describe the study and obtain 'consent to contact' from interested patients; if checked, the patient's name and medical record number were transmitted electronically to the research team, who would then contact patients by telephone and invite them to participate after verifying eligibility and providing a more detailed overview of the study. Those who agree would then formally enroll in person at their next scheduled appointment with their PCP. To optimize recruitment, a pretrial seminar was conducted with providers to notify them about the study and elicit support. Participating providers were also sent monthly e-mails reminders encouraging them to refer eligible patients.

Letter method (April through October 2005)

Because many potential subjects who expressed interest in participating after being recruited through the 2-visit *Click* method did not have scheduled follow-up appointments at the time of recruitment due to the use of an open access scheduling system, actual enrollment was severely compromised.

To circumvent this problem, the research team was granted approval by the IRB to generate provider-specific lists of patients due for screening with scheduled appointments during the upcoming month. If the treating provider felt that the patient was an appropriate candidate, a standardized referral letter briefly describing the study was prepared by the research team, signed by the provider and mailed to the patient along with an enclosed opt-in return postcard. Upon receipt of the postcard, the research team contacted patients by telephone and invited them to participate after verifying eligibility and providing an overview of the study.

Call method (November 2005 through April 2006)

The *Letter* method also proved to be ineffective due to feasibility issues related to the narrow time interval (1–4 weeks) between patient identification, receipt of the letter, return of an enclosed postcard, and follow-up phone call. To circumvent these issues, approval to use a verbal substitute for the referral letter was requested and granted by the IRB. As with the letter format, individual providers were given a list of their patients due for screening with scheduled appointments for the upcoming month. If deemed appropriate by the PCP, the research team contacted the patients directly by telephone and explored their interest in participating after verifying eligibility and providing an overview of the study. The telephone script explicitly stated that patients were being called to invite them to participate in an educational research study at the recommendation of their primary care provider. Contacted patients had the opportunity to opt-in or -out early in the interview process. Each patient was called up to five times at varying times of the day (morning, afternoon, evening, and weekends); those who could not be contacted were deemed nonenrollees.

Beginning in the fall of 2006, an IRB-approved automated patient identification process was initiated, whereby provider-specific lists of scheduled patients due for screening were generated by an independent senior database analyst at the beginning of each month and provided to the research team (BMC site only). This process eliminated the need for individual electronic medical record reviews by the research team. As with the *Letter* and *Call* methods, the lists were then disseminated to the individual providers for their review and approval. Even though this so-called information technology-assisted (*IT*)-*Call* method was identical to the *Call* method with respect to interactions between the research team and prescreened eligible patients, we felt compelled to incorporate it as a separate recruitment strategy in our cost analyses only to enhance the generalizability of our findings to research settings with well-integrated EMR systems.

Primary and secondary analyses

Accrual rates

The primary outcome was subject accrual rate, defined as the percent of subjects enrolled during the 6-month recruitment period based out of the number of prescreened eligible (i.e., due for screening) patients targeted minus known ineligible subjects. Given this definition, targeted subjects whose eligibility is unknown due to nonresponse were included in the denominator of the accrual rate.

Pairwise comparisons of accrual rates for the three recruitment methods were tested through chi-square analysis. Similar analyses were also performed to compare the demographic characteristics of the targeted samples (excluding known ineligibles) and to compare the demographic characteristics of patients enrolled through the three recruitment methods, acknowledging that the latter comparison lacks adequate power due to the small number of enrolled patients for the *Click* and *Letter* methods. Associations between demographic characteristics and enrollment were examined for the *Call* method only using multiple logistic regression to detect potential recruitment biases. Significance was defined at the $p \leq 0.05$ level. All statistical analyses were performed using SAS (version 8.2, Cary, North Carolina).

Cost

To estimate costs, a detailed bottom-up microcosting approach was applied to each recruitment method. Personnel costs were determined by first identifying the various activities performed by different staff (study coordinators, computer programmer, and database analyst) necessary for each method and then estimating the mean amount of time required for completion of these activities based on a minimum of 10 measurements; mean and high/low cost estimates were then calculated by multiplying the mean and range of time to complete each activity by average hourly staff salaries. The costs of consumable items (mailings, calls) were derived by multiplying per item costs by item number. To extrapolate results, costs were divided into one-time fixed development costs and variable costs, i.e., costs that directly increase with higher volume. Examples of one-time costs include developing patient or physician contact materials, such as letters, programming, and database costs for

linking electronic medical records with endoscopy reports. Examples of variable costs include the medical record review, phone calls, letters, or monthly computer reports. Personnel costs related to updating screening data in the EMR, primary care provider time and ancillary costs for developing and maintaining the hospital-based EMR, clinical data repository, and endoscopy database were excluded. Relevant cost estimates are provided in the Table 1. Using beta distributions for individuals contacted and enrolled and normal distributions for costs, the extrapolations transformed 6-month accruals into constant 1-month event rates and costs and extrapolated accruals for alternative time horizons. Monte Carlo bootstrap samples involving 10,000 simulations were performed with Decision Maker (Physicians of Tufts Medical Center, Boston, MA).

Table 1 Cost estimates for each recruitment method

Method	Item	Type	Mean (range) Time, h[a]	Item Cost
Click	Personnel time			
	Meetings to develop EMR CRC screening module[b]	One-time fixed	6.0 (5.0–7.0)	$19.18/h[c]
	Computer programming	One-time fixed	12 (10–14)	$38.00/h
	Generate monthly e-mail reminders to providers	Variable	1.50 (1.25–1.75)	$19.18/h[c]
	Generate bi-monthly reports of consenting patients	Variable	0.07 (0.05–0.08)	
	Review bi-monthly reports of consenting patients	Variable	1.0 (0.5–1.5)	
	Telephone calls: pitch and explain study	Variable	0.10 (0.08–0.17)	
	Telephone calls: no answer (no voicemail)	Variable	0.02 (0.01–0.02)	
	Telephone calls: leave message	Variable	0.05 (0.03–0.07)	
	Telephone calls: documentation	Variable	0.01 (0.01–0.01)	
	S			
	Telephone calls	Variable		$0.015
Letter	Personnel time			
	Draft and revise letter	One-time fixed	3.0 (1.5–4.5)	$19.18/h[c]
	Draft postcard	One-time fixed	0.75 (0.5–1.0)	
	Perform weekly chart reviews	Variable	60 (45–75)	
	Generate each letter	Variable	0.01 (0.01–0.02)	
	Obtain provider signatures (weekly)	Variable	6.0 (4.0–8.0)	
	Telephone calls: pitch and explain study	Variable	0.1 (0.08–0.17)	
	Telephone calls: no answer (no voicemail)	Variable	0.02 (0.01–0.02)	
	Telephone calls: leave message if no answer	Variable	0.05 (0.03–0.07)	
	Telephone calls: documentation	Variable	0.01 (0.01–0.01)	
	S			
	Letterhead, envelope, and postcard with postage	Variable		$0.79
	Telephone calls	Variable		$0.015
Call	Personnel time			
	Perform weekly chart reviews	Variable	60 (45–75)	$19.18/h[c]
	Telephone calls: pitch and explain study	Variable	0.1 (0.08–0.17)	
	Telephone calls: no answer (no voicemail)	Variable	0.02 (0.01–0.02)	
	Telephone calls: leave message if no answer	Variable	0.05 (0.03–0.07)	
	Telephone calls: documentation	Variable	0.01 (0.01–0.01)	
	S			
	Telephone calls	Variable		$0.015
IT-Call	Personnel time			
	Meetings to developing advanced chart audit[b]	One-time fixed	40 (30–60)	$19.18/h[c]
	Computer programming	One-time fixed	80 (60–100)	$38.00/h
	Generate monthly reports of 'due' patients	Variable	0.8 (0.07–0.10)	$38.00h
	Review monthly reports and e-mail providers	Variable	24 (18–30)	$19.18/h[c]
	Telephone calls: pitch and explain study	Variable	0.1(0.08–0.17)	
	Telephone calls: no answer (no voicemail)	Variable	0.02 (0.01–0.02)	
	Telephone calls: leave message if no answer	Variable	0.05 (0.03–0.07)	
	Telephone calls: documentation	Variable	0.01 (0.01–0.01)	
	S			
	Telephone calls	Variable		$0.015

EMR, electronic medical record; CRC, colorectal cancer.[a]Values, expressed in hours rounded to second decimal place, reflect total time for fixed one-time costs and per item time for variable costs.[b]Study coordinator time; computer analyst time included in estimate of computer programming time.[c]Mean hourly salary for the three study coordinators (range, $17.43/h to $22.60/h).

Results

Patient recruitment

Figure 1 summarizes patient flow for each recruitment method. During the defined 6-month recruitment periods, 100 potential subjects (i.e., patients due for CRC screening'') were identified using the *Click* method, 847 by the *Letter* method, and 758 by the *Call* method. After excluding ineligible prescreened patients, accrual rates were higher for the *Call* method (188 of 531 [35.4%]) than either the *Click* (12 of 72 [16.7%]; $p = 0.002$) or *Letter* (17 of 816 [2.1%]; $p < 0.001$) methods. The actual proportion of potential subjects who were successfully contacted, deemed eligible, and enrolled was relatively high for all three methods (12 of 18 [67%] for the *Click* method; 17 of 17 [100%] for the *Letter* method; 188 of 238 [79%] for the *Call* method), suggesting that direct contact with the patient was the critical determinant of successful recruitment and not the acquisition of more effective recruiting skills by the research team over time.

Characteristics of enrolled and nonenrolled patients for each recruitment method

Table 2 depicts the baseline characteristics of eligible patients for each recruitment method. Despite the use of a nonrandomized patient allocation scheme, there were no significant differences between the three groups with respect to age, race/ethnicity, sex, marital status, and employment status. There was, however, a difference with respect to insurance coverage. The target samples were predominantly older (≥age 65), non-White, and female with public or private medical insurance; a relative minority were married or employed. These demographic characteristics are similar to those of the participating sites, except for the relative predominance of older individuals (data not shown).

Table 3 depicts the demographic characteristics of enrolled patients for each recruitment method. No significant differences were observed between the three groups with respect to age, race/ethnicity, gender, marital status, insurance coverage, or employment status.

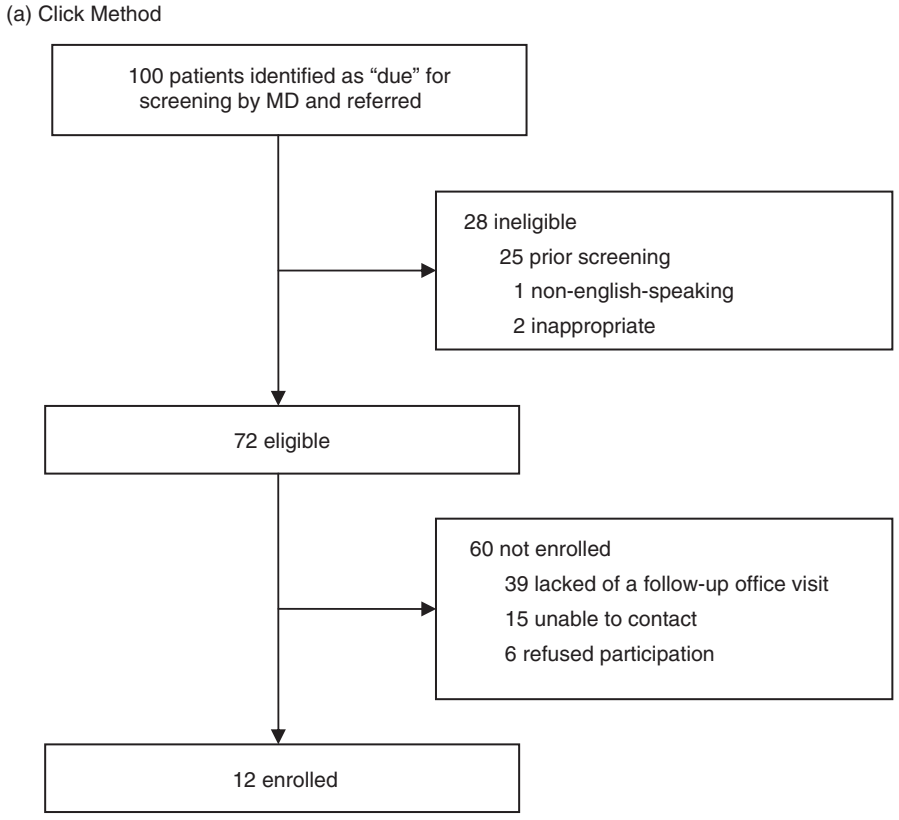


Figure 1 Flow of trial participants for each recruitment strategy. (a) *Click* method; (b) *Letter* method; (c) *Call* method. 'Due' indicates that a patient was not current with any colorectal cancer screening recommendations based on review of institutional electronic data repositories

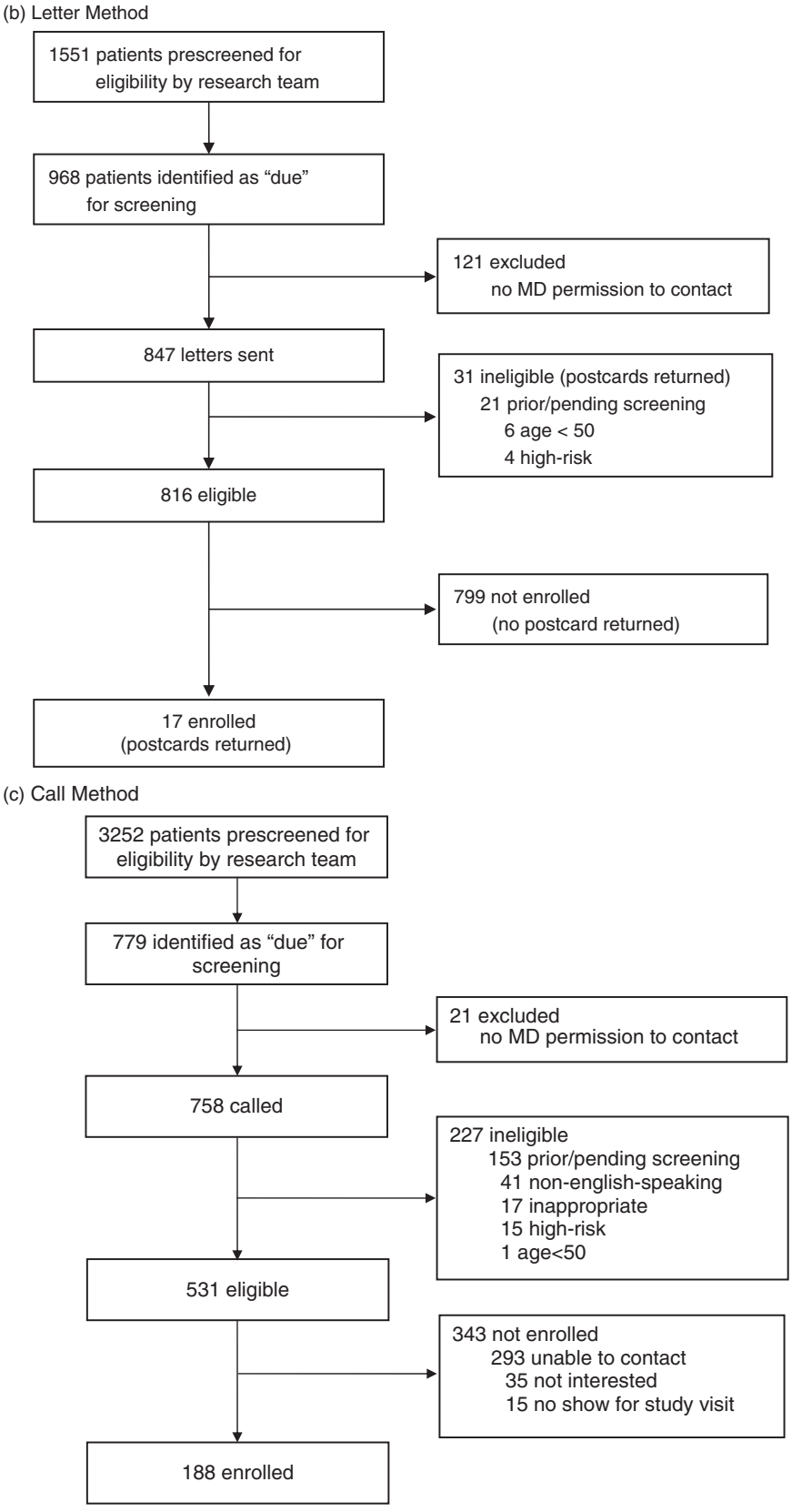


Figure 1 Continued

Table 2 Comparison of characteristics of targeted samples (enrolled and nonenrolled patients) for each recruitment method[a]

Characteristic	*Click* ($n = 72$)	*Letter* ($n = 816$)	*Call* ($n = 531$)	*P*-Value[b]
Age				0.06
<65y	73.6	69.3	75.1	
≥65y	26.4	30.8	24.9	
Race/Ethnicity				0.49
Black	65.3	69.3	54.2	
White	20.8	35.8	29.9	
Hispanic	4.2	6.4	7.5	
Other	9.7	10.7	8.3	
Sex				0.45
Female	54.2	58.8	55.7	
Male	45.8	41.2	44.3	
Marital status				0.22
Single	40.3	36.0	42.4	
Married	31.9	36.5	33.2	
Other	27.8	27.5	24.5	
Insurance				0.02
Private	44.4	40.1	37.7	
Medicare	33.3	33.6	27.9	
Medicaid	16.7	17.8	21.9	
Free care	4.2	7.8	10.2	
No coverage	1.4	0.7	2.5	
Employment				0.43
Full-time	19.7	19.2	19.9	
Part-time	8.5	7.1	9.4	
Unemployed	36.6	34.4	38.2	
Disabled	9.9	12.6	10.3	
Retired	23.9	22.3	16.6	
Other	1.4	2.6	3.4	

[a]Data expressed as percent of each group.[b]Chi-square analysis.

Table 3 Comparison of characteristics of enrolled samples for each recruitment method[a]

Characteristic	*Click* ($n = 12$)	*Letter* ($n = 17$)	*Call* ($n = 188$)	*P*-Value[b]
Age				0.98
<65y	75.0	76.5	77.7	
≥65y	25.0	24.5	22.3	
Race/Ethnicity				0.12
Black	58.3	35.3	58.0	
White	25.0	58.8	29.3	
Hispanic	0.0	5.9	6.9	
Other	16.7	0.0	5.9	
Sex				0.84
Female	58.3	52.9	60.1	
Male	41.7	47.1	39.9	
Marital status				0.39
Single	33.3	47.1	50.5	
Married	33.3	11.8	25.0	
Other	33.3	41.2	24.5	
Insurance				0.89
Private	33.3	41.2	37.2	
Medicare	50.0	41.2	32.5	
Medicaid	16.7	11.8	19.7	
Free care	0.0	5.9	9.6	
No coverage	0.0	0.0	1.1	
Employment				0.10
Full-time	0.0	6.2	19.7	
Part-time	8.3	0.0	12.2	
Unemployed	33.3	37.5	33.5	
Disabled	8.3	18.7	12.2	
Retired	50.0	25.0	14.9	
Other	0.0	12.5	4.3	

[a]Data expressed as percent of each group.[b]Chi-square analysis.

Table 4 depicts the results of univariate and multivariate analyses comparing enrolled and nonenrolled patients for the *Call* method only. The two groups were similar with respect to age, race/ethnicity, sex, insurance coverage, and employment status but differed with respect to marital status. Married patients were significantly less likely to enroll than single patients (unadjusted and adjusted O.R. [95%CI], 0.50 [0.32–0.76], and 0.53 [0.34–0.82], respectively). No significant differences were observed between enrolled and nonenrolled patients for both the *Click* and *Letter* methods, but the relatively small number of enrolled patients for each method precludes meaningful statistical analyses (data not shown).

Cost versus yield of different recruitment methods

Table 5 displays the cost and accrual rates for the different recruitment methods for three alternative time frames: 6, 12, and 24 months. For the purpose of these analyses, the *Call* method was subdivided into *Call* and information technology (IT)-assisted *Call* (*IT-Call*) methods to account for the different costs associated with the laborious electronic medical record review approach used by the study coordinators to identify potential subjects compared with the more efficient, automated method employed by the data analyst. Because of amortization, strategies with higher one-time costs (*Click* and *IT-Call* methods) have more dramatic reductions in average per-patient costs to screen for eligibility, contact (successfully or unsuccessfully) and enroll when the time period of interest increases from 6 to 24 months. The other strategies with relatively low one-time development costs have stable average per-patient costs to screen, contact and enroll, regardless of the time period of interest. During the initial 6-month period, the most labor-intensive recruitment strategy, the *Letter* method, had the highest overall cost ($31,480) yet lowest yield and hence the highest average cost per patient screened ($20), contacted ($37), and enrolled ($1967). The *Click* method, on the other hand, had the lowest overall cost ($1423) but only

Table 4 Characteristics associated with enrollment, Call method ($n = 531$)

Characteristic	% patients enrolled	Unadjusted OR	95% CI	Adjusted OR	95% CI
Age					
<65y	36.6	1.00	–	1.00	–
≥65y	31.8	0.82	0.54, 1.25	0.71	0.40, 1.2
Race/Ethnicity					
Black	37.9	1.14	0.76, 1.72	1.10	0.72, 1.70
White	34.6	1.00	–	1.00	–
Hispanic	32.5	0.92	0.44, 1.93	0.90	0.41, 1.96
Other	25.0	0.61	0.29, 1.31	0.63	0.28, 1.43
Sex					
Female	31.9	1.00	–	1.00	–
Male	38.2	1.33	0.93, 1.91	1.41	0.96, 2.07
Marital status					
Single	42.2	1.00	–	1.00	–
Married	26.7	0.50	0.32, 0.76	0.53	0.34, 0.82
Other	35.4	0.76	0.49, 1.19	0.75	0.47, 1.21
Insurance					
Private	35.0	1.00		1.00	–
Medicare	41.2	1.31	–	1.84	0.96, 3.51
Medicaid	31.9	0.84	0.84, 2.03	0.96	0.54, 1.73
Free care	33.3	0.93	0.52, 1.37	0.86	0.42, 1.74
No coverage	15.4	0.33	0.49, 1.75	0.31	0.06, 1.48
Employment					
Full-time	35.6	1.00	–	1.00	–
Part-time	46.9	1.60	0.80, 3.19	1.47	0.70, 3.09
Unemployed	31.5	0.83	0.50, 1.37	0.74	0.40, 1.35
Disabled	42.6	1.34	0.67, 2.63	0.79	0.33, 1.88
Retired	32.2	0.86	0.47, 1.57	0.70	0.31, 1.58
Other	44.4	1.45	0.53, 3.99	1.76	0.60, 5.19

OR, odds ratio; CI, confidence intervals. Multiple logistic regression controlling for other factors listed in the table.

the second lowest average cost per patient contacted ($14) and enrolled ($129) due to its low yield. The *Call* method was the second most personnel intensive recruitment strategy and as such had the second highest overall cost ($29,288); however, because of its superior yield, the average cost per patient screened ($9), contacted ($39), and enrolled ($156) was only marginally higher than the *Click* method. Finally, the *IT-Call* method had the second lowest overall cost ($9548) but lowest average cost per patient screened ($0.93), contacted ($7), and enrolled ($99), due to a higher absolute albeit lower proportional yield compared to the *Click* method. It is noteworthy that the target population for the *IT-Call* method was comprised of a much higher proportion of patients who had been unsuccessfully recruited using the *Call* method in the past mostly due to 'failure to contact,' thus suggesting that differences in patient mix rather than methodological differences were responsible for the lower enrollment rates (data not shown).

Typically, clinical trials must recruit a pre-specified number of patients based on power analysis. Table 6 displays the recruitment time and costs associated with the four alternative strategies to recruit 100, 300, 500 and our target sample of 900 patients, assuming a stable pool of eligible patients and a steady rate of enrollment. The *Click* method always has the lowest cost but always takes the longest to fulfill targeted patient recruitment goals. The *Letter* method is always the most expensive and requires nearly many years as the *Click* method. The *Call* method has higher costs than either the *Click* or *IT-Call* method (provided electronic health records are available) but takes the least time to fulfill recruitment. Thus, the higher recruitment cost may be offset by a shorter time to study completion because of higher proportion enrollment. We estimate that it would take 2.4 years at an overall cost of $138,518 to recruit our target sample of 900 patients by the *Call* method, 4.6 years at a cost of $56,520 for the *IT-Call* method, 40.5 years at a cost of $62,419 for the *Click* method, and 27.9 years at a cost of $1,737,757 for the *Letter* method. For previously cited reasons, however, we speculate that the recruitment time for the *IT-Call* method would

Table 5 Cost and accrual rate results

	Overall costs (95% CI)	Number of individuals			Average total cost per individual		
		Screened *N*	Contacted[a] (95% CI)	Enrolled (95% CI)	Screened (95% CI)	Contacted[a] (95% CI)	Enrolled (95% CI)
6 Months							
Letter	$31,480 (23,463–39,413)	1551	847 (808–885)	17 (10–26)	$20 (15–25)	$37 (29–47)	$1967 (1126–3330)
Click	$1423 (1193–1652)	–	100 (83–119)	12 (6–20)	–	$14 (11–18)	$129 (71–229)
Call	$29,288 (21,893–36,742)	3252	758 (712–805)	188 (163–215)	$9 (7–11)	$39 (29–49)	$156 (113–204)
IT-Call	$9548 (6384–12,264)	10,260	1344 (1278–1413)	98 (80–118)	$0.93 (0.67–1.20)	$7 (5–9)	$99 (66–136)
12 Months							
Letter	$62,964 (51,859–73,814)	3102	1694 (1573–1820)	34 (24–46)	$20 (17–24)	$37 (30–44)	$1909 (1279–2775)
Click	$2841 (2518–3159)	1530	201 (162–242)	24 (15–35)	$1.86 (1.65–2.07)	$14 (12–17)	$124 (81–186)
Call	$58,554 (47,797–68,919)	6504	1517 (1405–1629)	376 (340–413)	$9 (7–11)	$39 (32–46)	$156 (125–189)
IT-Call	$19,132 (15,319–22,995)	20,520	2689 (2544–2840)	196 (169–225)	$0.93 (0.75–1.12)	$7 (6–9)	$98 (75–124)
24 Months							
Letter	$125,862 (109,830–141,571)	6204	3388 (3309–3465)	68 (53–85)	$20 (18–23)	$37 (32–42)	$1878 (1434–2458)
Click	$5687 (5228–6136)	3060	400 (364–437)	48 (36–62)	$1.85 (1.71–2.01)	$14 (13–16)	$121 (91–161)
Call	$117,309 (102,206–132,137)	13,008	3032 (2937–3127)	752 (702–805)	$9 (8–10)	$39 (34–44)	$156 (134–179)
IT-Call	$38,217 (32,641–43,649)	41,040	5377 (5244–5511)	392 (354–431)	$0.93 (0.80–1.07)	$7 (6–8)	$98 (81–115)

[a]Contacted refers to number of prescreened eligible patients called or sent letters, regardless of whether direct contact was made.

have been similar to that of the *Call* method if the same patient population has been targeted.

Discussion

The primary objective of this study was to evaluate the effectiveness, as defined by patient accrual rates, and costs associated with three different recruitment methods. We found that the investigator-initiated direct contact opt-out (*Call*) method yields substantially higher accrual rates than either the provider-initiated electronic referral (*Click*) or provider-mediated referral letter (*Letter*) opt-in methods. We also found that the average cost per enrolled patient associated with the *Call* method was competitive with the *Click* method and substantially lower than the *Letter* method; the *Call* method was actually the least expensive if combined with an automated EMR-based patient identification strategy (*IT-Call*). Moreover, as recruitment demand and duration of the study increased from 6 to 24 months, the initial upfront cost investment in *IT-Call* resulted in substantial long-term savings when compared with *Letter* or *Call*.

Our findings corroborate those of prior studies demonstrating higher recruitment rates with direct investigator-mediated opt-out strategies compared with provider-mediated opt-in strategies [19, 20, 24, 25]. The relative success of the *Call* method may be attributable to a number of factors, most notably the ability to deliver a more intense and personalized recruitment message, answer questions, and dissuade any fears or concerns. Of the 426 patients contacted directly, none expressed concerns about a violation of privacy. Moreover, the refusal rates were similar for eligible patients contacted either after granting consent to contact by the *Click* method or contacted directly by investigators by the *Call* method (11% vs. 15%, respectively; $p = 0.37$). Thus, our study provides additional evidence that most patients do not perceive the investigator-initiated direct contact approach to be a violation of privacy or result in a loss of personal autonomy [26–29].

Apart from our findings related to the effectiveness of the different recruitment strategies, our cost-effectiveness analyses provide a critical perspective on the feasibility of each strategy. The *Letter* method remains economically the least efficient and suffers from low enrollment, presumably due to both logistical constraints imposed by our open access scheduling system and patient factors (e.g., lack of understanding about the study, disinterest, and distrust). While the *Click* method is comparatively inexpensive per patient enrolled,

Table 6 Time and cost to reach recruitment targets

Recruitment target	Letter	Click	Call	IT-Call
***N* = 100**				
Years to reach (95% CI)	3.1 (1.8–5.0)	4.5 (2.5–8.1)	0.23 (0.17–0.25)	0.47 (0.42–0.58)
Cost (95% CI)	$192,204 (107,837–327,898)	$8162 (4544–14,743)	$13,639 (8342–18,121)	$9218 (6404–12,345)
***N* = 250**				
Years to reach (95% CI)	7.7 (4.8–12.4)	11.2 (6.3–19.9)	0.62 (0.50–0.75)	1.2 (1.0–1.5)
Cost (95% CI)	$483,886 (271,083–820,818)	$19,687 (10,593–35,702)	$36,664 (26,267–48,459)	$18,016 (12,592–24,146)
***N* = 500**				
Years to reach (95% CI)	15.5 (9.6–25.1)	22.4 (12.6–39.4)	1.3 (1.1–1.5)	2.5 (2.1–3.1)
Cost (95% CI)	$969,133 (545,280–1,639,591)	$38,707 (20,514–70,570)	$75,648 (54,691–99,596)	$32,694 (22,351–44,249)
***N* = 900**				
Years to reach (95% CI)	27.9 (17.3–44.4)	40.5 (22.5–72.1)	2.4 (2.0–2.8)	4.6 (3.8–5.6)
Cost (95% CI)	$1,737,757 (985,818–2,894,467)	$62,419 (36,268–127,543)	$138,518 (100,298–181,098)	$56,520 (38,263–77,060)

accrual relies on PCP activation and would have necessitated an estimated 37.5 years to reach the enrollment target. Efforts to increase PCP activation through monthly e-mail reminders were ineffective. The extent to which incentive-laden enticements may have been more successful, however, is unknown. The *Call* method is clearly the most feasible recruitment strategy, requiring only 2.4 years to accrue the enrollment target, at a cost that compares favorably with that of the *Click* method if combined with IT support. Even in the absence of IT support, the *Call* method is more cost-effective due to substantially higher accrual rates.

Another important observation of our study was that the *Call* method afforded minimal selection bias. Prior studies have clearly demonstrated that both the Privacy Rule and mandatory opt-in recruitment policies not only challenge the feasibility of recruiting patients to clinical research, but also introduce so-called 'consent bias' or 'authorization bias' that compromises external validity [15–17]. By circumventing these requirements, we found that eligible patients enrolled in our study using the *Call* method were similar to those who did not enroll, with respect to all characteristics examined, except marital status. Although we also found no evidence of selection bias for either the *Click* or *Letter* methods, too few patients were enrolled to assess accurately their impact on selection bias.

The extent to which each of the three recruitment strategies employed in this study complied with current regulatory requirements is of paramount importance to the overall objectives of the study. The *Click* method was in strict compliance with both the Privacy Rule and the Common Rule, because it utilized treating providers to both identify potential patients from within their own practices and obtain a 'consent to contact' via an opt-in approach. Although abandoned primarily because of logistical constraints imposed by an open access scheduling system, this approach was largely ineffective as a recruitment strategy, thus corroborating a prior observation that the use of PCPs to recruit patients to clinical trials during a routine office visit is feasible but relatively ineffective, time-consuming, and impractical [30]. The *Letter* method was also acceptable to the extent that treating providers rather than the research team contacted potential subjects and facilitated recruitment using an opt-in referral letter process. Although less disruptive and less time-consuming than the *Click* method, this passive provider-mediated approach was also ineffective. The *Call* method was the most contentious strategy, requiring approval from the IRB to contact potential subjects directly using an opt-out approach. Approval was sought and granted on grounds that the phone call served as a verbal substitute for the letter because the treating provider was identified as the referral source in the introductory statement, the risk to loss of privacy was minimal, the benefits of the study far outweighed the risks, and patient recruitment using both the *Click* and *Letter* methods was sufficiently low to suggest that the research could not otherwise be practically conducted. The degree to which other IRBs or Privacy Boards would act similarly and approve an investigator-initiated direct contact opt-out approach under these circumstances is unknown [31].

Our study has several notable strengths. First, we systematically compared the relative costs and effectiveness of three different recruitment strategies targeting the same patient population. Second, the target population was comprised of a large minority population, thus enhancing the external

validity of our findings and potentially assisting other researchers attempting to bolster recruitment of this hard to reach group. Third, our study highlights the utility of electronic medical records for facilitating patient identification, recruitment and communication between treating providers and the research team [32].

Our study also has several noteworthy limitations. The first relates to the generalizability of our findings to other settings that may vary with respect to the availability of an EMR for identifying eligible patients, the level of motivation by primary care providers to assist in patient identification and/or recruitment, use of an open access system for patient scheduling, personnel salaries and the policies and procedures of other IRBs or Privacy Boards regarding opt-in versus opt-out recruitment strategies. It also remains unclear whether our findings have external validity with respect to other types of clinical trials where the relative risks and benefits of participation may strongly influence patient and provider interest, as well as IRB considerations. The second major limitation relates to our study's nonrandomized design. Although we demonstrate that the majority of baseline characteristics of eligible patients in each of the three groups were similar, it remains possible that the three groups may have differed with respect to unmeasured patient factors. A third limitation relates to use of the telephone to contact directly potential patients. Although our response rate was relatively high (63%), the widespread use of answering machines and caller ID have been shown to compromise participation rates and increase costs in other research settings [33]. Lastly, given our study design, it is possible that differential exposure to heightened publicity related to CRC screening during one of the time intervals may have influenced our results.

In conclusion, our study provides compelling evidence supporting the superior cost-effectiveness of the investigator-initiated direct contact opt-out approach over provider-initiated or provider-mediated opt-in strategies for patient recruitment to clinical trials. Moreover, our study finds little evidence supporting the widespread belief that patients perceive the direct contact opt-out approach to be a violation of privacy or result in a loss of personal autonomy. Future studies are needed to better define the generalizability of our findings to other types of clinical trials targeting distinct patient populations in diverse health care settings. In the interim, however, we encourage other IRBs or Privacy Boards to adopt more lenient policies regarding use of investigator-initiated direct contact opt-in recruitment strategies for low-risk clinical trials where the potential benefits of the research far exceed the risks of loss of confidentiality.

Acknowledgments

We thank Linda Rosen, MSEE, from the Office of Clinical Research, Boston University School of Medicine and Daniel Newman, MD, from the Department of Medicine, Boston Medical Center for their technical support; Suzette Levenson, MPH, and Jingshun Yang, MS, from the Data Coordinating Center, Boston University School of Public Health for data management services; and Richard Saitz, MD, MPH, from the Department of Medicine, Boston University School of Medicine for his advice and critical review of our manuscript.

Grant support: Funding was provided Agency for Healthcare Research and Quality grant RO1HS013912 (P.C.S.).

References

1. **Lovato LC, Hill K, Hertert S, *et al*.** Recruitment for controlled clinical trials: literature summary and annotated bibliography. *Control Clin Trials* 1997; **18**: 328–52.
2. **Ross S, Grant A, Counsell C, *et al*.** Barriers to participation in randomised controlled trials: a systematic review. *J Clin Epidemiol* 1999; **52**: 1143–56.
3. **Haidich AB, Ioannidis JP.** Effect of early patient enrollment on the time to completion and publication of randomized controlled trials. *Am J Epidemiol* 2001; **154**: 873–80.
4. **Haidich AB, Ioannidis JP.** Patterns of patient enrollment in randomized controlled trials. *J Clin Epidemiol* 2001; **54**: 877–83.
5. **U.S. Department of Health and Human Services.** Standards for privacy of individually identifiable health information, Final Rule (Privacy Rule). Code of Federal Regulations, Title 45, Parts 160 and 164. Federal Register, 2002, Vol. 67, pp. 53181-273.
6. **Nosowsky R, Giordano TJ.** The Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rule: implications for clinical research. *Annu Rev Med* 2006; **57**: 575–90.
7. **Gunn PP, Fremont AM, Bottrell M, *et al*.** The Health Insurance Portability and Accountability Act Privacy Rule: a practical guide for researchers. *Med Care* 2004; **42**: 321–27.
8. **Holt E.** The HIPAA Privacy Rule, Research, and IRBs. Applied Clinial Trials, June 2003. Available at: http://actmagazine.findpharma.com/appliedclinicaltrials/Feature±Article/The-HIPAA-Privacy-Rule-Research-and-IRBs/ArticleStandard/Article/detail/80209 (accessed 31 January 2008).
9. **Tu JV, Willison DJ, Silver FL, *et al*.** Impracticability of informed consent in the Registry of the Canadian Stroke Network. *N Engl J Med* 2004; **350**: 1414–21.
10. **Ness RB.** A year is a terrible thing to waste: early experience with HIPAA. *Ann Epidemiol* 2005; **15**: 85–86.
11. **Armstrong D, Kline-Rogers E, Jani SM, *et al*.** Potential impact of the HIPAA privacy rule on data collection in a registry of patients with acute coronary syndrome. *Arch Intern Med* 2005; **165**: 1125–29.

12. **Wolf MS, Bennett CL.** Local perspective of the impact of the HIPAA privacy rule on research. *Cancer* 2006; **106**: 474–79.
13. **Krousel-Wood M, Muntner P, Jannu A, *et al*.** Does waiver of written informed consent from the institutional review board affect response rate in a low-risk research study? *J Investig Med* 2006; **54**: 174–79.
14. **Ness RB.** Influence of the HIPAA Privacy Rule on health research. *JAMA* 2007; **298**: 2164–70.
15. **Woolf SH, Rothemich SF, Johnson RE, Marsland DW.** Selection bias from requiring patients to give consent to examine data for health services research. *Arch Fam Med* 2000; **9**: 1111–18.
16. **Junghans C, Feder G, Hemingway H, *et al*.** Recruiting patients to medical research: double blind randomised trial of ''opt-in'' versus ''opt-out'' strategies. *BMJ* 2005; **331**: 940.
17. **Al-Shahi R, Vousden C, Warlow C.** Bias from requiring explicit consent from all participants in observational research: prospective, population based study. *BMJ* 2005; **331**: 942.
18. **Strobl J, Cave E, Walley T.** Data protection legislation: interpretation and barriers to research. *BMJ* 2000; **321**: 890–92.
19. **Al-Shahi R, Warlow C.** Using patient-identifiable data for observational research and audit. *BMJ* 2000; **321**: 1031–32.
20. **Ward HJ, Cousens SN, Smith-Bathgate B, *et al*.** Obstacles to conducting epidemiological research in the UK general population. *BMJ* 2004; **329**: 277–79.
21. **Metcalfe C, Martin RM, Noble S, *et al*.** Low risk research using routinely collected identifiable health information without informed consent: encounters with the Patient Information Advisory Group. *J Med Ethics* 2008; **34**: 37–40.
22. **Hewison J, Haines A.** Overcoming barriers to recruitment in health research. *BMJ* 2006; **333**: 300–02.
23. **Office of Management and Budget.** Budget of the United States Government. Fiscal year 2009. Avaiable at: http://www.whitehouse.gov/omb/budget/fy2009/hhs.html. (accessed 6 February 2008).
24. **Foy R, Parry J, Duggan A, *et al*.** How evidence based are recruitment strategies to randomized controlled trials in primary care? Experience from seven studies. *Fam Pract* 2003; **20**: 83–92.
25. **Beskow LM, Sandler RS, Weinberger M.** Research recruitment through US central cancer registries: balancing privacy and scientific issues. *Am J Public Health* 2006; **96**: 1920–26.
26. **Sugarman J, Regan K, Parker B, *et al*.** Ethical ramifications of alternative means of recruiting research participants from cancer registries. *Cancer* 1999; **86**: 647–51.
27. **Beskow LM, Sandler RS, Millikan RC, Weinberger M.** Patient perspectives on research recruitment through cancer registries. *Cancer Causes Control* 2005; **16**: 1171–75.
28. **Beskow LM, Millikan RC, Sandler RS, *et al*.** The effect of physician permission versus notification on research recruitment through cancer registries (United States). *Cancer Causes Control* 2006; **17**: 315–23.
29. **Inskip HM, Godfrey KM, Robinson SM, *et al*.** Cohort profile: The Southampton Women's Survey. *Int J Epidemiol* 2006; **35**: 42–48.
30. **Mosis G, Dieleman JP, Stricker B, *et al*.** A randomized database study in general practice yielded quality data but patient recruitment in routine consultation was not practical. *J Clin Epidemiol* 2006; **59**: 497–502.
31. **Dziak K, Anderson R, Sevick MA, *et al*.** Variations among Institutional Review Board reviews in a multisite health services research study. *Health Serv Res* 2005; **40**: 279–90.
32. **Pakhomov S, Weston SA, Jacobsen SJ, *et al*.** Electronic medical records for clinical research: application to the identification of heart failure. *Am J Manag Care* 2007; **13**: 281–88.
33. **Kempf AM, Remington PL.** New challenges for telephone survey research in the twenty-first century. *Annu Rev Public Health* 2007; **28**: 113–26.

American Journal of Epidemiology
Vol. 170, No. 12
DOI: 10.1093/aje/kwp309
Advance Access publication:
November 12, 2009

Practice of Epidemiology

A Family Longevity Selection Score: Ranking Sibships by Their Longevity, Size, and Availability for Study

Paola Sebastiani, Evan C. Hadley*, Michael Province, Kaare Christensen, Winifred Rossi, Thomas T. Perls, and Arlene S. Ash

* Correspondence to Dr. Evan Hadley, Gateway Building, MSC 9205, National Institute on Aging, 7201 Wisconsin Avenue, Bethesda, MD 20892-9205 (e-mail: evan.hadley@nih.gov).

Initially submitted February 10, 2009; accepted for publication September 1, 2009.

Family studies of exceptional longevity can potentially identify genetic and other factors contributing to long life and healthy aging. Although such studies seek families that are exceptionally long lived, they also need living members who can provide DNA and phenotype information. On the basis of these considerations, the authors developed a metric to rank families for selection into a family study of longevity. Their measure, the family longevity selection score (FLoSS), is the sum of 2 components: 1) an estimated family longevity score built from birth-, gender-, and nation-specific cohort survival probabilities and 2) a bonus for older living siblings. The authors examined properties of FLoSS-based family rankings by using data from 3 ongoing studies: the New England Centenarian Study, the Framingham Heart Study, and screenees for the Long Life Family Study. FLoSS-based selection yields families with exceptional longevity, satisfactory sibship sizes and numbers of living siblings, and high ages. Parameters in the FLoSS formula can be tailored for studies of specific populations or age ranges or with different conditions. The first component of the FLoSS also provides a conceptually sound survival measure to characterize exceptional longevity in individuals or families in various types of studies and correlates well with later-observed longevity.

aged, 80 and over; family data; longevity; Shannon information

Abbreviations: FLoSS, family longevity selection score; FRS, family risk score.

Exceptional longevity strongly aggregates in families (1–4). Thus, studies of exceptionally long-lived families have the potential to identify genetic variants and other factors contributing to longevity, particularly factors that are too rare to detect in a population-based study (5, 6). Both a family's exceptionality of survival and its number of living siblings contribute to its value for a genetic study of longevity. A challenge in selecting families for such studies is that the most long-lived individuals may have few or no living siblings to provide DNA and other biologic material. To select families for a study of environmental and genetic factors contributing to longevity, we developed the family longevity selection score (FLoSS). The FLoSS combines a measure of the exceptionality of a family's survival with a bonus for the presence of living old members; it is computed from information that can be collected relatively easily from an immediate family member.

We compare the FLoSS with an alternative scoring method, the family risk score (FRS). The FRS generalizes the family history score (7) to quantitative phenotypes, creating an average score among family members adjusted for the distribution of the phenotype in the general population (8, 9). We compare the FLoSS and FRS with regard to characteristics of sibships that would be selected by choosing the highest scoring sibships according to the following: sibship size, numbers of living siblings, mean age, and exceptionality of sibship survival. We compare these characteristics in 2 US populations selected for longevity (the Long Life Family Study and the New England Centenarian Study) and in 1

not selected for longevity (the Framingham Heart Study), a geographically based US cohort.

MATERIALS AND METHODS

Metric of estimated survival exceptionality for a family

We define both a family measure of survival exceptionality that requires knowing the age at death of all members and a method for estimating this measure when all family members have not yet died.

We characterize an *individual's* ''exceptionality of longevity'' at age A via $p(A)$, the probability that a random person in the same birth cohort survives to at least age A. We use $-\ln(p(A))$, the *information content of the rarity of survival* measured by ''Shannon information'' (10), as the basis for an individual's exceptionality score. The increment in this function for each additional year of survival rises with age. For example, the increase in its value for survival past age 96 years versus 95 years is far greater than the increase for survival past age 71 years versus age 70 years.

The probability of survival past age A varies by cohort as specified by year, country of birth, and gender. For example, on the basis of the US Social Security Administration cohort life tables (http://www.ssa.gov/OACT/NOTES/as116/as116LOT.html), the same age can produce very different exceptionality scores: $2.7 = -\ln(0.065)$ for a woman aged 95 years born in 1910 versus $4.2 = -\ln(0.015)$ for a man aged 95 years born in 1900. Specifically, an individual's exceptionality score at age A is $-\ln(p(A|C))$, where $p(A|C)$ is the probability of survival past age A for those born into that person's country-, birth year-, and gender-specific cohort C. Of course, exceptionality of longevity is fully observed only at death.

We next wished to measure the *family's* survival exceptionality. Note that simply summing the $-\ln(p)$ values for individuals cannot be correct, because these are always positive; thus, each additional family member, even one who dies at a young age, would increase the score. Clearly, a family of 2, both living at age 95, is more exceptional than a family of 3, with 2 still living at age 95 plus one who died at age 65. This suggests that each person's score should be of the form, $-\ln(p) - k$, where k is chosen so that the expected value of the score for a randomly selected person is 0. In that case, a family member for whom $-\ln(p)$ exceeds k increases the family's longevity exceptionality, while one who dies younger subtracts. The desired value of k is 1 because, for each cohort, the probabilities p in a life table are uniformly distributed on the interval from 0 to 1, and the integral (i.e., mean value) of $-\ln(p)$ over this interval is 1. Thus, we choose $-\ln(p(A|C)) - 1$ to measure each member's survival exceptionality for including in a family score, defined as follows:

$$\mathrm{SE}_f = \sum\nolimits_{\text{all family members}} (-\ln(p(A|C)) - 1),$$

in which all A's are ages at death.

For a living person of age A, we can calculate A^*, an expected age at death calculated from life tables. Specifically, A^* is a cohort-specific expected age at death, conditional on survival to age A; then, $-\ln(p(A^*|C))$ is an *estimated* measure of the person's exceptionality of survival. Because exceptional longevity is strongly familial (1, 2, 4–6), those with long-lived siblings are likely to live longer than life-table calculations suggest. (In this sense, $-\ln(p(A^*|C))$ is a conservative estimate of expected exceptionality in families of interest in longevity studies.) This leads to the following *estimated survival exceptionality* for a family f, in which each member's contribution increases the family score only if $-\ln(p(A|C))$ for dead members or $-\ln(p(A^*|C))$ for living members exceeds 1:

$$\begin{aligned}\mathrm{est(SE)}_f = &\sum\nolimits_{\text{dead family members}} (-\ln(p(A|C)) - 1)\\ &+ \sum\nolimits_{\text{living family members}} \left(-\ln\left(p\left(A^*|C\right)\right) - 1\right).\end{aligned}$$

Metric incorporating the added value of old living family members

The score ''$\mathrm{est(SE)}_f$'' can be computed with any mix of living and dead family members. For families with living members, it is a plausible current estimate of SE_f, the family's (ultimate) survival exceptionality. When all have died, it equals SE_f. Although $\mathrm{est(SE)}_f$ should be useful for many purposes, it may not be ideal for *selecting* families into a study of familial longevity, because it does not particularly value additional older *living* siblings, who provide more biologic and phenotypic data. Thus, we add a living siblings' bonus score that satisfies the following principles:

1. Living family members whose age A is already somewhat exceptional, in the sense that $-\ln(p(A|C)) - 1$ is greater than 0, should add to the family's value for genetic studies on exceptional survival *beyond* the value already captured in their contribution to $\mathrm{est(SE)}_f$.
2. The amount of added value conferred by these living family members should be a function of $-\ln(p(A|C)) - 1$, where A is their current age, rather than their expected age at death (A^*), since only survival through age A is certain.
3. Living family members who are still so young that $-\ln(p(A|C)) - 1$ is negative should neither add to nor subtract from the bonus.

This leads us to define the following living sibling bonus:

$$\mathrm{LB}_f = \sum\nolimits_{\text{living family members}} \max\{0, [-\ln(p(A|C)) - 1]\}.$$

Adding this to $\mathrm{est(SE)}_f$ produces our *family longevity selection score*, or FLoSS:

$$\begin{aligned}\mathrm{FLoSS}_f = &\textstyle\sum_{\text{dead family members}} (-\ln(p(A|C)) - 1)\\ &+ \textstyle\sum_{\text{living family members}} \left[\left(-\ln\left(p\left(A^*|C\right)\right)\right) - 1\right)\\ &+ \max\{0, -\ln(p(A|C)) - 1\}].\end{aligned}$$

To illustrate the FloSS calculation, consider a sibship with an oldest brother born in 1920 and died (at age 88) in 2008,

while 2 other siblings remain living at the end of 2008: a brother born in 1921 (aged 87 years) and a sister in 1930 (aged 78 years). The US Social Security Administration cohort life tables indicate that only about 12.2% of the brothers' cohort (males born between 1916 and 1925) lives past 88 and 14.3% lives past 87, while 57.4% of the females of the sister's cohort (born between 1926 and 1935) lives past 78. Furthermore, the living brother is expected to achieve age $A^* = 91$, while the sister's A^* is 88. The FLoSS is the sum of their 3 scores. Note that $-\ln(0.122) - 1 = 2.104 - 1 = 1.104$ is the score for the dead sibling, while for each of the 2 living siblings the score is computed as $-\ln(p(A^*|C)) - 1 + \max\{0, -\ln(p(A|C)) - 1\}$. From the US cohort tables, $-\ln(p(A^* = 91|1920, m)) - 1 = 1.685$, while $-\ln(p(A = 87|1920, m)) - 1 = 0.942$, so that $6 = \max\{0, -\ln(p(A = 87|1920, m)) - 1\} = 0.942$. Similarly, $-\ln(p(A^* = 88|1930, f)) - 1 = 1.253 - 1 = 0.253$, while $6 = -\ln(p(A = 78|1930, f)) - 1 = -0.445$ and $\max\{0, -\ln(p(A = 78|1930, f)) - 1\} = 0$. These values lead to an est(SE) for the sibship of $3.042 = 1.104 + 1.685 + 0.253$, a living siblings' bonus of $0.942 + 0$, and a FLoSS of 3.984. Additional sibships with various survival patterns and high FLoSS scores are described and illustrated in supplementary material (Web Figure 1). (This information is posted on the *Journal*'s website (http://aje.oxfordjournals.org/).)

An additional consideration is whether to exclude from scoring those family members who died before some age A_0 and, if so, what age. Factors to consider include the quality of data on deaths in infancy and the survival phenotype of principal interest, for example, survival from birth, survival conditional on reaching maturity, or survival conditional on reaching a specified older age. In the Long Life Family Study, given our focus on exceptional familial survival over the life span, we calculated survival conditional on reaching age 40 (rather than a younger age), on the basis of several considerations. First, in the early 20th century, deaths below age 2 were often not recorded. Furthermore, in our study cohorts, infectious disease epidemics and wars were prominent contributors to mortality in infancy, adolescence, and early maturity. Moreover, the role of familial factors in survival below age 40 is unclear (2). Hence, we calculated $p(A|C)$ in our analyses as the "probability of survival to age A, conditional upon survival to age $A_0 = 40$." We computed this from population birth cohort life tables, using only the subjects alive at age 40 as the referent population. We provide gender- and US birth cohort-specific scores for those who survived to at least age 40 in the online supplementary material (Web Table 1; refer to the *Journal*'s website) and on the Long Life Family Study website (https://longlifefamilystudy.wustl.edu/FLoSS).

Regarding the actual or expected ages at death required for individuals to make a positive contribution to this score, we note that $-1 \approx \ln(0.37)$, so that $-\ln(p(A|C)) - 1$ is only greater than 0 when p is smaller than 0.37, so that a positive score is achieved only when *longevity* exceeds the 63rd percentile for members of the same gender/birth year cohort who survived until at least age 40. In the 1900 birth cohort, for example, this corresponds to 77 years for males and 87 for females.

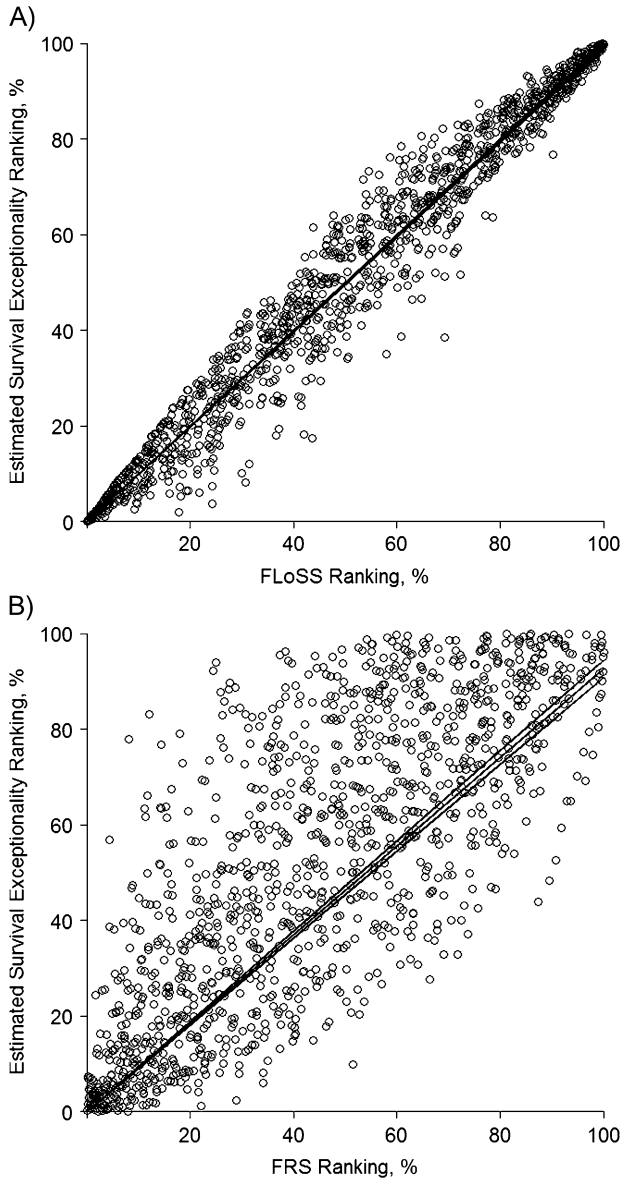


Figure 1. Characteristics of Long Life Family Study sibships in 2007 as ranked by the family longevity selection score (FLoSS) (A) and the family risk score (FRS) (B). These plots show the relation between the percentile ranking of the sibships screened for the Long Life Family Study based on the estimated survival exceptionality, est(SE), on the *y* axis and the percentile ranking based on the family longevity selection score on the *x* axis of part A and the family risk score on the *x* axis of part B. Diagonal lines represent least-square regression lines and 95% confidence bounds. Note that the 3 lines essentially overlap in the plot of part A.

Comparing the FLoSS and the FRS

A FRS can be generated for quantitative phenotypes, adjusting for family size and the distribution of the phenotype in the general population (8, 9). When applied to longevity in sibships, the FRS is the *average* over all siblings in a family of their individual scores, S, where:

Table 1. Age and Size Distributions of Low- and High-ranked Sibship Cohorts in the Long Life Family Study, by Ranking Method, in 2007[a]

Siblings Included and Ranking Method	Sibship Age, years					
	Lowest Ranked 40%			Highest Ranked 5%		
	Mean	Median	% >90 Years	Mean	Median	% >90 Years
All						
FLoSS	80.6	81.0	6	92.8	92.5	67
FRS	78.9	79.5	0	99.9	99.0	100
Living only						
FLoSS	85.0	85.0	16	95.8	95.8	87
FRS	84.8	84.5	13	100.0	100.0	100
	Sibship Size, no.					
	Lowest Ranked 40%			Highest Ranked 5%		
	Mean	Median	% >1 Sibling	Mean	Median	% >1 Sibling
All						
FLoSS	3.0	3.0	75	5.3	5.0	95
FRS	4.0	4.0	91	1.3	1.0	20
Living only						
FLoSS	1.7	1.0	44	3.0	3.0	79
FRS	2.4	2.0	64	1.0	1.0	3

Abbreviations: FLoSS, family longevity selection score; FRS, family risk score.

[a] Summary statistics of size and age of sibships ranked in the lowest 40% and the top 5% by the FLoSS or the FRS in the Long Life Family Study. In each of the 4 pairs of rows, the first describes sibships as ranked by the FLoSS and the second, by the FRS. Age statistics describe the average achieved ages (either age at death or current age), for either all siblings or only living ones, as indicated, in the included sibships. All differences in sibship age and size for the FRS versus FLoSS top groups were significant ($P < 0.001$; Wilcoxon test).

$$S = \frac{(A - \tilde{A})}{\sqrt{\tilde{A}}}, \text{if the sibling is dead, and}$$

$$= \max(0, \frac{(A - \tilde{A})}{\sqrt{\tilde{A}}}), \text{if the sibling is alive,}$$

where A denotes the current age (or age at death) of the sibling, and $\tilde{A}$ is the sex/birth year cohort-adjusted life expectancy of that sibling. Note that, in long-lived sibships, each still-living but younger-than-average sibling pulls the FRS toward 0. In addition, the FRS scores a family with 2 exceptional siblings identically to a family with 4 or 8 siblings, each pair of which is just as exceptional as the 2.

Characteristics of populations used to compare FLoSS and FRS

We calculated these scores in 3 sibship samples. Only the first 2 are enriched for longevity.

Sample 1: 660 sibships from the New England Centenarian Study—a nationwide-based sample of centenarians, their siblings, and offspring. In all selected sibships, the proband was aged $\geq$100 years. In 2005, 49% of all siblings were still living, 33% of the sibships contained at least 1 living member, and the average attained age was 91 years.

Sample 2: 1,671 US sibships from the screening pilot phase of the Long Life Family Study—a US–Danish study of long-lived sibships and their offspring (https://longlifefamilystudy.wustl.edu/). Sibships of living probands who were at least 80 years of age were screened in the US component of the Long Life Family Study. In 2007, 58% of all siblings were still living, all sibships contained at least 1 living member (the proband who provided data on the other siblings), and the average age was 83 years.

Sample 3: 766 sibships from the Framingham Heart Study (11)—a sample of longitudinally followed subjects and their families. The Framington Heart Study has enrolled subjects born from 1888 onward; all our sibships had at least 1 sibling born before 1925. In 2004, 13% of all siblings were still living, 21% of the sibships contained at least 1 living member, and the average attained age was 74 years.

Cohort, survivorship, and sibship data

For the New England Centenarian Study, Framington Heart Study, and Long Life Family Study sibships, we estimated gender/birth-year survival functions and calculated the

expected age at death (A^*) using the Social Security Administration cohort life tables for successive birth decades, with the 1900 tables for births from 1895 through 1904, the 1910 tables for births from 1905 through 1914, and so on. These tables are available at http://www.ssa.gov/OACT/NOTES/as116/as116LOT.html. In the Long Life Family Study data, siblings with unknown vital status were considered to be dead, with the age at death given (conservatively) as the age at last contact. For an unknown birth year, we used the youngest sibling's birth year. Both siblings and half-siblings were included in FLoSS score calculations.

Statistical analysis

We estimated Spearman correlations between the percentile ranking based on $\text{est}(\text{SE})_f$ and the FLoSS and FRS. We used summary statistics to illustrate differences between sibships in the lowest 40 percentile groups and in the top 5%, as defined by each score, and Wilcoxon and t statistics to test for their significance. Reported P values are 2 sided.

RESULTS

As shown in Figure 1, the percentile ranking of families screened for the Long Life Family Study based on $\text{est}(\text{SE})_f$, the estimated "survival exceptionality," is more strongly correlated with the rankings based on the FLoSS ($r = 0.98$) (Figure 1A) than with those based on the FRS ($r = 0.71$) (Figure 1B). As seen in Table 1, the sibships ranked in the top 5% by the FLoSS and FRS have quite different characteristics. Sibships ranked in the top 5% by the FRS average between 4 and 7 years older but are far smaller than those ranked in the top 5% by the FLoSS. Note that 79% of the sibships ranked in the top 5% by the FLoSS have more than 1 living sibling compared with only 3% of the sibships ranked in the top 5% by the FRS.

Figure 2 shows the distribution of the FLoSS and FRS in the sibships of the Framington Heart Study (left), the Long Life Family Study (middle), and the New England Centenarian Study (right). For comparability, the FLoSS and FRS scores in the Long Life Family Study and the New England Centenarian Study are each standardized to their distributions in the Framington Heart Study, a sample that is not enriched for longevity. In contrast, the New England Centenarian Study recruited families with centenarians, resulting in a strongly shifted distribution of the standardized FLoSS scores. The FLoSS distribution for the Long Life Family Study screening population is intermediate, because it includes families with some evidence of longevity. The standardized FRS score distributions also shift to the right in the Long Life Family Study and New England Centenarian

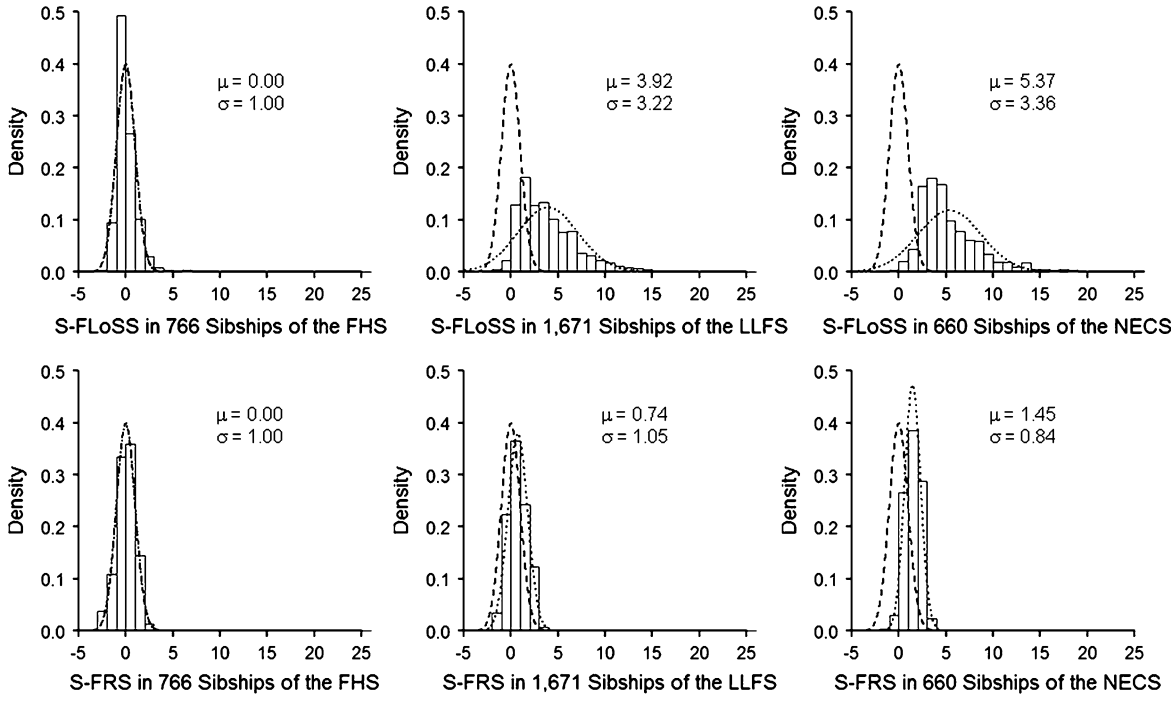


Figure 2. Distribution of the standardized family longevity selection score (S-FLoSS) and the standardized family risk score (S-FRS) for sibships from the Framingham Heart Study (FHS), the Long Life Family Study (LLFS) screenees, and the New England Centenarian Study (NECS). S-FLoSS is defined as $(\text{FLoSS} - M)/S$, where $M = -0.24$ and $S = 1.47$ are, respectively, the mean and standard deviation of FLoSS scores in the FHS cohort. S-FRS is defined analogously (i.e., as a z score in the same cohort, with $M = 1.98$ and $S = 1.30$). Parameters μ and σ in each panel are the mean and standard deviation of standardized scores. Dashed lines represent the standard normal density, and dotted lines represent the normal density with mean and standard deviation of the S-FLoSS and S-FRS. The S-FLoSS values for the LLFS and NECS sibships are significantly different from the S-FLoSS values for the FHS sibships ($P < 0.0000$; Wilcoxon test). Similarly, the S-FRS values for the LLFS and NECS sibships are significantly different from the S-FRS values for the FHS sibships ($P < 0.0000$; Wilcoxon test).

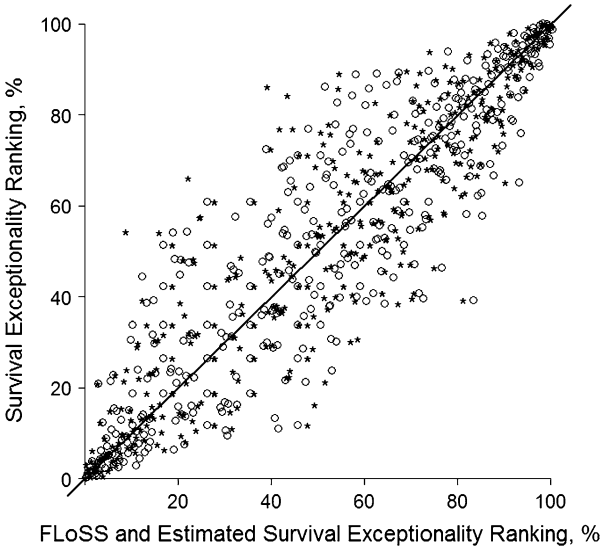


Figure 3. Rankings for survival exceptionality versus family longevity selection score (FLoSS) and estimated survival exceptionality (est(SE)) among 442 New England Centenarian Study sibships with final survivors' deaths within 10 years after 1995. For each of the 442 New England Centenarian Study sibships with at least 1 living member in 1995 and none in 2004, an open circle relates its percentile ranking based on observed survival exceptionality in 2004 on the *y* axis to its 1995 family longevity selection score percentile ranking on the *x* axis. Stars show the same relation for a sibship's percentile ranking based on survival exceptionality in 2004 (on the *y* axis) and percentile ranking based on its 1995 estimated survival exceptionality on the *x* axis.

Study. Indeed, all 3 cohort pairs differ in score distributions (all $P < 0.001$; Wilcoxon test). However, the FLoSS distinguishes the populations far more sharply. Note that less than 1% of Framington Heart Study sibships have a FLoSS of >7 (standardized FLoSS, >4.95). To focus on unusually long-lived families, the Long Life Family Study enrolled only families with a FLoSS of 7 or more.

We also examined the degree to which the FLoSS and estimated survival exceptionality score predict actual survival exceptionality (SE_f), observed only after all have died. For those 442 sibships of the New England Centenarian Study in which at least 1 sibling was alive in 1995 but all had died by 2004, we computed the FLoSS using data known in 1995 and the (fully observed) SE_f in 2004. The scatterplot in Figure 3 shows the relation between the sibships' rankings based on the 2004 SE_f (*y* axis) and those based on the FLoSS and on the estimated survival exceptionality longevity score, $\text{est}(SE)_f$, in 1995. The ranking based on the FLoSS and on $\text{est}(SE)_f$ in 1995 correlates equally strongly with the ranking based on the SE_f measured a decade later (both $r = 0.90$).

DISCUSSION

A sibship's FLoSS is a good predictor of its true (not yet fully known) exceptionality of survival. Thus, it is useful for selecting long-lived families for studies of genetic and non-genetic factors contributing to longevity. In particular, the FLoSS can identify desirable sibships among families being screened for genetic epidemiologic studies of exceptional longevity. Sibships with a high FLoSS have high ages of the total sibship and of living siblings, as well as high numbers of total and living members.

A high FLoSS can be achieved by the presence of a single individual with extremely long survival (if the other siblings are not short lived) or by the presence of many long (but not extremely)-lived siblings. The modes of transmission of longevity (if any) associated with these types of patterns may differ. For example, the former might reflect a recessive trait; the latter, a dominant. The fact that a high FLoSS value captures both of these patterns is one of its strengths, because it does not exclude longevous families who might not be ascertained by other methods (e.g., setting a single extremely exceptional age for inclusion of individuals or siblings). However, the additional potential heterogeneity of genetic factors captured in high-FLoSS families implies a need for attention to potential differences among subgroups in genetic transmission.

We used the FLoSS to score families defined as members of a single sibship. Although it could be used to score multigenerational groups of relatives, gathering complete information on the parents of old siblings or on widely dispersed groups of near relatives could be challenging. Furthermore, young people contribute little information regarding ultimate survival.

In the Long Life Family Study, we set the minimum FLoSS for a family to be eligible at 7. This threshold was chosen by observing that only 0.2% of the FLoSS sibships of the Framington Heart Study meet this threshold, in contrast to over 30% of the Long Life Family Study screening families and over 40% of families enrolled in the New England Centenarian Study. Thus, families with a FloSS as large as 7 are extremely rare but findable. Calculation of FLoSS scores in families from additional population-based samples can provide further guidance about appropriate selection thresholds.

The FLoSS is the sum of $\text{est}(SE)_f$ and LB_f (the living bonus). Its first component is a current estimate of a family's ultimate, "fully observed," survival exceptionality, SE_f. Although the living bonus enables the FLoSS to *select* families that are more desirable for genetic epidemiologic studies of exceptional longevity, $\text{est}(SE)_f$ may be a better intrinsic measure of family longevity. As such, $\text{est}(SE)_f$ may be particularly useful for examining relations of genotypes and other risk factors to phenotypes of family members and their family's exceptionality of survival, or for finding subsets of populations that are similar with respect to the exceptionality of their families. Although in our comparison the FLoSS and $\text{est}(SE)_f$ correlated equally well with the observed SE_f 10 years later, this may not apply over longer intervals. Note that, when all family members have died, all 3 measures coincide.

The FLoSS as used in the Long Life Family Study can be viewed as a member of a class of family scores that combine an estimated exceptionality of survival with a bonus for living siblings, as in the following:

$$\text{Longevity selection score} = \text{est(SE)}_f + w \times \text{LB}_f,$$

where est(SE)_f is of the form,

$$\sum_{\text{dead family members}} (-\ln(p(A|C(a))) - k) + \sum_{\text{living family members}} \left(-\ln\left(p\left(A^*|C\left(a\right)\right)\right) - k\right).$$

Here, w and k are nonnegative constants, $C(a)$ is a person-specific reference cohort of those born around the same time who survived to some minimum age, a, and LB_f is a nonnegative bonus for older living siblings. Each parameter can be tailored to the particular needs of other studies.

For the FLoSS, we chose $w = 1$, $C(a)$ = all people in the same birth-year, gender, and national cohort who survived to at least age 40, $k = 1$, and

$$\text{LB}_f = \sum_{\text{living family members}} \max\{0, [-\ln(p(A|C(a))) - 1]\}.$$

An age threshold other than 40 years could be chosen depending on a study's focus (e.g., age 70 to study factors influencing survival only in advanced age). Choosing $w = 1$ weights the estimated survival exceptionality and living bonus equally; $w = 0$ yields est(SE)_f, a measure of family survival exceptionality alone. Note that the expected value of est(SE)_f for a ''pseudofamily'' constructed by grouping N randomly selected individuals from the US population is $N \times (1 - k)$. Choosing $k = 1$ makes est(SE)_f neutral to sibship size, because its expected value should be 0 for randomly selected (not particularly long-lived) people. The ''neutrality'' of the FLoSS was validated by the fact that its average in the geography-based Framingham Heart Study cohort was quite close to 0. Because –ln(0.37) is approximately equal to 1, only ages in the top 37th percentile add to est(SE)_f. More generally, values of k smaller than 1 favor larger families (because each ''typical'' person's expected score is positive), while choosing k larger than 1 favors smaller ones. If greater exceptionality is sought, a larger cutoff for the FLoSS could be used.

We chose $w = 1$, giving equal weight to the estimated exceptionality score and the ''bonus'' for living older family members. This reflected our interest in both survival exceptionality and the availability of old living study subjects. With this choice, the FLoSS still correlates strongly with est(SE)_f as we wanted. Larger w's will give more weight to the living bonus and reduce the correlation with est(SE)_f.

In summary, we have introduced and examined the consequences of a conceptually attractive framework for family longevity studies. These include 1) a measure of an individual's exceptionality of survival, 2) a feasible way to estimate that exceptionality for those still alive, 3) a size-neutral way to combine individual scores into a family score, 4) a plausible bonus measure for the additional value of already exceptional living family members, and 5) a way to balance interest in older living relatives and family survival exceptionality in a single score such as the FLoSS. This framework should be useful in many settings. We also note that formulas for the FLoSS and est(SE)_f could be adapted to measure exceptionality of survival until other events besides death. Thus, we could quantify family risk for the onset of conditions (such as stroke or onset of diabetes or disability) whose incidence rises with age. We are investigating this idea.

ACKNOWLEDGMENTS

Author affiliations: Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts (Paola Sebastiani); Division of Geriatrics and Clinical Gerontology, National Institute on Aging, Bethesda, Maryland (Evan C. Hadley, Winifred Rossi); Department of Biostatistics, Washington University, St. Louis, Missouri (Michael Province); the Danish Aging Research Center, Institute of Public Health, University of Southern Denmark, Odense, Denmark (Kaare Christensen); Geriatrics Division, Department of Medicine, Boston University Medical Center, Boston, Massachusetts (Thomas Perls); and Health Care Research Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts (Arlene S. Ash).

Funded by grants U01AG023755 (P. S., T. P., A. S. A.), U01AG023746 (M. P.), and U01AG023712 (K. C.). The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with Boston University.

The authors thank the staff and investigators of the Framingham Heart Study for providing access to their data.

This article was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or the National Heart, Lung, and Blood Institute.

Conflict of interest: none declared.

REFERENCES

1. Kerber RA, O'Brien E, Smith KR, et al. Familial excess longevity in Utah genealogies. *J Gerontol A Biol Sci Med Sci.* 2001;56(3):B130–B139.
2. Perls TT, Wilmoth J, Levenson R, et al. Life-long sustained mortality advantage of siblings of centenarians. *Proc Natl Acad Sci U S A*. 2002;99(12):8442–8447.
3. Willcox BJ, Willcox DC, He Q, et al. Siblings of Okinawan centenarians share lifelong mortality advantages. *J Gerontol A Biol Sci Med Sci.* 2006;61(4):345–354.
4. Gudmundsson H, Gudbjartsson DF, Frigge M, et al. Inheritance of human longevity in Iceland. *Eur J Hum Genet.* 2000;8(10):743–749.
5. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet.* 2006;7(6):436–448.
6. Perls T, Kunkel L, Puca A. The genetics of aging. *Curr Opin Genet Dev.* 2002;12(3):362–369.
7. Williams RR, Hunt SC, Barlow GK, et al. Health family trees: a tool for finding and helping young family members of coronary and cancer prone pedigrees in Texas and Utah. *Am J Public Health.* 1988;78(10):1283–1286.

8. Li R, Bensen JT, Hutchinson RG, et al. Family risk score of coronary heart disease (CHD) as a predictor of CHD: the Atherosclerosis Risk in Communities (ARIC) Study and the NHLBI Family Heart Study. *Genet Epidemiol*. 2000;18(3):236–250.
9. Pereira MA, Schreiner PJ, Pankow JS, et al. The family risk score for coronary heart disease: associations with lipids, lipoproteins, and body habitus in a middle-aged bi-racial cohort: the ARIC Study. *Ann Epidemiol*. 2000;10(4): 239–245.
10. Shannon CE. The mathematical theory of communication. 1963. *MD Comput.* 1997;14(4):306–317.
11. Brand FN, Kiely DK, Kannel WB, et al. Family patterns of coronary heart disease mortality: the Framingham Longevity Study. *J Clin Epidemiol*. 1992;45(2):169–174.

DOI: 10.1111/j.1475-6773.2009.00994.x
RESEARCH ARTICLE

Quality and Safety

Comparing Safety Climate between Two Populations of Hospitals in the United States

Sara J. Singer, Christine W. Hartmann, Amresh Hanchate, Shibei Zhao, Mark Meterko, Priti Shokeen, Shoutzu Lin, David M. Gaba, and Amy K. Rosen

Objective. To compare safety climate between diverse U.S. hospitals and Veterans Health Administration (VA) hospitals, and to explore the factors influencing climate in each setting.
Data Sources. Primary data from surveys of hospital personnel; secondary data from the American Hospital Association's 2004 Annual Survey of Hospitals.
Study Design. Cross-sectional study of 69 U.S. and 30 VA hospitals.
Data Collection. For each sample, hierarchical linear models used safety-climate scores as the dependent variable and respondent and facility characteristics as independent variables. Regression-based Oaxaca–Blinder decomposition examined differences in effects of model characteristics on safety climate between the U.S. and VA samples.
Principal Findings. The range in safety climate among U.S. and VA hospitals overlapped substantially. Characteristics of individuals influenced safety climate consistently across settings. Working in southern and urban facilities corresponded with worse safety climate among VA employees and better safety climate in the U.S. sample. Decomposition results predicted 1.4 percentage points better safety climate in U.S. than in VA hospitals: −0.77 attributable to sample-characteristic differences and 2.2 due to differential *effects* of sample characteristics.
Conclusions. Results suggest that safety climate is linked more to efforts of individual hospitals than to participation in a nationally integrated system or measured characteristics of workers and facilities.

Key Words. Safety culture, safety climate, survey research, hospitals, integrated hospital networks, decomposition

Based on mounting evidence that better safety climate is related to lower incidence (Naveh, Katz-Navon, and Stern 2005; Hofmann and Mark 2006; Neal and Griffin 2006; Vogus and Sutcliffe 2007; Singer et al. 2008b) and

greater reporting (Cohen et al. 2004; Weingart et al. 2004; Gandhi et al. 2005) of adverse events and to increased communication among managers and staff (Hofmann and Morgeson 1999), considerable effort among hospitals is being focused on improving safety climate. Along with hospitals' own efforts, several voluntary, collaborative initiatives that could improve safety climate (e.g., Leapfrog Group's patient safety leaps, and Institute for Healthcare Improvement's 5 Million Lives campaign) have garnered substantial participation among both public and private hospitals. "Benchmarking" of safety-climate survey results through participation in such collaborative initiatives is an effective way for hospitals to target quality improvement efforts. Benchmarking enables a hospital to compare its survey results with those of other hospitals, thereby facilitating identification of relative strength and weakness. It is being encouraged by numerous organizations. Since 2002, the Joint Commission's performance improvement standard (PI.01.01.01) has encouraged hospitals to collect data on staff perceptions of safety risks and improvement opportunities and to compare data with external sources (Joint Commission on Accreditation of Health Care Organizations 2002). The Agency for Healthcare Research and Quality (AHRQ) established the *Hospital Survey on Patient Safety Culture Comparative Database* for this purpose in 2006. Its 2009 database included safety-climate results from 622 hospitals (Sorra et al. 2009). Independent investigators engaged in benchmarking safety climate have also identified systematic differences in safety climate within and among hospitals, which provide clues to improving safety climate more generally (Singer et al. 2003; Thomas, Sexton, and Helmreich 2003; Makary et al. 2006; Sexton et al. 2006a, b, c; Vogus and Sutcliffe 2007; Hartmann et al. 2008; Singer et al.

Address correspondence to Sara J. Singer, M.B.A., Ph.D., Department of Health Policy and Management, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115; e-mail: ssinger@hsph.harvard.edu. Sara J. Singer, M.B.A., Ph.D., is also with the Institute for Health Policy, Massachusetts General Hospital, Boston, MA. Sara J. Singer, M.B.A., Ph.D., Shoutzu Lin, M.S., and David M. Gaba, M.D., are with the Center for Health Policy and Center for Primary Care and Outcomes Research, Stanford University, Stanford, CA. Christine W. Hartmann, Ph.D., is with the Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, MA. Christine W. Hartmann, Ph.D., Mark Meterko, Ph.D., and Amy K. Rosen, Ph.D., are with the Department of Health Policy and Management, Boston University School of Public Health, Boston, MA. Amresh Hanchate, Ph.D., is with the Section of General Internal Medicine, Boston University School of Medicine, Boston, MA. Shibei Zhao, M.P.H., Mark Meterko, Ph.D., Priti Shokeen, M.S., and Amy K. Rosen, Ph.D., are with the Center for Organization, Leadership, and Management Research, Boston VA Medical Center, Boston, MA. Shoutzu Lin, M.S., is with the Chronic Heart Failure (CHF) Quality Enhancement Research Initiative, and David M. Gaba, M.D., is with the Patient Safety Culture Institute, VA Palo Alto Health Care System, Palo Alto, CA.

2008a, 2009). Benchmarking safety-climate survey results across health care systems is difficult due to coordination challenges.

Our own effort to measure and benchmark safety climate is unique in that we used essentially the same survey instrument and sampling and administration procedures at approximately the same time in two discrete populations of hospitals in the United States: the Veterans Health Administration (VA) health care system and a national sample, excluding VA hospitals. This provided a novel opportunity to compare safety climate across two populations encompassing great diversity in both individual hospital characteristics and overall organizational structure, and potentially to identify any hospital features systematically related to safer care.

Differences in safety climate between VA and other U.S. hospitals may inherently exist because the VA is a nationally integrated network, while few U.S. hospitals come from large, integrated systems and none come from nationally integrated systems. As a system, the VA enjoys distinct advantages in broadly implementing and enforcing compliance with standardized safety activities. The VA conducts several initiatives with potential to improve safety climate. For example, the VA National Center for Patient Safety was established specifically to promote a systems approach to preventing and reducing harm to patients and to encourage hospitals to conduct root cause analyses after safety incidents (http://www.va.gov/ncps/vision.html, accessed on August 26, 2008, for NCPS).

In this paper, we examine differences in safety climate between 69 diverse U.S. hospitals and 30 VA hospitals using cross-sectional employee surveys. Given potential advantages in promoting strong safety climate in a nationally integrated network, we hypothesized that safety climate among VA hospitals would be stronger than among U.S. hospitals.

Differences in safety climate between U.S. and VA samples may arise from multiple sources. First, there may be variation between the two samples in measured characteristics associated with safety climate—for instance, one sample may contain more large hospitals than the other. The residual difference in safety climate between the two samples would be attributable to differential *effects* of the sample characteristics between the two health care systems (e.g., hospital size may impact VA hospitals differently than U.S. hospitals). We performed comparisons of safety climate in these two settings that allowed us to discern the relative impact of these potential sources of difference. We hypothesized that variance in observed sample characteristics would explain more of the difference in safety climate between U.S. and VA hospitals than would differential *effects* of sample characteristics on the two groups.

METHODS

Data Sources

We used the Patient Safety Climate in Healthcare Organizations (PSCHO) survey to collect data on employees' perceptions of safety climate. While various instruments exist to measure hospital safety climate (Colla et al. 2005; Flin et al. 2006), the PSCHO instrument is the only one with established reliability and validity in both U.S. and VA hospital settings (Singer et al. 2007; Hartmann et al. 2008). PSCHO survey items use a five-point, Likert scale ranging from "strongly agree" to "strongly disagree," with a neutral midpoint. Items reflect 12 dimensions that capture various aspects of safety climate. We divided these dimensions into three categories, based on the extent to which they described hospital (e.g., "Organizational Resources for Safety"), work-unit (e.g., "Unit Safety Norms"), and interpersonal (e.g., "Fear of Blame and Punishment") contributions to safety climate (Singer et al. 2007).

Because of modifications resulting from psychometric testing, two slightly different versions of the PSCHO survey were used in this study. In U.S. hospitals we used a 45-item instrument, while in the VA we used a 42-item instrument. The two versions have 41 common items, 39 of which map onto the 12 safety-climate dimensions. Both versions of the PSCHO also contained six close-ended demographic items.

Because the development of a strong safety climate necessitates a homogenous focus on preventing safety failures, the PSCHO instrument is scored to highlight responses opposed to safety, which we refer to as "problematic responses." We generated scores for items, dimensions, and safety climate overall. First, we calculated the mean percent problematic response ("PPR") for a given item across all respondents. We then calculated the mean of all item means in a dimension and the mean of all item means in the survey. A lower mean indicates a better perception of safety climate. This method of scoring identifies areas of nonuniformity in safety focus that are of potential concern and that might benefit from interventions to improve the safety climate.

Data for characteristics of respondents' jobs were obtained from the 2004 American Hospital Association (AHA) Annual Survey of Hospitals. Using these data, we determined hospitals' nurse staffing ratios, bed size, teaching status, national census region, and urban or nonurban location (see Table 1).

Approval from relevant Institutional Review Boards was granted before conducting the studies.

Table 1: Respondents' Individual and Facility Characteristics

	U.S. (N= 13,841)		*VA (N= 4,581)*	
Variables	*N*	*%*	*N*	*%*
Individual characteristics				
Age category (years)				
18–30	1,249	9.3	180	4.1
31–50	6,973	52.0	2,066	46.7
>50	5,197	38.7	2,179	49.3
Time at facility				
≤ 10 years	7,817	57.6	2,319	51.8
>10 years	5,749	42.4	2,154	48.2
Male				
Yes	4,708	35.8	2,150	49.8
No	8,451	64.2	2,170	50.2
Nurse				
Yes	3,444	25.5	1,016	22.7
No	10,050	74.5	3,461	77.3
Senior manager				
Yes	2,267	17.8	544	12.6
No	10,455	82.2	3,778	87.4
Employment in high hazard unit (OR, ER, ICU, PACU)				
Yes	2,716	22.6	731	17.5
No	9,308	77.4	3,454	82.5
Facility characteristics*				
Region				
East	3,891	28.1	1,344	29.3
Midwest	2,628	19.0	1,234	27.0
South	3,663	26.5	1,201	26.2
West	3,659	26.4	802	17.5
Teaching status†				
Major teaching	4,900	35.4	3,155	68.9
Minor teaching	3,176	23.0	1,045	22.8
Non-teaching	5,765	41.6	381	8.3
Urban‡				
Yes	11,643	84.1	4,452	97.2
No	2,198	15.9	129	2.8
Bedsize				
Small (≤ 99)	1,483	10.7	525	11.5
Medium (100–249)	2,166	15.7	981	21.4
Large (≥ 250)	10,192	73.6	3,075	67.1
Nurse staffing ratio§				
Mean (SD)	12.1	(4.2)	9.7	(4.1)

Note. *T*-test (nursing staffing ratio) or χ^2 tests (all other variables) were conducted to test for differences between non-VA and VA respondents. *p*-values were significant (<.001) for all comparisons.

*Facility characteristic variables were created using American Hospital Association Annual Survey Database FY'04.

†"Major teaching" hospital category includes members of the Council of Teaching Hospitals of the Association of American Medical Colleges; "minor teaching" hospitals have residency training programs approved by the Accreditation Council for Graduate Medical Education or by the American Osteopathic Association; and "non-teaching" hospitals are neither of the above.

‡AHA CBSA (Core-Based Statistical Area) type is "division" or "metropolitan."

§Full-time equivalent registered nurse hours per total facility inpatient days.

ER, emergency room; ICU, intensive care unit; OR, operating room; PACU, postanesthesia care unit; SD, standard deviation; VA, Veterans Health Administration.

Samples

We used stratified random sampling strategies in both populations. The U.S. hospital sample represented non-VA public and private acute-care hospitals, approximately equally divided among U.S. census regions and size categories. The VA sample represented a balanced geographic distribution of VA hospitals in four performance strata based on AHRQ's Patient Safety Indicators (PSIs) (low, medium, high, and other), to minimize selection bias. Details of the sampling strategies have been summarized elsewhere (Hartmann et al. 2008; Singer et al. 2009).

Although we did not stratify the U.S. sample based on performance, the sample included 69 hospitals whose PSI rates were similar to those of all U.S. hospitals. Our recruitment strategy, however, dictated that average size and related characteristics would differ from the U.S. average (Singer et al. 2009). In addition, despite recruitment efforts, hospitals from the Midwest were underrepresented in our sample compared with the U.S. average.

The VA sample included 30 hospitals, including eight facilities each from the high, medium, and other PSI rate strata, and six from the low stratum. The VA facilities represented a balanced geographic distribution within each PSI stratum, with the exception of no low PSI hospitals in the West (Rosen et al. 2008).

Administration of Surveys

U.S. hospital survey administration took place from July 2006 to May 2007; the VA administration was conducted from December 2005 to May 2006. In both groups we sampled 100 percent of senior managers, defined as department head or above; 100 percent of active hospital-based physicians; and a random 10 percent of all other employees. Senior managers and physicians were over-sampled because of their relatively small numbers and their potentially low response rates, respectively.

For U.S. hospitals, we also sampled 100 percent of employees in three work areas in 12 larger hospitals with relatively high response rates in a 2004 survey administration so as to permit work-area-level analyses while maintaining respondent confidentiality. In these hospitals, we over-sampled employees in work areas that in 2004 were least likely to meet our 10-respondent minimum reporting requirement: laboratories (lab), operating rooms (ORs), and intensive care units (ICUs). Budget constraints drove this selection approach.

In the VA, to allow for analysis of work areas in which employees conduct work of intrinsically greater hazard, we also sampled 100 percent of employees in certain work areas in 10 randomly selected hospitals.

The specific work areas were the OR, postanesthesia care unit, ICU, and emergency department. In this paper, we refer to these as "high hazard units" (HHUs).

The sampling frames in U.S. and VA hospitals consisted of 36,375 and 9,309 personnel, respectively. Both samples excluded individuals who no longer worked at the facility and those who used a survey response postcard to indicate that they did not wish to participate.

Analysis of Data

Weighting of Data. Two U.S. hospitals were excluded from analysis because AHA data suggested improbably high nurse staffing ratios. One VA hospital was dropped because it returned data for physicians only. For the remaining hospitals, we employed weighting techniques to reflect the two sampling frames accurately (Singer et al. 2003; Hartmann et al. 2008). Identical weighting calculations were performed for each sample. First, we determined separate sampling and nonresponse weights. Regarding the latter, for the U.S. hospital sample we calculated a nonresponse weight for each workgroup (senior managers, physicians, and other employees) within each hospital. In VA hospitals, we calculated four nonresponse weights: for senior managers, physicians, HHU employees, and regular staff for each hospital. Then, in both samples, we multiplied the nonresponse and sampling weights and used the resulting "combined weight" to calculate a proportional weight that accounted for hospital size differences.

Statistical Analysis. For all analyses, the unit of analysis was the individual. Initially, we compared sample characteristics of respondents in U.S. and VA hospitals. We compared overall mean PPR in each hospital, graphically distinguishing hospitals from the U.S. and VA samples. We assessed internal consistency reliability for the 12 dimensions of the PSCHO instrument by calculating Cronbach's α coefficients for proposed dimensions for each sample. We compared average PPR among U.S. and VA hospitals for each item and dimension.

The dependent variable for all statistical models was PPR for each individual across all 39 PSCHO survey items, a summary measure we call "safety climate overall." All models included variables describing individual respondents (i.e., gender, age, length of time at institution, job type, management category, and employment in HHU) and the facilities in which they worked (i.e., geographic region, hospital size, urban location, and nurse staffing ratio). Teaching status was not included in the models because major teaching status was correlated with large hospital size ($r = 0.5$, $p < .001$) in both samples.

We examined the relationship between PPR and respondent characteristics in the United States and VA by estimating a separate hierarchical linear model (HLM) for each sample. To test the appropriateness of using two-level HLM to account for nesting of individuals within hospitals, we first ran random effects ANOVA "empty" models that included no independent variables (Snijders and Bosker 1999). Comparison of the two-level models with the linear regressions revealed significant differences at the hospital level in both samples ($\chi^2 = 449$ and 21 for U.S. and VA hospitals, respectively; both $p < .001$), indicating that there were meaningful differences in PPR among staff from different hospitals and that two-level random intercept models were preferred. The models did not assume that PPR was uniformly represented within a facility; rather, they allowed for variation within and across facilities at the individual level. We did not use three-level HLMs to account for work-area variance due to limitations of the work-area data.

To examine variance in observed sample characteristics between the U.S. and VA hospitals and differential effects of sample characteristics on PPR in the two groups, we conducted a regression-based decomposition approach developed by Oaxaca and Blinder (Blinder 1973; Oaxaca 1973). Oaxaca–Blinder decomposition has been broadly applied in the economics literature and more recently in health services research (Kirby, Taliaferro, and Zuvekas 2006; Shen and Long 2006; Hudson, Miller, and Kirby 2007). We refer to systematic variance between the two samples in characteristics associated with safety climate as the "sample-characteristics component" because it is explained by observable variation in sample characteristics. We estimated the sample-characteristics component by using the U.S. hospital model estimates as the reference model. The residual difference in PPR between the two samples is called the "unexplained component" and includes (a) differential *effects* of the sample characteristics in the model between the two health care systems, and (b) differences in unobserved factors such as differences in characteristics of patients. Thus, the sample-characteristics component in our Oaxaca–Blinder decomposition measures the expected difference in PPR assuming that the same model is applicable to both systems, and the unexplained component measures the extent to which the *effects* of observed and unobserved characteristics in the models differ between U.S. and VA hospitals. In other words, the unexplained component indicates how the U.S. and VA samples would differ if the distribution of the sample characteristics were exactly the same.

Analyses were conducted using *Stata* (version 9.2), including the *Oaxaca* module (Jann 2008) for the Oaxaca–Blinder decomposition.

RESULTS

Survey Response

Among the 67 U.S. hospitals studied, 13,841 individuals responded to the survey (41 percent). Response rates for individual hospitals ranged from 13 to 100 percent. Response also varied by type of personnel, with 62 percent of senior managers, 20 percent of physicians, and 50 percent of frontline employees responding.

For the 29 VA hospitals, we obtained an overall response rate of 50 percent (4,581 respondents). Response rates varied among hospitals (26–73 percent) and among personnel (69 percent for senior managers, 38 percent physicians, 38 percent HHU personnel, and 60 percent other staff).

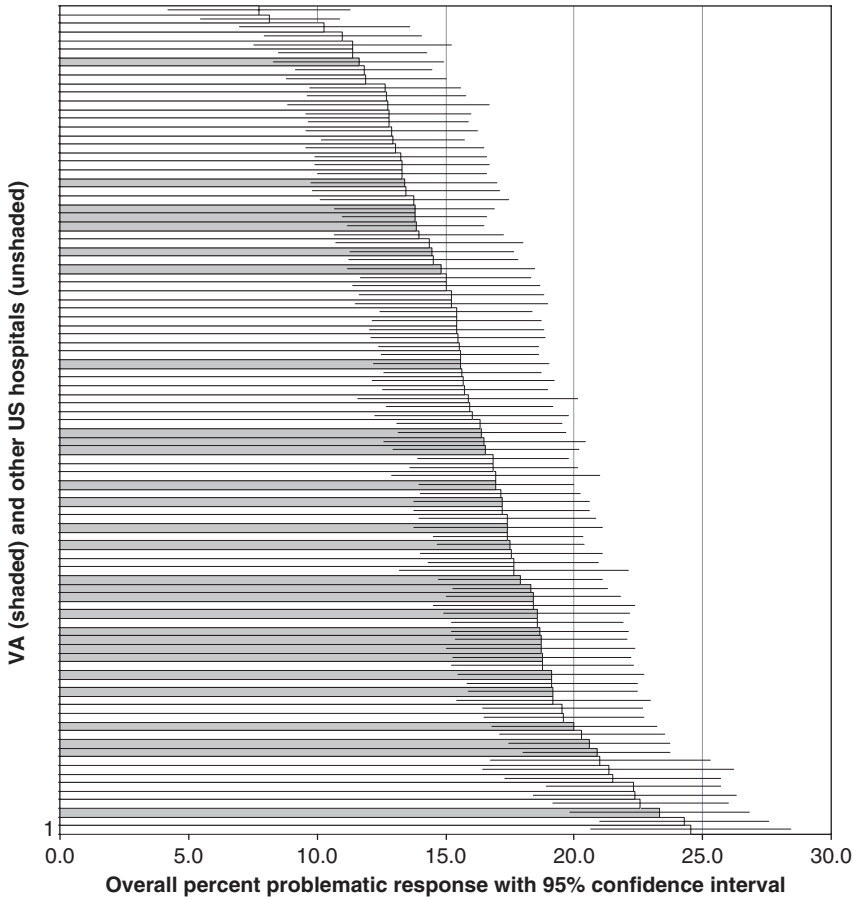
Comparison of Sample Characteristics

All comparisons between the demographic characteristics of U.S. and VA samples revealed statistically significant differences ($p < .001$; Table 1). Respondents in U.S. hospitals were considerably younger than those in VA hospitals. U.S. hospital personnel also worked less time at their facility and were less likely to be male. They were more likely than VA personnel to work in HHUs and to be nurses and senior managers. In the U.S. sample, respondents more often worked in hospitals from the West, categorized as large, and with higher nurse staffing ratios. They were less often in major teaching hospitals and urban areas.

Safety-Climate Perceptions by Hospital

The graph displays the overall PPR and 95 percent confidence interval for each of the 96 hospitals in our study, displayed from lowest PPR (best safety climate) to highest (worst safety climate), differentiating between U.S. and VA hospitals. The range in safety climate among U.S. hospitals is larger than among VA hospitals, based on point estimates. In U.S. hospitals PPR varied from 7.7 to 24.5 percent, and in VA hospitals the range was 11.6–23.3 percent. More than twice as many VA hospitals fell in the bottom half of the distribution ($n = 21$) than in the top ($n = 8$). The results, however, place individual VA hospitals among both the top 10 and bottom 10 hospitals surveyed, and uncertainty in the point estimates suggests few meaningful differences between U.S. and VA hospitals (Figure 1).

Figure 1: Overall Percent Problematic Response (PPR) by Hospital, U.S. and VA Hospitals 2006



Comparison of Safety-Climate Perceptions

Table 2 presents survey results by item and dimension. Cronbach's α's for all dimensions were within an acceptable range (0.6–0.8) except for "Fear of Shame" (0.4) and "Fear of Blame and Punishment" (0.5). The low reliabilities for the latter two scales reflect the reduced number of common items remaining after dropping those items that were not phrased identically in the U.S. and VA surveys. Results for these dimensions are presented because the domains represent potentially important aspects of safety climate; however, they should be regarded as tentative and interpreted with caution.

Table 2: Mean Percent Problematic Response (PPR) among All Respondents by Item and Dimension: U.S. and VA*

	Problematic Response Rate (Cronbach's α Coefficient)	
Dimensions† *and Text of Item*	*U.S.*	*VA*
Hospital contributions to safety climate		
Senior managers' engagement	14.7(0.80)	18.6(0.82)
Senior management has a clear picture of the risks associated with patient care.	14.2	17.7
Senior management has a good idea of the kinds of mistakes that actually occur in this facility.	17.6	20.4
Senior management supports a climate that promotes patient safety.	9.6	12.2
Senior management considers patient safety when program changes are discussed.	10.8	14.6
Patient safety decisions are made by the most qualified people, regardless of rank or hierarchy.	22.1	29.7
Good communication flow exists up and down the chain of command regarding patient safety issues.	13.8	17.2
Organizational resources for safety	13.6(0.65)	16.4(0.63)
I have enough time to complete patient care tasks safely.	18.1	20.5
I am provided with adequate resources (personnel, budget, and equipment) to provide safe patient care.	17.9	24.4
I have received sufficient training to enable me to address patient safety problems.	6.9	8.1
This facility devotes sufficient resources to follow up on identified safety problems.	11.4	12.5
Overall emphasis on patient safety	9.0(0.57)	9.7(0.57)
Compared with other facilities in the area, this facility cares more about the quality of patient care it provides.	10.6	11.0
Overall, the level of patient safety at this facility is improving.	7.4	8.4
Work-unit contributions to safety climate		
Unit managers' support	18.8(0.59)	21.7(0.59)
Management in my unit helps me overcome problems that make it hard for me to provide safe patient care.	19.4	25.1
In my unit, management puts safety at the same level of importance as meeting the schedule and productivity.	19.8	19.6
Whenever pressure builds up, management in my unit wants us to work faster, even if it means taking shortcuts that might negatively affect patient safety.	17.2	20.5
Unit safety norms	9.9(0.57)	10.6(0.61)
My unit takes the time to identify and assess risks to ensure patient safety.	6.7	7.9

continued

Table 2. *Continued*

Dimensions[†] *and Text of Item*	*Problematic Response Rate (Cronbach's α Coefficient)* U.S.	VA
My unit does a good job managing risks to ensure patient safety.	5.7	6.9
In my unit, there is significant peer pressure to discourage unsafe patient care.	23.2	20.6
In my unit, anyone found to intentionally violate standards or safety rules is corrected.	7.2	10.0
Deliberate violations of standard operating procedures are rare in my unit.	6.6	7.5
Unit recognition and support for safety efforts	28.7(0.63)	31.0(0.64)
My unit recognizes safety achievement through rewards and incentives.	47.8	48.3
I am rewarded for taking quick action to identify a serious mistake.	34.8	39.1
My unit provides training on teamwork in order to improve patient care performance and safety.	22.0	24.4
My performance is evaluated against defined safety standards.	10.8	12.0
Collective learning	8.9(0.69)	10.3(0.70)
Mistakes have led to positive changes in my unit.	9.8	11.7
On my unit, we identify and fix safety problems before an incident actually occurs.	9.4	11.3
Our process of accident and incident investigation is effective at identifying root causes.	9.6	10.7
In my unit, patient safety problems and errors are communicated to the right people so that the problem can be corrected.	6.9	7.4
Psychological safety	12.2(0.63)	14.4(0.65)
Staff feel comfortable questioning the actions of those with more authority when patient safety is at risk.	19.0	22.1
Staff freely speak up if they see something that may negatively affect patient care.	10.3	11.2
I am comfortable reporting safety concerns without fear of being punished by management.	7.3	9.7
Problem responsiveness	12.5(0.69)	15.2(0.70)
Bringing patient safety concerns to management's attention usually results in the problem being addressed.	13.8	17.4
When I take the time to communicate about patient safety problems there is appropriate follow-up.	11.1	13.0
Interpersonal contributions to safety climate		
Fear of shame	4.8(0.44)	5.3(0.41)
Asking for help is a sign of incompetence.	5.9	6.1

continued

Table 2. *Continued*

Dimensions[†] *and Text of Item*	*Problematic Response Rate (Cronbach's α Coefficient)* U.S.	VA
If I make a mistake that has significant consequences and nobody notices, I do not tell anyone about it.	3.6	4.4
Fear of blame and punishment	32.2(0.54)	23.1(0.53)
If people find out that I made a mistake, I will be disciplined.	35.6	25.9
Clinicians who make serious mistakes are usually punished.	28.8	20.4
Other aspects of safety climate		
Provision of safe care	36.4(0.67)	36.0(0.63)
In the last year, I have witnessed a coworker do something that appeared to me to be unsafe for the patient.	31.1	29.7
I have never witnessed a coworker do something that appeared to me to be unsafe patient care.	41.7	42.2
Overall average[‡]	15.9	17.2

*All means were calculated using weights.

[†]Mean of all items in dimension averaged to calculate dimension mean.

[‡]Overall means for U.S. and VA not significantly different from each other (*t*-test, $p = .83$).

VA, Veterans Health Administration.

The overall average PPR (i.e., mean of individual item means) was not significantly different in U.S. (mean = 15.9, SD = 1.61) and VA hospitals (mean = 17.2, SD = 1.56; $p = .55$). For 10 of the 12 individual dimensions, mean PPR was lower in U.S. than in VA hospitals. However, a smaller percentage of VA than U.S. respondents indicated fear of blame or punishment and that they witnessed or participated in unsafe care.

Relationship of Safety Climate and Sample Characteristics

All respondent characteristics, with the exception of time worked in the hospital, related significantly and in the same direction to safety climate overall in both samples (Table 3a). Being male and a nurse were positively related to PPR (worse safety climate), while age >50, being a senior manager and working in an HHU were negatively correlated with PPR (better safety climate). The magnitude of these correlations, however, differed somewhat by sample. PPR among men was higher than among women by 0.8 and 2.6 percentage points in U.S. and VA hospitals, respectively. PPR among senior managers was lower than among nonsenior managers by 4.6 percentage points in U.S. hospitals and by 7.3 percentage points in the VA.

Table 3a: Association of Individual Mean Percent Problematic Response (PPR) with Individual and Facility Characteristics

	Hierarchical Linear Models Coefficient (SE)			
Variable	*U.S.*		*VA*	
Individual characteristics				
Male	0.843*	− 0.37	2.628**	− 0.63
Age > 50 years	− 1.990**	− 0.34	− 1.724**	− 0.6
Time at facility > 10 years	0.253	− 0.33	1.115	− 0.6
Being nurse	3.717**	− 0.34	4.616**	− 0.66
Being senior manager	− 4.596**	− 0.87	− 7.250**	− 1.82
Employment in HHU	3.078**	− 0.38	2.739**	− 1.01
Facility characteristics				
Region South	− 0.947*	− 0.42	2.970**	− 0.93
Region Midwest	− 2.596**	− 0.47	− 1.401	− 0.91
Region East	− 1.547**	− 0.4	− 0.919	− 0.88
Bedsize (large)	1.731**	− 0.36	1.068	− 0.72
Urban location	− 0.381	− 0.4	2.860**	− 1.08
Nurse staffing ratio	− 0.119**	− 0.04	0.051	− 0.1
Intercept	15.71**	− 0.58	10.88**	− 1.94

**$p < .01$, *$p < .05$.

HHU, high hazard unit; SE, standard error; VA, Veterans Health Administration.

In contrast, characteristics of the facilities in which respondents worked related to safety climate in considerably different ways in the two samples. For U.S. hospitals, all facility characteristics with the exception of urban location related significantly to safety climate. For example, larger size was associated with higher PPR while a higher nurse staffing ratio was associated with lower PPR. In the VA, only working in the South and in an urban location were significantly correlated with safety climate, and in both instances the direction of correlation was opposite that of U.S. hospitals. VA employees working in the South had higher PPR than VA employees in the West. However, employees in Western U.S. hospitals had higher PPR than all other regions. VA employees working in urban hospitals had higher PPR than employees working in nonurban locations; the opposite was true for U.S. hospital employees.

Oaxaca–Blinder Decomposition of Safety-Climate Results

The Oaxaca–Blinder decomposition analysis allowed us to quantify the extent to which the predicted difference in PPR between the U.S. and VA hospitals

was due to (a) variation in the two systems' distributions of sample characteristics (sample characteristics component) and (b) differences in the U.S. and VA model characteristics as expressed by the values of the coefficients (unexplained component). The model calculated the difference between sample characteristics and unexplained components based on the coefficients from the U.S. hospital model for each variable and the difference between the distributions of each characteristic for each sample. That is, $\hat{Y}_{\mathrm{VA}} - \hat{Y}_{\mathrm{U.S.}} = \hat{\beta}_{\mathrm{U.S.}}(X_{\mathrm{VA}} - X_{\mathrm{U.S.}})$ gives the effect of the difference in characteristic X between U.S. and VA hospitals on the predicted PPR, using the U.S. hospital model (i.e., $\hat{\beta}_{\mathrm{U.S.}}$) as the reference.

The net difference in safety climate, based on the predicted means of the U.S. and VA models, was 1.39 higher predicted PPR for the VA (Table 3b). Some of this difference was attributable to observed sample characteristics. For example, male respondents had higher PPR in VA than U.S. hospitals

Table 3b: Differences in U.S. and VA Safety Climate, Decomposition Results

	Mean	*[95% CI]/(SE)*
Predicted mean PPR for		
Non-VA	15.3	[16.1–17.2]
VA	16.7	[14.9–15.6]
Total difference between U.S. and VA (based on predicted means)	1.39**	− 0.33
Difference attributable to variation in†		
Male	0.108*	− 0.048
Age > 50 years	− 0.298**	− 0.054
Time at facility > 10 years	0.036	− 0.046
Being nurse	− 0.115**	− 0.034
Being senior manager	0.032*	− 0.015
Employment in HHU	− 0.373**	− 0.05
Region South	0.041*	− 0.02
Region Midwest	− 0.194**	− 0.04
Region East	− 0.147**	− 0.041
Bedsize (large)	− 0.028	− 0.017
Urban location	− 0.066	− 0.07
Nurse staffing ratio	0.237**	− 0.07
Difference attributable to variation in sample characteristics	− 0.766**	− 0.16
Difference attributable to differences in model coefficients (differential effects of sample characteristics)	2.160**	− 0.35

**$p < .01$, *$p < .05$.

†These use U.S. model estimates from Table 3a as reference.

CI, confidence interval; HHU, high hazard unit; PPR, percent problematic response; SE, standard error; VA, Veterans Health Administration.

(Table 3a). Because the VA also had more male respondents, the result was a higher predicted PPR for the VA sample—on average by 0.108 percentage points (Table 3b). On the other hand, the impact of the VA sample having more respondents older than 50 than the U.S. sample was that the VA's predicted PPR for this factor was −0.298, because this characteristic was associated with lower PPR. The aggregate impact of variation between U.S. and VA hospitals in the distribution of sample characteristics was −0.766, suggesting that the predicted VA PPR should be 0.77 percentage points lower than that for U.S. hospitals (based on the U.S. hospital sample as the reference). However, the unexplained component accounted for a larger portion of the difference in safety climate between U.S. and VA hospitals than the sample characteristics component. The differential *effects* of observed characteristics (i.e., differences in model coefficients) plus differences in unobserved characteristics predicted average PPR in the VA to be 2.160 percentage points higher than in U.S. hospitals.

DISCUSSION

This study is the first to compare hospital safety climate between two fundamentally different sets of hospitals: one a nationally integrated hospital network, the other predominantly independent general acute-care hospitals. The study summarizes safety climate in the VA and other U.S. hospitals and factors influencing safety climate in each setting. Results also show how sample characteristics contribute to differences in safety climate between settings.

Overall, we found no difference in safety climate between U.S. and VA hospitals on average, based on descriptive statistics. Differences with respect to specific dimensions were significant, generally favoring U.S. hospitals. However, the range in safety-climate results among U.S. hospitals substantially overlapped, suggesting that neither population has achieved superior safety climate. In addition, relative to high reliability organizations, such as naval aviation, which serve as the "gold standard" for safety achievement despite hazardous and demanding conditions, safety climate in both U.S. and VA settings was considerably worse (Gaba et al. 2003). This finding does not support our first hypothesis, that participating in a nationally integrated hospital network would be associated with stronger safety climate. It appears that potential advantages associated with the system's intense focus on safety improvement and its ability to implement uniformly its improvement program may have been outweighed by local considerations. While institutional pro-

grams may facilitate the ability of local managers to improve safety, they may not be targeted closely enough to the actual challenges of the workplace to make a difference alone.

We also found that characteristics of individuals influenced safety climate consistently across settings when controlling for other factors. Older age and more seniority corresponded to more positive perceptions of safety climate, while working as a nurse or in an HHU were associated with more negative perceptions. These findings are consistent with studies showing perceptions of safety climate differ by workgroup and management level (Pronovost et al. 2003; Sexton et al. 2006c; Singer et al. 2008a, 2009). In contrast, facility characteristics influenced safety climate differently in U.S. and VA samples. Working in southern and urban facilities corresponded with higher PPR among VA employees and lower PPR in the U.S. sample. Other studies have found similarly mixed results regarding effects of geographical and structural characteristics within non-VA hospitals (Baldwin et al. 2004; Coburn et al. 2004; Loux, Payne, and Knott 2005; Longo et al. 2007). Also consistent with prior studies (Aiken et al. 2002; Stone et al. 2007; Weissman et al. 2007), we found that higher nurse staffing ratios were associated with lower PPR in U.S. hospitals.

Decomposition analysis examined the influence of (1) variation in the distribution of observed sample characteristics among personnel in an integrated network compared with other U.S. hospitals and (2) differential *effects* of sample characteristics in each group. The overall difference between the samples, that is, the influence of (1) and (2) together, was a 1.4 percentage point higher PPR for the VA. We hypothesized that variations in sample characteristics between settings would explain more of this difference in safety climate than would differences in effects of those sample characteristics. Our results do not support this hypothesis. Instead, it was the differential effects of sample characteristics that explained more of the difference in safety climate between U.S. and VA hospitals. The difference based on the distribution of all the VA sample characteristics compared with U.S. characteristics was negative, indicating that the VA would be expected to have a 0.77 percentage point lower PPR based on observed sample characteristics alone. The unexplained difference, indicating the differential *effect* of sample characteristics, was 2.2 percentage points higher PPR in VA than in U.S. hospitals. This second difference was driven primarily by two factors: region and location, both of which act in opposite directions on PPR in the U.S. and VA models, and by unobserved characteristics. Decomposition of the residual suggests that our model explained just 5.9 percent of the variation in the outcome measure. Future

research should explore additional characteristics of hospitals and factors driving the effects of region and location in order to determine whether some modifiable factors may be involved that could provide leverage for change.

Our results suggest that characteristics of respondents and their work facilities influence safety-climate scores. Thus, in comparing safety climate among hospitals or over time in hospitals whose respondent characteristics may have changed, it is important to include known characteristics in analyses. Such longitudinal studies would also provide opportunity for research on how the effects of respondent characteristics on PPR change over time.

Results should be interpreted within the context of several limitations. This was a cross-sectional study; thus, we cannot make assertions about causality. We cannot explain the mechanisms underlying effects of various factors on safety climate. Nor can we differentiate the effect on safety climate of observed from unobserved characteristics in the unexplained component of the difference between samples. We cannot rule out nonresponse bias as a factor in our results. The methodology in both settings aimed to maximize response rates while maintaining the voluntary and anonymous nature of the surveys. While the VA sample achieved a response rate that is similar to that of other studies of this type (Asch, Jedrziewski, and Christakis 1997; Jepson et al. 2005), the overall response rate in the U.S. sample was lower. We adjusted for nonresponse and sampling bias through the use of weights in our analysis; however, it is possible that results do not accurately represent the facilities or populations intended. A related issue is the representativeness of the hospitals in each sample. We conducted a stratified random sampling strategy in both settings, but since participation was voluntary, sampled facilities may differ from facilities in their respective populations in unanticipated ways. As noted, administration dates and recruitment and sampling strategies also differed slightly between U.S. and VA samples. Although recruited on the basis of size and region rather than PSI rates, those rates among the U.S. hospital sample did not differ from those of U.S. hospitals overall. In addition, within the U.S. hospital sample we found no difference when we compared overall mean PPR between over-sampled hospitals and the other hospitals in that sample. Finally, while our models included variables associated with safety climate in the literature, we were limited by variables available in our datasets.

Nevertheless, the methodology employed in our study represents an advance over prior research. In particular, the decomposition analysis provides information about systematic differences in sample characteristics and the effects of specific characteristics on safety climate in different settings. By

achieving a more thorough understanding of what is driving apparent differences in safety-climate survey results among hospitals we can proceed more clearly toward developing effective improvement interventions.

The results presented suggest that continued efforts are needed to improve safety climate in hospitals. While participation in systems can provide some advantages in this regard, the large unexplained component of safety climate from the regression estimates suggests that other factors, such as hospitals' emphasis on creativity and innovation and their leaders' abilities to motivate, implement, and sustain improvement, may matter more.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Financial support for this research was provided by the VA Health Services Research and Development Service grant no. IIR 03-303-2. The authors would also like to acknowledge research support from Ms. Alyson Falwell.

Disclosures: None.

Disclaimers: None.

REFERENCES

Aiken, L. H., S. P. Clarke, D. M. Sloane, J. Sochalski, and J. H. Silber. 2002. "Hospital Nurse Staffing and Patient Mortality, Nurse Burnout, and Job Dissatisfaction." *Journal of the American Medical Association* 288 (16): 1987–93.

Asch, D., M. Jedrziewski, and N. Christakis. 1997. "Response Rates to Mail Surveys Published in Medical Journals." *Journal of Clinical Epidemiology* 50 (10): 1129–36.

Baldwin, L. M., R. F. MacLehose, L. G. Hart, S. K. Beaver, N. Every, and L. Chan. 2004. "Quality of Care for Acute Myocardial Infarction in Rural and Urban US Hospitals." *Journal of Rural Health* 20 (2): 99–108.

Blinder, A. S. 1973. "Wage Discrimination: Reduced Form and Structural Variables." *Journal of Human Resources* 8: 436–55.

Coburn, A. F., M. Wakefield, M. Casey, I. Moscovice, S. Payne, and S. Loux. 2004. "Assuring Rural Hospital Patient Safety: What Should Be the Priorities?" *Journal of Rural Health* 20 (4): 314–26.

Cohen, M., N. Kimmel, M. Benage, C. Hoang, T. Burroughs, and C. Roth. 2004. "Implementing a Hospitalwide Patient Safety Program for Cultural Change." *Joint Commission Journal on Quality and Safety* 30 (8): 424–31.

Colla, J., A. Bracken, L. Kinney, and W. Weeks. 2005. "Measuring Patient Safety Climate: A Review of Surveys." *Quality and Safety in Health Care* 14 (5): 364–6.

Flin, R., C. Burns, K. Mearns, S. Yule, and E. Robertson. 2006. "Measuring Safety Climate in Health Care." *Quality and Safety in Health Care* 15 (2): 109–15.

Gaba, D., S. Singer, A. Sinaiko, and J. Bowen. 2003. "Differences in Safety Climate between Hospital Personnel and Navy Aviators." *Human Factors* 45 (2): 173–85.

Gandhi, T., E. Graydon-Baker, C. Huber, A. Whittemore, and M. Gustafson. 2005. "Closing the Loop: Follow-Up and Feedback in a Patient Safety Program." *Joint Commission Journal on Quality and Patient Safety* 31 (11): 614–21.

Hartmann, C., A. Rosen, M. Meterko, P. Shokeen, S. Zhao, S. Singer, A. Falwell, and D. Gaba. 2008. "An Overview of Patient Safety Climate in the VA." *Health Services Research* 43 (4): 1263–84.

Hofmann, D., and F. Morgeson. 1999. "Safety-Related Behavior as a Social Exchange: The Role of Perceived Organizational Support and Leader–Member Exchange." *Journal of Applied Psychology* 84 (2): 286–96.

Hofmann, D. A., and B. Mark. 2006. "An Investigation of the Relationship between Safety Climate and Medication Errors as Well as Other Nurse and Patient Outcomes." *Personnel Psychology* 59 (4): 847–69.

Hudson, J. L., G. E. Miller, and J. B. Kirby. 2007. "Explaining Racial and Ethnic Differences in Children's Use of Stimulant Medications." *Medical Care* 45 (11): 1068–75.

Jann, B. 2008. "The Blinder–Oaxaca Decomposition." ETH Zurich Sociology Working Paper No. 5 [accessed on June 9, 2009]. Available at http://ideas.repec.org/p/ets/wpaper/5.html

Jepson, C., D. Asch, J. Hershey, and P. Ubel. 2005. "In a Mailed Physician Survey, Questionnaire Length Had a Threshold Effect on Response Rate." *Journal of Clinical Epidemiology* 58 (1): 103–5.

Joint Commission on Accreditation of Health Care Organizations. 2002. *Hospital Accreditation Standards*. Oakbrook Terrace, IL: Joint Commission Resources.

Kirby, J. B., G. Taliaferro, and S. H. Zuvekas. 2006. "Explaining Racial and Ethnic Disparities in Health Care." *Med Care* 44 (5, suppl): I64–72.

Longo, D. R., J. E. Hewett, B. Ge, and S. Schubert. 2007. "Hospital Patient Safety: Characteristics of Best-Performing Hospitals." *Journal of Healthcare Management* 52 (3): 188–204; discussion 04–5.

Loux, S., S. Payne, and A. Knott. 2005. "Comparing Patient Safety in Rural Hospitals by Bed Count." In *Advances in Patient Safety: From Research to Implementation*, Vol. 1, edited by K. Henriksen, J. B. Battles, and D. I. Lewin, pp. 391–404. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ) and Department of Defense (DoD).

Makary, M., J. Sexton, J. Freischlag, C. Holzmueller, E. Millman, L. Rowen, and P. Pronovost. 2006. "Operating Room Teamwork among Physicians and Nurses: Teamwork in the Eye of the Beholder." *Journal of the American College Surgeons* 202 (5): 746–52.

Naveh, E., T. Katz-Navon, and Z. Stern. 2005. "Treatment Errors in Healthcare: A Safety Climate Approach." *Management Science* 51 (6): 948–60.

Neal, A., and M. Griffin. 2006. "A Study of the Lagged Relationships among Safety Climate, Safety Motivation, Safety Behavior, and Accidents at the Individual and Group Levels." *Journal of Applied Psychology* 91 (4): 946–53.

Oaxaca, R. 1973. "Male–Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14: 693–709.

Pronovost, P., B. Weast, C. Holzmueller, B. Rosenstein, R. Kidwell, K. Haller, E. Feroli, J. Sexton, and H. Rubin. 2003. "Evaluation of the Culture of Safety: Survey of Clinicians and Managers in an Academic Medical Center." *Quality and Safety in Health Care* 12 (6): 405–10.

Rosen, A. K., D. M. Gaba, M. Meterko, P. Shokeen, S. Singer, S. Zhao, A. Labonte, and A. Falwell. 2008. "Recruitment of Hospitals for a Safety Climate Study: Facilitators and Barriers." *Joint Commission Journal on Quality and Patient Safety* 34 (5): 275–84.

Sexton, J., R. Helmreich, T. Neilands, K. Rowan, K. Vella, J. Boyden, P. Roberts, and E. Thomas. 2006a. "The Safety Attitudes Questionnaire: Psychometric Properties, Benchmarking Data, and Emerging Research." *BMC Health Services Research* 6 (44).

Sexton, J., C. Holzmueller, P. Pronovost, E. Thomas, S. McFerran, J. Nunes, D. Thompson, A. Knight, D. Penning, and H. Fox. 2006b. "Variation in Caregiver Perceptions of Teamwork Climate in Labor and Delivery Units." *Journal of Perinatology* 8: 463–70.

Sexton, J., M. Makary, A. Tersigni, D. Pryor, A. Hendrich, E. Thomas, C. Holzmueller, A. Knight, Y. Wu, and P. Pronovost. 2006c. "Teamwork in the Operating Room: Frontline Perspectives among Hospitals and Operating Room Personnel." *Anesthesiology* 105 (5): 887–94.

Shen, Y. C., and S. K. Long. 2006. "What's Driving the Downward Trend in Employer-Sponsored Health Insurance?" *Health Services Research* 41 (6): 2074–96.

Singer, S., A. Falwell, D. Gaba, and L. Baker. 2008a. "Hospital Patient Safety Climate: Variation by Management Level." *Medical Care* 46 (11): 1149–56.

Singer, S., D. Gaba, A. Falwell, S. Lin, J. Hayes, and L. Baker. 2009. "Patient Safety Climate in 92 US Hospitals: Differences by Work Area and Discipline." *Medical Care* 47 (1): 23–31.

Singer, S., M. Meterko, L. Baker, D. Gaba, A. Falwell, and A. Rosen. 2007. "Workforce Perceptions of Hospital Safety Culture: Development and Validation of the Patient Safety Climate in Healthcare Organizations Survey." *Health Services Research* 42 (5): 1999–2021.

Singer, S. J., D. M. Gaba, J. J. Geppert, A. D. Sinaiko, S. K. Howard, and K. C. Park. 2003. "The Culture of Safety in California Hospitals." *Quality and Safety in Health Care* 12 (2): 112–8.

Singer, S. J., S. Lin, A. Falwell, D. Gaba, and L. C. Baker. 2008b. "Relationship of Safety Climate and Safety Performance in Hospitals." *Health Services Research* 44 (2, Part I): 399–421.

Snijders, T., and R. Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* Thousand Oaks, CA: SAGE Publications.

Sorra, J., V. Nieva, T. Famolaro, and N. Dyer. 2009. *Hospital Survey on Patient Safety Culture: 2007 Comparative Database Report. (Prepared by Westat, Rockville, MD, under contract No. 233-02-0087, Task Order No. 18.) AHRQ Publication No. 07-0025.* Rockville, MD: Agency for Healthcare Research and Quality.

Stone, P. W., C. Mooney-Kane, E. L. Larson, T. Horan, L. G. Glance, J. Zwanziger, and A. W. Dick. 2007. "Nurse Working Conditions and Patient Safety Outcomes." *Medical Care* 45 (6): 571–8.

Thomas, E., J. Sexton, and R. Helmreich. 2003. "Discrepant Attitudes about Teamwork among Critical Care Nurses and Physicians." *Critical Care Medicine* 31 (3): 956–9.

Vogus, T. J., and K. M. Sutcliffe. 2007. "The Safety Organizing Scale: Development and Validation of a Behavioral Measure of Safety Culture in Hospital Nursing Units." *Medical Care* 45 (1): 46–54.

Weingart, S., K. Farbstein, R. Davis, and R. Phillips. 2004. "Using a Multihospital Survey to Examine the Safety Culture." *Joint Commission Journal on Quality and Safety* 30 (3): 125–32.

Weissman, J. S., J. M. Rothschild, E. Bendavid, P. Sprivulis, E. F. Cook, R. S. Evans, Y. Kaganova, M. Bender, J. David-Kasdan, P. Haug, J. Lloyd, L. G. Selbovitz, H. J. Murff, and D. W. Bates. 2007. "Hospital Workload and Adverse Events." *Medical Care* 45 (5): 448–55.

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.

ORIGINAL ARTICLES

Primary Care Validation of a Single-Question Alcohol Screening Test

Peter C. Smith, MD, MSc[1,5], Susan M. Schmidt[1], Donald Allensworth-Davies, MSc[2], and Richard Saitz, MD, MPH[3,4]

[1]Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA; [2]Data Coordinating Center, Boston University School of Public Health, Boston, MA, USA; [3]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA; [4]Youth Alcohol Prevention Center and Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; [5]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA.

BACKGROUND: Unhealthy alcohol use is prevalent but under-diagnosed in primary care settings.

OBJECTIVE: To validate, in primary care, a single-item screening test for unhealthy alcohol use recommended by the National Institute on Alcohol Abuse and Alcoholism (NIAAA).

DESIGN: Cross-sectional study.

PARTICIPANTS: Adult English-speaking patients recruited from primary care waiting rooms.

MEASUREMENTS: Participants were asked the single screening question, "How many times in the past year have you had X or more drinks in a day?", where X is 5 for men and 4 for women, and a response of >1 is considered positive. Unhealthy alcohol use was defined as the presence of an alcohol use disorder, as determined by a standardized diagnostic interview, or risky consumption, as determined using a validated 30-day calendar method.

MAIN RESULTS: Of 394 eligible primary care patients, 286 (73%) completed the interview. The single-question screen was 81.8% sensitive (95% confidence interval (CI) 72.5% to 88.5%) and 79.3% specific (95% CI 73.1% to 84.4%) for the detection of unhealthy alcohol use. It was slightly more sensitive (87.9%, 95% CI 72.7% to 95.2%) but was less specific (66.8%, 95% CI 60.8% to 72.3%) for the detection of a current alcohol use disorder. Test characteristics were similar to that of a commonly used three-item screen, and were affected very little by subject demographic characteristics.

CONCLUSIONS: The single screening question recommended by the NIAAA accurately identified unhealthy alcohol use in this sample of primary care patients. These findings support the use of this brief screen in primary care.

KEY WORDS: alcohol screening test; alcoholics; primary care validation; NIAAA.

Received August 17, 2008
Revised January 14, 2009
Accepted January 23, 2009
Published online February 27, 2009

J Gen Intern Med 24(7):783–8
DOI: 10.1007/s11606-009-0928-6

INTRODUCTION

Unhealthy alcohol use, the spectrum from risky consumption to the alcohol use disorders, alcohol abuse and dependence, is prevalent in the primary care setting and is under-diagnosed[1]. Screening and brief intervention by primary care physicians for those with unhealthy alcohol use reduces risky consumption[2]. Because of this, practice guidelines recommend universal screening[3]. Time is limited, however, and commonly-used alcohol screening instruments are comprised of multiple questions, often do not cover the full spectrum of unhealthy use, can be time consuming to administer and may require scoring[4,5]. Consequently, many patients are not screened[6,7]. Single-question screening tests for unhealthy alcohol use may help to increase the frequency of screening in primary care. The National Institute on Alcohol Abuse and Alcoholism (NIAAA) recommends, in its clinician's guide, one such single-question screen for unhealthy alcohol use[8]. The recommended question asks "How many times in the past year have you had X or more drinks in a day?" (where X is 5 for men and 4 for women, and a response of $\geq$ 1 is considered positive). While similar single-question screens (which used different phrasing, alcohol quantity and time cutoffs) have been validated in various settings, the NIAAA recommended screening test has not been validated in the primary care setting[9–12]. Because of the wide dissemination of this guide and practice recommendation, we attempted to validate this version of the screening question in a sample of primary care patients.

SUBJECTS AND METHODS

Subjects

Subjects were selected by a research associate who systematically approached patients in the waiting room of a primary care clinic in an urban safety net hospital. 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 under the age of 18 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 not themselves being patients of the clinic were also excluded. The Institutional Review Board of Boston University Medical Center reviewed and approved all study procedures.

Data Collection

Interviews were conducted by trained research staff in a private setting and data were recorded anonymously, unaccompanied by any unique identifiers.

Screening Question. In accordance with the strategy recommended in the NIAAA Clinician's Guide, subjects were first asked a pre-screening question, "Do you sometimes drink alcoholic beverages?", and then the single screening question, "How many times in the past year have you had X or more drinks in a day?" (where X is 5 for men and 4 for women, and a response of $\geq$1 is considered positive). Subjects responding negatively to the pre-screening question were still asked the single screening question. If asked to clarify, the research associate provided definitions of a standard drink (12 ounces of beer, 5 ounces of wine, or 1.5 ounces of 80 proof spirits). For comparison purposes, the three-item Alcohol Use Disorders Identification Test-Consumption (AUDIT-C) was administered following the single-question screen and before the other assessments[13]. After this, alcohol consumption and the presence or absence of an alcohol use disorder were assessed using reference standards.

Risky Alcohol Consumption Amounts. A validated calendar-based method (the timeline followback) was used for the measurement of alcohol consumption[14]. Using this method, subjects estimated the amount of alcohol consumed on each of the 30 days preceding the interview (summarized in analyses using the same standard drink definitions that appear above). Subjects were considered to have consumed risky amounts of alcohol if their average weekly alcohol intake over the preceding 30 days exceeded recommended limits (>14 drinks per week for men and >7 drinks per week for women) or if they reported exceeding recommended daily limits (>4 drinks per occasion for men and >3 drinks per occasion for women) on any of the 30 days[15]. Average weekly alcohol intake was calculated by multiplying by 7 the average number of standard drinks consumed per day during the 30 days.

Alcohol Related Problems. Subjects were then asked if they had ever experienced any of a list of problems related to alcohol use, from the Short Inventory of Problems (SIP)[16]. Subjects were considered to have alcohol related problems if they consumed risky amounts of alcohol and responded positively to any of the 15 SIP questions.

Alcohol Use Disorders. The computerized version of the Composite International Diagnostic Interview (CIDI) Substance Abuse Module was used for the assessment of current (12-month) alcohol use disorders (abuse and dependence)[17]. This structured questionnaire was administered by the research assistant, and subject responses were recorded electronically. The responses were then analyzed, using an algorithm, to yield a Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnosis of alcohol abuse or dependence. Subjects with alcohol abuse or dependence as determined by the CIDI and who reported experiencing symptoms within the past 12 months were considered to have a current alcohol use disorder.

Unhealthy Alcohol Use. Subjects with unhealthy alcohol use either consumed risky amounts of alcohol (with or without associated alcohol problems), or had a current alcohol use disorder based on (as defined above) the timeline followback and the CIDI, respectively.

STATITICAL ANALYSIS

We calculated the sensitivity (proportion of subjects with the condition of interest who tested positive), specificity (proportion of subjects without the condition of interest who tested negative) and likelihood ratios for the single-question screen for the detection of risky alcohol consumption amounts, risky consumption associated with problems, a current alcohol use disorder, or for unhealthy alcohol use in general (either the consumption of risky amounts or a disorder, the usual target for universal screening). A positive likelihood ratio is determined by dividing the sensitivity by (1 - specificity), while a negative likelihood ratio is (1 - sensitivity) divided by the specificity. The NIAAA clinician's guide recommends a two-step screening process: subjects are asked if they sometimes drink alcoholic beverages, and only those who answer affirmatively are asked the screening question. In the main analysis we determined the test characteristics of this approach. In a sensitivity analysis we also determined the test characteristics of the single-question alone, without regard to the response to the pre-screening question. For comparison with the single-question screen, we calculated the sensitivity, specificity and likelihood ratios of the AUDIT-C for the detection of the same conditions. The AUDIT-C, which consists of three items, each with four possible responses, yields a score between 0 and 12. A total of more than three points is considered a positive test[18]. We calculated 95% confidence intervals using published formulas[19]. Statistical analyses were performed using Version 9.1 of the SAS System (copyright SAS Institute Inc.).

RESULTS

Subject recruitment. Of the 1,781 people approached, 903 (51%) agreed to be screened for study eligibility (Fig. 1). Of these, 509 (56%) 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, 4 (1%) refused to participate and 87 (22%) did not show up for the planned interview after the visit with their physician. Of the 303 subjects who arrived and gave consent to participate, 3 (1%) were unable to complete the interview. The data of 14 subjects

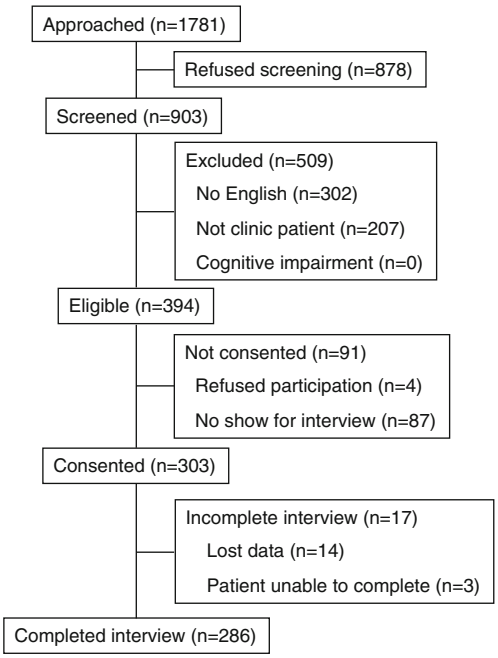


Figure 1. Recruitment of subjects.

(5%) were lost due to an electronic error, leaving 286 subjects whose data were analyzed (73% of those eligible).

Subject Characteristics. Of the 286 subjects, 54% were women, and the median age was 49 (range 21–86) (Table 1). The majority of subjects (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. Unhealthy alcohol use was reported by 31% of subjects: 6% consumed risky amounts but did not have alcohol-related problems or a disorder; 13% consumed risky amounts and had problems but no current disorder; and 12% had a current alcohol use disorder (2% of subjects reported a past year alcohol use disorder, but not past month risky consumption). The lifetime prevalence of alcohol use disorders (44%) and drug use disorders (47%) was high.

Test Characteristics. The single-question screen was 81.8% sensitive (95% CI 72.5% to 88.5%) and 79.3% specific (95% CI 73.1% to 84.4%) for the detection of unhealthy alcohol use (Table 2). It was slightly more sensitive (87.9%, 95% CI 72.7% to 95.2%) and was less specific (66.8%, 95% CI 60.8% to72.3%) for the detection of a current alcohol use disorder. The longer AUDIT-C screen was slightly less sensitive (73.9%, 95% CI 63.8% to 81.9%) for the detection of unhealthy alcohol use and slightly more specific (82.8%, 95% CI 77.0% to 87.4%) for the detection of an alcohol use disorder, but overall its test characteristics were similar to those of the single-question screen. Subject gender, ethnicity, education and primary language affected point estimates of the sensitivity and specificity of the single item screen very little, though some groups had small sample sizes and so larger differences could not be excluded (Table 3). In a sensitivity analysis, disregarding the pre-screening question response resulted in slightly increased sensitivity (87.5%, 95% CI 79.0% to 92.9%), and decreased specificity (73.7%, 95% CI 67.2% to 79.4%) for the detection of unhealthy alcohol use, when compared to analysis of responses to the single item screening question only when the patient reported sometimes using alcohol.

CONCLUSION

A single-question screen was sensitive and specific for the detection of unhealthy alcohol use in a sample of primary care patients. Its test characteristics were similar to those of a

Table 1. Subject Characteristics

Characteristic	(n = 286) (%, n)
Female	54.2 (155)
Age	
Mean ± SD	49.0 ± 12.3
Median (Range)	49.0 (21–86)
Education	
Some high school	28.3 (81)
High school graduate	37.4 (107)
Some college	20.6 (59)
College graduate	9.8 (28)
Post-graduate education	3.9 (11)
Race	
American Indian/Alaskan Native	2.8 (8)
Asian	2.5 (7)
Black or African American	62.6 (179)
Native Hawaiian/PI	1.0 (3)
White	17.1 (49)
Unknown	14.0 (40)
Hispanic or Latino ethnicity	16.1 (46)
English is first language	78.0 (223)
Drug use	
Past year drug use*	34.6 (99)
Past year drug use disorder†	12.2 (35)
Lifetime drug use disorder†	46.5 (133)
Alcohol use	
Unhealthy alcohol use‖	30.8 (88)
Risky consumption amounts‡	28.7 (82)
Without alcohol related problems or current disorder†§	6.3 (18)
With alcohol related problem, but no disorder†§	12.9 (37)
Problem use or a disorder†§	24.5 (70)
Current (12 month) alcohol use disorder†	11.5 (33)
Current alcohol abuse†	2.8 (8)
Current alcohol dependence †	8.7 (25)
Any lifetime alcohol use disorder (abuse or dependence)†	44.1 (126)
Any lifetime alcohol problems	50.0 (143)

**As part of the CIDI interview subjects are asked about their use, during the past 12 months, of illicit drugs or of prescription drugs for non-medical reasons*

†Lifetime and current alcohol and drug use disorders as determined by responses to the CIDI

‡For men, an average > 14 drinks per week over the past 30 days, or > 4 drinks on any one day during the past 30 days (for women, >7 drinks per week, or >3 drinks per occasion)

§Subjects were considered to have alcohol related problems if they consumed risky amounts of alcohol and responded positively to any of the 15 Short Inventory of Problems (SIP) questions

‖ Hazardous consumption amounts, problem use, or current disorder. Some subjects reported a current (past year) disorder but not (past month) hazardous consumption amounts

¶A positive response to any of the questions from the SIP questionnaire

Table 2. Sensitivity, Specificity and Likelihood Ratios for the Detection of Unhealthy Alcohol Use: Single Screening Question and AUDIT-C (n=286)

For detection of:	Sensitivity (95% CI)		Specificity (95% CI)	
	Single Question	AUDIT-C	Single Question	AUDIT-C
Risky consumption amounts	84% (75%, 91%)	74% (64%, 83%)	78% (72%, 84%)	81% (76%, 86%)
Alcohol related problems or disorder	84% (74%, 91%)	80% (69%, 88%)	75% (69%, 80%)	80% (74%, 85%)
Current alcohol use disorder	88% (73%, 95%)	88% (73%, 95%)	67% (61%, 72%)	72% (67%, 78%)
Unhealthy alcohol use (risky amounts or disorder)	82% (73%, 89%)	73.9% (64%, 82%)	79% (73%, 84%)	83% (77%, 87%)
For detection of:	Positive LR (95% CI)		Negative LR (95% CI)	
	Single Question	AUDIT-C	Single Question	AUDIT-C
Risky consumption amounts	3.9 (3.0, 5.2)	4.0 (2.9, 5.5)	0.2 (0.1, 0.3)	0.3 (0.2, 0.4)
Alcohol related problems or disorder	3.4 (2.6, 4.3)	4.0 (3.0, 5.4)	0.2 (0.1, 0.4)	0.3 (0.2, 0.4)
Current alcohol use disorder	2.6 (2.1, 3.3)	3.2 (2.5, 4.0)	0.2 (0.1, 0.4)	0.2 (0.1, 0.4)
Unhealthy alcohol use (risky amounts or disorder)	4.0 (3.0, 5.3)	4.3 (3.1, 6.0)	0.2 (0.1, 0.4)	0.3 (0.2, 0.4)

longer screening tool in this sample, as well as in numerous studies reported in the literature[13,20–23].

Unhealthy alcohol use is prevalent in primary care, and brief intervention in this setting effectively reduces consumption among those without dependence, and improves patient outcomes[1,2]. Lack of detection of unhealthy alcohol use, however, stands as a barrier to such treatment[6]. Time constraints in the primary care setting have been cited as a reason for non-adherence to screening and prevention guidelines in general, and for the under-diagnosis of unhealthy alcohol use specifically (according to one estimate, providing all recommended preventive services to an average primary care panel would require 7.4 hours out of each work day)[24]. Among the best validated options for alcohol screening in primary care settings are the CAGE questionnaire, the AUDIT, and the MAST[25–27]. More recently, and therefore with fewer validation studies in general care settings, researchers have tested instruments as short as single items and as long as 80 items requiring scoring algorithms and keys for interpretation[10,28]. One widely known brief screening tool, the CAGE questionnaire, while accurately identifying more severe unhealthy alcohol use (i.e. dependence), was not developed to detect risky consumption amounts or alcohol problems that are more amenable to brief interventions in primary care[29]. The MAST similarly identifies alcohol dependence and is less well validated for detecting risky use and at 25 items (or 10 items for a briefer version) does not present advantages in length[25]. The 10-item AUDIT, although well-validated for detecting risky drinking, is less well known or used by primary care physicians, likely in part because it requires scoring and it is not easily memorized for incorporation into the medical interview. The AUDIT and ASSIST may have promise as electronic record systems with decision support become more widespread (and as evidence for the validity of the ASSIST accumulates). The ASSIST has one other major limitation — it does not directly identify risky consumption amounts. The single-question screen proposed by Williams et al. is not identical to that recommended by the NIAAA but it too has proven to be accurate for identifying unhealthy alcohol use among emergency department patients, in primary care, and among respondents to a household survey[9,10,12]. In summary, in terms of brevity, ease of scoring, and validity for detecting the conditions of interest in primary care, and therefore, likely greater ease for widespread implementation as recommended by practice guidelines, the single item recommended by NIAAA appears to have favorable characteristics.

The results we report are similar to those from studies using different populations and different formulations of the single-question alcohol screen. This study adds to existing literature by validating the version recommended by the NIAAA in a sample of primary care patients — one of the main populations in which it was intended to be used. This version of the single-question screen was derived from a national household survey on alcohol use, the results of which were reported by Dawson, et al.[15]. While they did not report test characteristics, and although the subjects were not primary care patients, analysis of their published results yields a sensitivity of 89.8% and a specificity of 68.3% for the detection of a current alcohol use disorder, results which were very close to those reported in the current study. In addition to being recommended for widespread use by a health authority, the question phrasing normalizes drinking of large amounts likely increasing honesty in replies, and it directly queries amounts that are defined as risky by national guidelines. The similar single-question screen proposed by Williams et al. that used different cut-off values ('When was the last time you had more than X drinks in

Table 3. Single question Screen for the Detection of Unhealthy Alcohol Use, in Selected Subgroups

	n	Sensitivity (95% CI)	Specificity (95% CI)
Male	131	82.5% (70.6%, 90.2%)	71.6% (60.5%, 80.6%)
Female	155	80.6% (63.7%, 90.8%)	83.9% (76.4%, 89.3%)
Non-Hispanic white	45	78.6% (52.4%, 92.4%)	87.1% (71.2%, 94.9%)
Non-Hispanic black	176	79.0% (66.7%, 87.5%)	79.0% (70.8%, 85.4%)
Hispanic	46	93.3% (70.2%, 98.8%)	71.0% (53.4%, 83.9%)
English primary language	223	80.0% (69.6%, 87.5%)	77.7% (70.3%, 83.7%)
English not primary language	63	92.3% (66.7%, 98.6%)	84.0% (71.5%, 91.7%)
High school graduate	205	78.7% (66.9%, 87.1%)	79.9% (72.6%, 85.6%)
Not high school graduate	81	88.9% (71.9%, 96.2%)	77.8% (65.1%, 86.8%)

1 day,' with X=4 for women and 5 for men, and a response of less than 3 months ago considered a positive screen) yielded sensitivities of between 80% and 85% and specificities of between 70% and 77% for the detection of unhealthy alcohol use, and was validated in a sample of primary care patients by Seale, et al.[10,12]. A third formulation of the single-question screen, using the third question of the AUDIT and its multiple response options ('How often in the last year have you had 6 or more drinks on one occasion' with a response other than 'never' considered a positive screen), had a sensitivity of 77% and a specificity of 83% for the detection of unhealthy alcohol use in a sample of male veteran primary care patients, though the sensitivity was lower in a separate study of female veterans (both findings confirmed in subsequent studies of non-veterans)[11,13,20,21]. These comparisons suggest that using slightly different cut-offs or changing the phrasing of the question affects the test characteristics to only a small degree.

In order for a screening test for unhealthy alcohol use to be useful, it must be applicable to the broad range of people seen in primary care. The diversity of our subject sample allowed us to examine the effect of gender, 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. The single-question screen performed well in an urban, predominately minority population, a population different from those in which single-question screens had been tested previously. This, taken together with the results of the other studies, conducted in a number of different settings, of the other single-question screens that similarly ask about heavy drinking, lends strong support to their use.

Our study has several limitations. Almost half of the patients approached in the primary care waiting room refused to be screened for eligibility in the study, and approximately one fourth of eligible subjects did not complete the study. A lack of information about those who did not participate raises the possibility that those studied were not representative of primary care patients, potentially limiting the generalizability of our results. A higher than expected proportion of subjects reported substance use disorders, likely reflecting the fact that they were recruited from an urban safety-net hospital located in a community where the prevalence of such problems is high, but potentially also reflecting selection bias. The evaluation of a test in an atypical population can result in spectrum bias if, for instance, the unusual severity of the condition renders it more or less easily detectable. While the very close approximation of our results to those of this question and similar questions in other settings suggests that such bias, if present, is small, further study of the question's test characteristics in a more affluent, lower-risk population may be justified. A limitation of the NIAAA recommended question, and, as far as we know, of the other single-question screens, is that they have not yet been validated languages other than English. This represents another potential future area of study. Subjects were also assured anonymity, a condition which improves the accuracy of the reference standard interview but which may also serve to over-estimate the accuracy of the screening test itself. This is consistent, however, with the methodology of most other alcohol screening test studies.

The single-question screen accurately identified subjects with unhealthy alcohol use. Some patients who screen positive will have severe alcohol use disorders requiring referral to substance abuse treatment, while those who consume excessive amounts of alcohol but have not experienced severe health or interpersonal problems would benefit from brief intervention by the primary care provider. The lack of an efficient way to distinguish these two groups (the NIAAA Clinician's Guide recommends following up a positive screening test with 13 questions about drinking amounts and alcohol problems), is a challenge that must be addressed when implementing screening for unhealthy alcohol use. The AUDIT and ASSIST, in providing scores, provide a measure of severity. Even though they may be too long for universal screening in many settings, they might be done as brief assessments after a single-item screening question is answered in the affirmative. But this approach has not been tested or validated. Vinson et al. found that two follow-up questions (about drinking in hazardous situations and drinking more or for longer than intended) could identify alcohol use disorders among those with a positive response to a single-question screen[30]. This approach, if validated, might represent a more efficient solution than applying a longer test to all patients.

The single-question screen recommended by the NIAAA accurately identified unhealthy alcohol use in this sample of primary care patients. The sensitivity and specificity of this single question was comparable to that reported for longer instruments in other studies. These findings of validity support the use of this brief screen in primary care as recommended by NIAAA, which should, in turn, help with the implementation of universal screening for unhealthy alcohol use as recommended by national practice guidelines.

Acknowledgements: *Presented in abstract form at the annual meeting of the Society of General Internal Medicine, April 26, 2007 and at the annual meeting of the Research Society on Alcoholism, July 7, 2007. This research was supported by a grant from the National Institute on Alcohol Abuse and Alcoholism: NIAAA R01-AA010870.*

Conflicts of interest: *None disclosed.*

Corresponding Author: *Peter C. Smith, MD, MSc; Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, 2nd Floor, Crosstown Center, 715 Albany Street, Boston, MA 02118, USA (e-mail: peter.smith@bmc.org).*

REFERENCES

1. **Fiellin DA, Reid MC, O'Connor PG.** Screening for alcohol problems in primary care: a systematic review. Arch Intern Med. 2000;160: 1977–89.
2. **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.
3. United States Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. Ann Intern Med. 2004;140:554–6.
4. **Mayfield D, McLeod G, Hall P.** The CAGE questionnaire: validation of a new alcoholism screening instrument. Am J Psychiatry. 1974; 131:1121–3.
5. **Bohn MJ, Babor TF, Kranzler HR.** The Alcohol Use Disorders Identification Test (AUDIT): validation of a screening instrument for use in medical settings. J Stud Alcohol. 1995;56:423–32.
6. **Friedmann PD, McCullough D, Chin MH, Saitz R.** Screening and intervention for alcohol problems. A national survey of primary care physicians and psychiatrists. J Gen Intern Med. 2000;15:84–91.

7. **Spandorfer JM, Israel Y, Turner BJ.** Primary care physicians' views on screening and management of alcohol abuse: inconsistencies with national guidelines. J Fam Pract. 1999;48:899–902.
8. National Institute on Alcohol Abuse and Alcoholism. Helping patients who drink too much: a clinician's guide, 2005 Edition. Bethesda, MD, 2007.
9. **Canagasaby A, Vinson DC.** Screening for hazardous or harmful drinking using one or two quantity—frequency questions. Alcohol Alcohol. 2005;40:208–13.
10. **Williams R, Vinson DC.** Validation of a single screening question for problem drinking. J Fam Pract. 2001;50:307–12.
11. **Taj N, Devera-Sales A, Vinson DC.** Screening for problem drinking: does a single question work? J Fam pract. 1998;46:328–35.
12. **Seale JP, Boltri JM, Shellenberger S, et al.** Primary care validation of a single screening question for drinkers. J Stud Alcohol. 2006;67: 778–84.
13. **Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA.** The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol use disorders identification test. Arch Intern Med. 1998;158:1789–95.
14. **Sobell L, Sobell M.** Timeline Follow-back: a technique for assessing self-reported ethanol consumption. In: Allen J, Litten R, eds. Measuring Alcohol Consumption: Psychosocial and Biological Methods. Totowa, NJ: Humana Press; 1992:41–72.
15. **Dawson DA, Grant BF, Li TK.** Quantifying the risks associated with exceeding recommended drinking limits. Alcohol Clin Exp Res. 2005;29:902–8.
16. **Feinn R, Tennen H, Kranzler H.** Psychometric properties of the short index of problems as a measure of recent alcohol-related problems. Alcohol Clin Exp Res. 2003;27:1436–41.
17. **Haro JM, Arbabzadeh-Bouchez S, Brugha TS, et al.** Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health Surveys. Int J Methods Psych Res. 2006;154:167–80.
18. **Bradley KA, McDonell MB, Bush K, Kivlahan DR, Diehr P, Fihn SD.** The AUDIT alcohol consumption questions: reliability, validity, and responsiveness to change in older male primary care patients. Alcohol Clin Exp Res. 1998;22:1842–9.
19. **Altman DG, Gardner MJ.** Confidence intervals for research findings. Br J Obstet Gynaecol. 1992;99:90–1.
20. **Bradley KA, Bush KR, Epler AJ, et al.** Two brief alcohol-screening tests From the Alcohol Use Disorders Identification Test (AUDIT): validation in a female Veterans Affairs patient population. Arch Intern Med. 2003;163:821–9.
21. **Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR.** AUDIT-C as a brief screen for alcohol misuse in primary care. Alcohol Clin Exp Res. 2007;31:1208–17.
22. **Gordon AJ, Maisto SA, McNeil M, et al.** Three questions can detect hazardous drinkers. J Fam Pract. 2001;50:313–20.
23. **Rumpf H-J, Hapke U, Meyer C, John U.** Screening for alcohol use disorders and at-risk drinking in the general population: psychometric performance of three questionnaires. Alcohol Alcohol. 2002;37:261–8.
24. **Yarnall KSH, Pollak KI, Ostbye T, Krause KM, Michener JL.** Primary care: is there enough time for prevention? Am J Pub Health. 2003;93:635–41.
25. **Selzer ML.** The Michigan alcoholism screening test: the quest for a new diagnostic instrument. Am J Psychiatry. 1971;127:1653–8.
26. **Saunders JB, Aasland OG, Babor TF, 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.
27. **Buchsbaum DG, Buchanan RG, Lawton MJ, Elswick RK Jr., Schnoll SH.** A program of screening and prompting improves short-term physician counseling of dependent and nondependent harmful drinkers. Arch Intern Med. 1993;153:1573–7.
28. WHO ASSIST Working Group. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. Addiction. 2002;97:1183–94.
29. **Buchsbaum DG, Buchanan RG, Centor RM, Schnoll SH, Lawton MJ.** Screening for alcohol abuse using CAGE scores and likelihood ratios. Ann Intern Med. 1991;115:774–7.
30. **Vinson DC, Kruse RL, Seale JP.** Simplifying Alcohol Assessment: Two Questions to Identify Alcohol Use Disorders. Alcohol Clin Exp Res. 2007;31:1392–8.

J Community Health (2009) 34:321–327
DOI 10.1007/s10900-009-9153-9

ORIGINAL PAPER

Asthma 1-2-3: A Low Literacy Multimedia Tool to Educate African American Adults About Asthma

Rina M. Sobel · Michael K. Paasche-Orlow · Katherine R. Waite · Sarah S. Rittner · Elizabeth A. H. Wilson · Michael S. Wolf

Published online: 8 April 2009

Abstract Asthma 1-2-3 is a newly-developed low-literacy multimedia education tool designed to promote asthma self-care concepts among African American adults. An expert panel ($n = 10$) informed content development for the tool. The video script and storyboard imagery were shown to 30 African Americans recruited from the American Lung Association, whose reactions and comments guided further revisions. The final version was pilot tested in three diverse community settings in Chicago to determine the efficacy of Asthma 1-2-3 at improving patient understanding of asthma and its symptoms. In all, 130 adults participated in the pilot test. Knowledge scores significantly improved from pretest to posttest following presentation of the developed tool for subjects across all literacy levels (Pretest: Mean = 4.2 [SD = 1.6]; Posttest: M = 6.8 [SD = 2.0], $P < 0.001$). Symptom pathophysiology concepts were the least understood. Individuals with low literacy had less total knowledge score gains compared to those with marginal and adequate literacy (1.8, 2.6, and 3.2 respectively; $P = 0.002$). The multimedia tool significantly improved understanding of asthma. Individuals with limited literacy may require additional instruction, repeated viewing, or added tangible cues (i.e. supplementary print materials) to support knowledge retention. In general, feedback from the target population was particularly helpful in the development of the tool and its initial evaluation, and should be considered as a necessary step in the creation of other patient education materials.

Keywords Asthma · Education · Knowledge · Multimedia · Health literacy

R. M. Sobel (✉) · K. R. Waite · E. A. H. Wilson · M. S. Wolf
Health Literacy and Learning Program, Center for Communication in Healthcare, Institute for Healthcare Studies, Division of General Internal Medicine, Northwestern University Feinberg School of Medicine, 750 N. Lake Shore Drive, 10th Floor, Chicago, IL 60611, USA
e-mail: r-sobel@northwestern.edu

S. S. Rittner
Center for Healthcare Equity, Institute for Healthcare Studies, Northwestern University Feinberg School of Medicine, 750 N. Lake Shore Drive, 10th Floor, Chicago, IL 60611, USA

M. K. Paasche-Orlow
Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, 801 Massachusetts Ave, Crosstown Center, 2nd Floor, Boston, MA 02118, USA

Introduction

African American men and women in the United States are disproportionately likely to be living with asthma when compared to other groups [1, 2]. Several studies have documented this racial disparity, such as Centers for Disease Control and Prevention (CDC) data that has shown that between 1998 and 2006 African Americans suffered from higher rates of asthma than nearly any other population [3]. Suggested causes for this disparity include poor living conditions, presence of known triggers such as dust and environmental pollutants, and inadequate disease self-management [1].

Recent studies have found limited 'health literacy'—one's capacity to obtain, process, and understand health information in order to make appropriate decisions—to be associated with poorer patient understanding of and inadequate self-care for chronic diseases, including asthma [4–7]. In these and other health literacy studies, African American race has been linked to less education and inadequate literacy skills [8]. Recent evidence further

suggests that certain noted racial disparities in health outcomes are mediated by patients' literacy abilities [9–11].

Many presently available asthma education programs target children and young adults through the use of technology such as multimedia education tools [12–14]. While this style of intervention has generally been proven effective for this age group, such studies focusing on adults, particularly African Americans, are lacking. We sought to develop and pilot test a novel education tool called Asthma 1-2-3, that incorporates best practices derived from adult literacy and cognitive factors research and is culturally tailored to the African American community [15, 16]. The goal of the tool is to simplify and standardize asthma education with the aim of assuring that patients have a functional understanding of the disease and how to prevent and manage symptoms. Partnered with a clinic-based dissemination strategy, the long-term goal for tools such as Asthma 1-2-3 is to support sustained improvement in health promoting behaviors and aid larger efforts targeting noted disparities between the African American population and other non-minority populations. In this paper, we (1) review the patient-centered development process of Asthma 1-2-3, and (2) report preliminary evidence of its efficacy based on pilot testing of the tool among community-dwelling African American adults with and without asthma.

Methods

Conceptual Framework

Asthma 1-2-3 was developed based on the belief that proper communication of asthma-related concepts is a necessary precondition for maintaining appropriate health behaviors [17]. Even with information acquisition as a precursor, it is understood that knowledge alone does not always directly link to recommended actions. Asthma 1-2-3 utilizes a 'health literacy perspective', which includes adult learning and education principles as well as cognitive factors 'best practices' to increase comprehension among patients, regardless of literacy skills [18–23]. Doak et al. [16] provided early guidance through their Suitability Assessment of Materials (SAM) criteria in an effort to offer an objective means for effectively presenting medical information to patients with limited literacy. Criteria are specific to content, literacy demand, graphics, layout and typography, learning stimulation/motivation, and the cultural appropriateness of materials. Furthermore, our multimedia tool utilizes aspects from cognitive factors research, which recommends reducing the cognitive load placed on working memory by minimizing the number of new concepts introduced simultaneously, and by avoiding distracting, 'off-message' information [18, 23–29].

Development of Asthma 1-2-3

Asthma 1-2-3 is a multimedia curriculum designed to parsimoniously deliver actionable patient information promoting recommended asthma self-care behaviors. It is designed to be versatile, for use either by clinic staff in a primary care setting or by workers and nurses in the healthcare community at-large. The video is not interactive, but rather it introduces a limited number of asthma concepts at a time to be followed by the clinic staff or a community worker leading a 'teach-to-goal' strategy to reinforce learning. As a research tool there are many possibilities for its use: the internet, clinic kiosks, a link within an electronic medical record for physician administration, or even as a DVD available for patients to take home.

Working with an asthma expert panel (N = 10; general internist, family practitioner, allergist, pharmacist, health literacy expert, two asthma nurses, and three African American adults living with asthma), content was generated and categorized into three chapters, ranging from 1½ to 5 minute in length each. Agreed upon chapter titles were: (1) What is Asthma, (2a) Manage Your Asthma: Triggers, (2b) Manage Your Asthma: Medication, and (3) Monitor Your Asthma. Together, these chapters were viewed as a core curriculum for patients, although additional topics (e.g. Talking with Your Doctor, Your Action Plan) will be developed in the near future.

Cognitive Interviews

A preliminary video script with sample storyboard imagery was generated by the research team and reviewed by the expert panel. The research team revised each chapter's script based on panel feedback, and early video prototypes of the chapters were then developed. African American adult caregivers of children with asthma (N = 13) were recruited through the American Lung Association to provide comment on the language used in the video and chapter storyboards. Segments of the prototype video were shown to each participant, followed by a series of probes to ascertain their functional understanding of the conveyed messages and their perspective on the language, narration, and visual imagery used. Comments were audio-recorded and reviewed by the research team to support specific revisions.

All proposed revisions to the video were presented for approval to the expert panel. Once a version was approved by the research team and panel members, a second wave of cognitive interviews was conducted. Seventeen African Americans recruited from one adult basic education

program course were interviewed individually for approximately 20 min after viewing the latest version of Asthma 1-2-3. Once the individual cognitive interviews were complete, a 45 min targeted discussion group was held. The group discussion addressed participants' general feedback of the tool, its effectiveness, and recommended changes. Comments were audio-taped for later transcription. Following the research teams' consideration of all participants' suggestions, the program went through a final revision.

Pilot Test

The final version of the introductory chapter of the Asthma 1-2-3 curriculum was pilot tested among 130 African American adults, with and without asthma and with varying literacy skills, in three diverse settings in the Chicago area to ascertain this chapter of the tool's efficacy for basic knowledge dissemination. A single-group, pretest–posttest design was implemented for this initial evaluation.

Sample

Subjects were African American adults recruited from three unique study sites in the Chicago area; a faith-based organization, an adult basic education center, and a general internal medicine ambulatory care clinic. Recruitment took place from August 2007 to January 2008. Participants were ineligible if they met any of the following exclusion criteria: (1) blindness or severely impaired vision not correctable with eyeglasses; (2) deafness or hearing problems uncorrectable with a hearing aid; (3) too ill to participate; (4) non-English speaking. Approval for human subjects research was obtained from the Institutional Review Board at Northwestern University prior to consenting participants to the study.

Measurement

A structured interview protocol was developed to assess functional understanding of basic asthma concepts presented in the first chapter of Asthma 1-2-3. Participants were questioned regarding their understanding of asthma as a disease, body parts affected, identification of asthma symptoms, recognition of the link between symptoms and disease control, comprehension of the pathophysiology of asthma symptoms, and perception of the seriousness of the disease. Identical questionnaires were administered both before and after the viewing of Asthma 1-2-3 for use in the pretest/posttest comparison. A total knowledge score was generated based on these assessments, with a range of 0 to 12 for possible scores (each distinct concept worth one point). Classifying asthma as a disease, identifying the body parts affected, and recognizing the link between asthma symptoms and disease control were worth one point each. Comprehension of the pathophysiology of asthma symptoms provided up to three points; one each for swelling, mucus, and airway tightness. Identification of asthma symptoms was worth up to six points, one point per symptom recognized in the video (Table 2).

Participant literacy was assessed using the Rapid Estimate of Adult Literacy in Medicine (REALM), a health word recognition test used most frequently to assess adult health literacy in medical settings [30]. Participants were asked to read aloud 66 medical terms. Scores were determined based on the total number of words pronounced correctly, with dictionary pronunciation being the scoring standard. Raw scores were then converted into one of four reading grade levels: 3rd grade or less (0–18), 4th–6th grade (19–44), 7th–8th grade (45–60), and 9th grade and above (61–66). In health care studies, such as this one, where participants need only to be categorized as low (scores 0–44), marginal (scores 45–60) or adequate (scores 61–66) readers, the information provided by the REALM is generally sufficient. The REALM is highly correlated with standardized reading tests and the Test of Functional Health Literacy in Adults (TOFHLA) [31].

Procedure

A total of 239 African American adults were approached between August 2007 and January 2008. In all, 130 participants were eligible and consented to the study among the three study sites ($n = 27$ [faith based organization]; $n = 26$ [adult basic education center]; $n = 78$ [general internal medicine clinic]). Once consented, demographic data was collected and a pretest measure of basic asthma concepts and a health literacy assessment (REALM) were administered. Participants then viewed the first chapter of Asthma 1-2-3 using a PC laptop computer equipped with Windows Media Player and an individual headset. After viewing the program, an immediate posttest measure of understanding and recall, with questions identical to those presented in the pretest assessment, was conducted and participants received $10 for their participation.

Analysis Plan

The mean and standard deviation were calculated for participant age, and descriptive statistics were determined for all other categorical data. A McNemar's test for categorical or dichotomous (i.e. correct/incorrect) data, or a paired *t*-test for continuous data was conducted to examine differences in knowledge before and after viewing chapter 1 of Asthma 1-2-3 (as assessed through pre and posttest scores) and in participant characteristics (age, gender,

education, literacy level, asthma diagnosis). Multivariate linear regression was used to examine the association between literacy and posttest knowledge score, while adjusting for pretest score, age, gender, education, and asthma diagnosis (self and/or relative). All statistical analyses were performed using STATA, version 9.0 (College Station, TX, USA).

Results

Sample Characteristics

The mean age of participants was 50.2 years and 76.2% were female (Table 1). Nearly one-fourth (22.3%) of participants had been diagnosed with asthma, and 63.8% had family members who suffered from the disease. More than 50% of participants reported having at least some college education. More than half of participants had limited literacy skills; 26.2% read at or below a 6th grade level (low literacy) and one-third (33.0%) read at a 7th to 8th grade level (marginal literacy).

Asthma Knowledge

Baseline functional understanding was assessed among participants, and overall comprehension of asthma and its symptoms was relatively low prior to viewing the video tool (Table 2). Approximately one quarter (26%) of participants did not perceive asthma to be a disease, and one-third (38%) could not accurately identify the body parts affected. Asthma pathophysiology of what happens to lungs and airways when symptoms occur was the least understood set of concepts. Despite this lack of knowledge among subjects, baseline perceived seriousness of asthma was relatively high [M = 4.5 (SD = 0.6) on a 5-point scale]. Participants with low literacy skills were less able to identify body parts affected compared to those with marginal and adequate literacy skills (41.2%, 58.1%, 77.4% respectively; $P = 0.003$). In addition, individuals with low literacy were less able to identify asthma symptoms compared to those who had adequate literacy (M = 1.2 (SD = 1.0) symptoms versus M = 2.0 (SD = 1.0) symptoms; $P = 0.001$). Participants most often recognized difficulty breathing (71%) and wheezing (43%) as asthma symptoms. No significant differences in baseline understanding were noted by presence or absence of asthma in either participants or their relatives.

Table 1 Demographic information ($N = 130$)

Variable	Summary value
Age, mean (SD)	50.2 (15.3)
Gender (%)	
Female	76.2
Education (%)	
<High school	22.5
High school graduate	22.3
>High school	53.9
Asthma diagnosis (%)	
Yes	22.3
Family member has asthma (%)	
Yes	63.8
Literacy level (%)	
Low (≤6th grade level)	26.2
Marginal (7th–8th grade level)	33.0
Adequate (≥9th grade level)	40.1

Table 2 Basic knowledge pre and post-asthma 1-2-3 intervention ($N = 130$)

Variable	Pre-test (correct)	Post-test (correct)	*P*-value
Classify asthma as a disease (%)	74	95	.008
Identify parts of body affected (%)	62	84	<.001
Name asthma symptoms (%)			
Itchy/scratchy throat	1	32	<.001
Tired/worn out	2	28	<.001
Chest feels tight	10	43	<.001
Cough	24	65	<.001
Whistling/wheezing sound	43	49	.229
Hard to breathe	72	86	.001
Recognize link between symptoms and disease control (%)	80	96	.059
Pathophysiology of asthma symptoms (%)			
Swelling	5	25	<.001
Mucus	8	26	<.001
Airway constriction	35	48	.020
Asthma seriousness, mean (SD)	4.5 (0.6)	4.7 (0.5)	.001
Total score, mean (SD)	4.2 (1.6)	6.8 (2.0)	<.001

Intervention

Participants' understanding of asthma significantly improved after watching Asthma 1-2-3. The total knowledge score (range 0–12) improved more than 60%, from M = 4.2 (SD = 1.6) at pretest to M = 6.8 (SD = 2.0) at posttest ($P < 0.001$). Although improvement in understanding of one concept—that asthma symptoms are related to disease control—did not reach significance, it did show a strong trend ($P = 0.06$, Table 2). Furthermore, while overall understanding from pretest to posttest significantly improved across all literacy levels, individuals with low literacy had smaller total knowledge gains

Table 3 Multivariate regression model examining asthma knowledge by literacy level

Measure	Literacy level		
	Adequate ($n = 52$)	Marginal ($n = 46$)	Low ($n = 31$)
Posttest asthma knowledge score, mean (SD)	7.8 (1.7)	6.6 (1.9)	5.6 (1.8)
Crude difference (95% CI)	–	−1.1 (−1.9, −0.4)***	−2.1 (−2.9, −1.4)***
Adjusted difference (95% CI)[a,b]	–	−0.8 (−1.5, −0.1)*	−1.5 (−2.3, −0.6)***

[a] Differences adjusted for subject age, gender, education, asthma diagnosis (self and/or relative), and baseline knowledge score

[b] Model fit statistics: $F_{(8,118)} = 8.8$, $P < 0.001$, adjusted $r^2 = 0.33$

* $P < 0.05$

*** $P < 0.001$

compared to those with marginal and adequate literacy (1.8, 2.6, and 3.2 respectively; $P = 0.002$).

Multiple linear regression analysis was used to examine crude and adjusted differences in posttest knowledge scores by literacy level while controlling for relevant covariates (Table 3). Both low and marginal literacy remained significant independent predictors of asthma knowledge post-exposure to Asthma 1-2-3. Differences among mean scores approximated nearly 1 point lower (β −0.8, 95% CI −1.5 to −0.1) for those with marginal literacy and 1.5 points lower (β −1.5, 95% CI −2.3 to −0.6) for those with low literacy compared to subjects with adequate literacy skills. No significant associations were noted by age, gender, education, or whether subjects' had an existing asthma diagnosis.

Discussion

Asthma 1-2-3 was developed utilizing cognitive principles and health literacy 'best practices' to create a video tool promoting asthma understanding and self-management. This version targets African American adults living with asthma, and community members were included in its development, refinement, and pilot testing. This input contributed to the specific tailoring of the tool, increasing its relevance to this population. The preliminary pilot test indicates that the tool significantly improved subjects' comprehension of the disease, its etiology, and the importance of self-care vigilance. Relative to baseline knowledge, participants' understanding of almost all concepts improved after watching the video. Understanding the link between asthma symptoms and disease control was not statistically significant; however, this lack of significance could be explained by the high baseline score for this concept, which suggests an already high level of understanding before viewing the multimedia tool.

Notably, scores improved only modestly for recall of asthma symptoms and their etiology. In the tool, the content relating to these two concepts is recited in list form by the narrator, yet no textual cues are presented. This presentation strategy is supported by cognitive principles of multimedia learning, as detailed by Mayer, which state that information processing may be more difficult when presented with simultaneous yet redundant auditory and visual content [32]. One of the multimedia guidelines set forth by Mayer allows for the use of text-based cues in close proximity to content-related images when the audience may have low prior knowledge. In these two concepts, subjects had only a minimal baseline understanding.

While participants across all literacy levels demonstrated significant improvement in their overall knowledge of asthma after watching Asthma 1-2-3, participants with adequate literacy skills learned the most while those with low literacy skills gained the least (Table 3). With multimedia, subjects are not able to review content after it is shown, nor can they control the pace of its presentation. For declarative content such as asthma symptoms and disease etiology, participants might have actually performed better and achieved deeper encoding of the material had they been able to review sections or use virtual 'page-turns' before moving on to new concepts. Video-based education does have many benefits, such as facilitating the demonstration of desired behaviors and explicitly showing pathophysiology. Yet patients, particularly those with lower literacy abilities, may benefit from the addition of a more tangible delivery modality, such as a summary card to accompany the video to support learning and retention. Further studies should test a combined video and supplemental enhanced print approach to best address the broad range of literacy needs.

There are several limitations to this study. Both the relatively small sample size and recruitment via targeted community exchanges (adult basic education center, faith-based organization) and one general internal medicine clinic limit the generalizability of our findings. Furthermore, our pilot test outcome utilized immediate recall, but neither retention of acquired asthma knowledge over an extended period nor behavior change were tested. A future study should assess longer-term effectiveness of Asthma

1-2-3. Finally, the results presented offer preliminary support and direction for refining the multimedia tool in order to optimize efficacy. The next stage of research should more properly evaluate the tool's efficacy in a controlled trial design.

Many studies have examined the advantages of using video as a medium for chronic condition education [13, 33–36]. To our knowledge, Asthma 1-2-3 is one of the first tools developed according to health literacy and cognitive factors 'best practices' to target asthma education within an African American adult population. Health materials that are specifically tailored to the interests and needs of target populations can help initiate patient-provider discussion and increase knowledge and awareness. The digital medium itself allows it to be embedded in exam room computers linked to an Electronic Medical Record, giving providers an opportunity to initiate use of the tool and document if it has been viewed by a patient. As each chapter is brief, patients can be educated prior to the clinician encounter, priming them for more meaningful discussions. If a medical practice is not yet equipped with this technology, the tool can be made easily available to patients in waiting rooms through relatively inexpensive computer-equipped kiosks or portable DVD players.

Asthma 1-2-3 has the potential to teach patients about asthma in a relevant, comprehensible, and parsimonious way, especially if paired with an appropriate provider interaction and perhaps tangible and complimentary print materials for patients with limited literacy. Such an approach can help more easily identify knowledge deficits and/or correct misinformation before and/or during an appointment. The goal for any patient education program like Asthma 1-2-3 should not be a stand-alone intervention. Rather, these tools are best viewed as part of a more comprehensive strategy to confirm patient comprehension of relevant health content, encourage dialogue between patients and their healthcare providers, and achieve uptake of recommended behaviors that will lead to optimal health outcomes across all literacy levels.

Acknowledgements This research was supported by a Centers for Disease Control and Prevention Career Development Award (Wolf: K01 EH000067-01). Authors acknowledge Quinn M. Stephens for his technical assistance.

References

1. Boudreaux, E. D., Emond, S. D., Clark, S., & Camargo, C. A., Jr. (2003). Acute asthma among adults presenting to the emergency department: The role of race/ethnicity and socioeconomic status. *Chest, 124*(3), 803–812.
2. Grant, E. N., Lyttle, S. C., & Weiss, K. B. (2000). The relationship to socioeconomic factors and racial/ethnic differences in US asthma mortality. *American Journal of Public Health, 90*(12), 1923–1925.
3. U.S. Department of Health and Human Services. (2006). *Health data for all ages (HDAA): Asthma and chronic obstructive pulmonary disease, all ages: US, 1998–2006* (Source: NHIS). Retrieved March 24, 2008, from http://205.207.175.93/HDAA/TableViewer/tableView.aspx?ReportId=261
4. Gazmararian, J. A., Williams, M. W., Baker, D. W., & Peel, J. (2003). Health literacy and patient knowledge of chronic disease. *Patient Education and Counseling, 51*(3), 267–275.
5. Williams, M. V., Baker, D. W., Honig, E. G., Lee, T. M., & Nowlan, A. (1998). Inadequate literacy is a barrier to asthma knowledge and self-care. *Chest, 114*(4), 1008–1015.
6. Wolf, M. S., Davis, T. C., Arozullah, A., Penn, R., Arnold, C., Sugar, M., et al. (2005). Relation between literacy and HIV treatment knowledge among patients on HAART regimens. *AIDS Care, 17*(7), 863–873.
7. Davis, T. C., Williams, M. V., Marin, E., Parker, R. M., & Glass, J. (2002). Health literacy and cancer communication. *CA: A Cancer Journal for Clinicians, 52*(3), 134–149.
8. U.S. Department of Education, National Center for Education Statistics. (2006). *The health literacy of America's adults: Results from the 2003 National assessment of adult literacy*. Washington D.C.: Kutner M, Greenberg, E., JinY., & Paulsen, C. NCES Publication No. 2006–483.
9. Wolf, M. S., Knight, S. J., Lyons, A., Durazo-Arzivu, R., Pickard, S. A., Arseve, A., et al. (2006). Literacy, race, and PSA level among low income men newly diagnosed with prostate cancer. *Urology, 68*(1), 89–93.
10. Sentell, T. L., & Halpin, H. A. (2006). Importance of adult literacy in understanding health disparities. *Journal of General Internal Medicine, 21*(8), 862–866.
11. 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.
12. Krishna, S., Francisco, B. D., Balas, E. A., Konig, P., Graff, G. R., & Madsen, R. W. (2003). Randomized trial. Internet-enabled interactive multimedia asthma education program: A randomized trial. *Pediatrics, 111*(3), 503–510.
13. McPherson, A., Forster, D., Glazebrook, C., & Smyth, A. (2002). The asthma files: Evaluation of a multimedia package for children's asthma education. *Paediatric Nursing, 14*(2), 32–35.
14. Bartholomew, L. K., Gold, R. S., Parcel, G. S., Czyzewski, D. I., Sockrider, M. M., Fernandez, M., et al. (2000). Watch, discover, think, and act: Evaluation of computer-assisted instruction to improve asthma self-management in inner city children. *Patient Education and Counseling, 39*(2–3), 269–280.
15. Seligman, H. K., Wallace, A. S., DeWalt, D. A., Schillinger, D., Arnold, C. L., Shilliday, B. B., et al. (2007). Facilitating behavior change with low literacy patient education materials. *American Journal of Health Behavior, 31*(Supp 1), S69–S78.
16. Doak, C. C., Doak, L. G., & Root, J. H. (1996). *Teaching patients with low literacy skills* (2nd ed.). Philadelphia: Lippincott.
17. National Cancer Institute. (1995). *Clear & simple: Developing effective print materials for low literate readers*. Washington, DC: HHS. NIH Publication No. 95-3594.
18. Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*(3), 339–354.
19. van den Broek, P., & Kremer, K. (2000). The mind in action: What it means to comprehend. In B. Taylor, P. van den Broek, & M. Graves (Eds.), *Reading for meaning* (pp. 1–31). New York: Teacher's College Press.
20. Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension: An integrative view of reading. *Current Directions in Psychological Science, 14*, 276–279.

21. Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363–394.
22. Morrow, D., & Leirer, V. O. (1999). Designing medication instructions for older adults. In D. C. Park, R. W. Morrel, & K. Shifren (Eds.), *Processing of medical information in aging patients. Cognitive and human factors perspectives* (pp. 249–266). Mahwah, MJ: Lawrence Erlbaum Associates Publishers.
23. Paasche-Orlow, M. K., Cheng, D. M., Palepu, A., Meli, S., Faber, V., & Samet, J. H. (2006). Health literacy, antiretroviral adherence, and HIV-RNA suppression: A longitudinal perspective. *Journal of General Internal Medicine, 21*(8), 835–840.
24. Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.
25. Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). New York: Cambridge University Press.
26. Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors, 46*(3), 567–581.
27. Schoo, A. M. M., Morris, M. E., & Bui, Q. M. (2005). The effects of mode of exercise instruction on compliance with a home exercise program in older adults with osteoarthritis. *Physiotherapy, 91*(2), 79–86.
28. Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9*(2), 75–82.
29. Tabbers, H. K., Martens, R. L., & van Merrienboer, J. J. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*(1), 71–81.
30. 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. *Family Medicine, 25*, 391–395.
31. Parker, R. M., Baker, D. W., Williams, M. V., & Nurss, J. R. (1995). The test of functional health literacy in adults: A new instrument for measuring patients' literacy skills. *Journal of General Internal Medicine, 10*(10), 537–541.
32. Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.
33. Pignone, M., Harris, R., & Kinsinger, L. (2000). Videotape-based decision aid for colon cancer screening. A randomized, controlled trial. *Annals of Internal Medicine, 133*, 761–769.
34. Morgan, M. W., Deber, R. B., Llewellyn-Thomas, H. A., Gladstone, P., Cusimano, R. J., O'Rourke, K., et al. (2000). Randomized controlled trial of an interactive videodisc decision aid for patients with ischemic heart disease. *Journal of General Internal Medicine, 15*(10), 685–693.
35. Bader, J. L., & Strickman-Stein, N. (2003). Evaluation of new multimedia formats for cancer communications. *Journal of Medical Internet Research, 5*(3), 16–19.
36. Meade, C. D., McKinney, P. W., & Barnas, G. P. (1994). Educating patients with limited literacy skills: The effectiveness of print and videotaped materials about colon cancer. *American Journal of Public Health, 84*(1), 119–121.

Binge Drinking Among U.S. Active-Duty Military Personnel

Mandy A. Stahre, MPH, Robert D. Brewer, MD, MSPH, Vincent P. Fonseca, MD, MPH, Timothy S. Naimi, MD, MPH

Background: Binge drinking (drinking on a single occasion ≥5 drinks for men or ≥4 drinks for women) is a common risk behavior among U.S. adults that is associated with many adverse health and social consequences. However, little is known about binge drinking among active-duty military personnel (ADMP). The objectives of this study were to quantify episodes of binge drinking, to characterize ADMP who binge-drink, and to examine the relationship between binge drinking and related harms.

Methods: The prevalence of binge drinking and related harms was assessed from responses to the 2005 Department of Defense Survey of Health Related Behaviors Among Military Personnel (n=16,037), an anonymous, self-administered survey. The data were analyzed in 2007 after the release of the public-use data.

Results: In 2005, a total of 43.2% of ADMP reported past-month binge drinking, resulting in 29.7 episodes per person per year. In all, 67.1% of binge episodes were reported by personnel aged 17–25 years (46.7% of ADMP), and 25.1% of these episodes were reported by underage youth (aged 17–20 years). Heavy drinkers (19.8% of ADMP) were responsible for 71.5% of the binge-drinking episodes and had the highest number of annual per-capita episodes of binge drinking (112.6 episodes). Compared to nonbinge drinkers, binge drinkers were more likely to report alcohol-related harms, including job performance problems (AOR=6.5; 95% CI=4.65, 9.15); alcohol-impaired driving (AOR=4.9; 95% CI=3.68, 6.49); and criminal justice problems (AOR=6.2; 95% CI=4.00, 9.72).

Conclusions: Binge drinking is common among ADMP and is strongly associated with adverse health and social consequences. Effective interventions (e.g., the enforcement and retainment of the minimum legal drinking age) to prevent binge drinking should be implemented across the military and in conjunction with military communities to discourage binge drinking.
(Am J Prev Med 2009;36(3):208–217) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Excessive alcohol consumption resulted in an average of approximately 79,000 deaths and 2.3 million years of potential life lost (about 30 years of life lost per death) in the U.S. from 2001 to 2005 (https://apps.nccd.cdc.gov/ardi/Homepage.aspx), making it the third leading preventable cause of death.[1] Binge drinking, usually defined as the consumption on a single occasion of ≥5 drinks for men or ≥4 drinks for women,[2,3] typically leads to acute impairment; it accounted for more than half of these 79,000 deaths (https://apps.nccd.cdc.gov/ardi/Homepage.aspx). Binge drinking is a common risk factor among U.S. adults, with approximately 1.5 billion episodes of binge drinking reported by U.S. adults in 2001 alone.[4] Additionally, binge drinking is associated with many adverse health and social consequences, including interpersonal violence, motor vehicle crashes, sexually transmitted diseases, unintended pregnancy, fetal alcohol syndrome, lost productivity, and suicidal behavior.[5–13]

Various studies have reported that excessive drinking and related harms are common among military personnel[14–16] and that a higher percentage of active-duty military personnel (ADMP) misuse alcohol compared to civilian populations, even after controlling for age and gender.[16–19] Studies have also shown that high levels of alcohol use, such as binge drinking, are associated with a high percentage of noncombat-related hospitalizations and deaths—usually the result

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 (Stahre, Brewer, Naimi), Atlanta, Georgia; and the Texas Department of State Health Services (Fonseca), Austin, Texas

At the time the research was completed, Fonseca was at the Population Health Support Division, Air Force Medical Support Agency, U.S. Air Force, Brooks City-Base, San Antonio, Texas.

Address correspondence and reprint requests to: Mandy A. Stahre, MPH, CDC/Alcohol Team, c/o University of Minnesota, School of Public Health, Division of Epidemiology, 1300 S. Second Street, Suite 300, Minneapolis MN 55454-1015. E-mail: mstahre@cdc.gov.

0749-3797/09/$–see front matter
doi:10.1016/j.amepre.2008.10.017

of unintentional injuries—among military personnel.[19–24] Further, binge drinking by ADMP can adversely affect military readiness, workplace productivity, safety, and healthcare expenditures, particularly given the equipment and the dangerous environments commonly encountered by ADMP.[9,25,26]

Despite these known risks, no study has assessed the frequency and the per-capita rate of binge-drinking episodes among ADMP in the U.S., nor has a study examined the relationship between binge drinking and other health and social consequences (e.g., problems with job performance). This research is needed because people who binge-drink often do so frequently, increasing the likelihood of alcohol-attributable harms. Data were used from a worldwide, cross-sectional survey of ADMP to assess the frequency and rate of binge-drinking episodes and the potential relationship between binge drinking and various health and social consequences in this population.

Methods

The Department of Defense Survey of Health Related Behaviors Among Military Personnel is an anonymous, self-administered survey of health outcomes and health risk behaviors among current ADMP stationed in the U.S., overseas, and onboard ships. The survey is conducted by the Research Triangle Institute (RTI) staff every 2–3 years and is the only population-based health survey involving all branches of the military. Military personnel are randomly selected, using a multilevel sampling frame to be representative of the entire active-duty military population. Survey participants from a particular base may be asked to complete the survey as a group at a specific base location, or may complete the survey on their own, and then return it to RTI, depending on the nature of their assignment (e.g., overseas). Those who are ineligible to participate include recruits; service academy cadets; personnel who are transferring to another base during data collection; those who have left the military; those who are absent without leave (AWOL); or those who have an unknown status. Data collection for the present survey occurred from April 2005 to August 2005, the most recent year for which survey data were available. Because of nonresponse, poststratification adjustments were made by branch of service, age, and race/ethnicity to maintain the representativeness of the sample. Details of the nested sampling, purpose, method, and analysis have been published elsewhere.[27–29]

Based on questions related to alcohol use in the past 30 days, respondents were characterized as either abstainers (no alcohol use in the past 30 days) or current drinkers. Average daily alcohol consumption was calculated, using responses to questions about the frequency of alcohol consumption and the usual quantity consumed in the last 30 days for beer, wine, and liquor. Heavy drinkers were defined as those consuming an average of >2 drinks per day for men or an average of >1 drink per day for women. Nonheavy drinking was an average daily alcohol consumption level less than that used to define heavy drinking.

Binge drinking was assessed using the following question: *During the past 30 days, on how many days did you have 5 or more drinks of beer, wine, or liquor on the same occasion (4 or more if you are a woman)?* Binge drinkers were defined as those who reported at least one ≥5/≥4 drinking episode in the last 30 days. Responses to this question were reported categorically as once, 2–3 days, 4–10 days, 11–19 days, 20–27 days, and 28–30 days. The midpoint of each response range was used to calculate total binge-drinking episodes, a technique commonly used when calculating alcohol consumption with frequency-range data.[30] Binge-drinking patterns were assumed to be fairly stable in this population, so the number of monthly episodes was multiplied by 12 to estimate total annual episodes. The use of this constant also helped to smooth out the slight month-to-month variation in alcohol consumption that occurred during the 5-month sampling period (April–August). Per-capita episodes of binge drinking (episodes per person per year) were calculated by dividing the total number of binge episodes per year for a given stratum by the total weighted number of people in that stratum.

To further characterize the public health impact of binge drinking, the prevalence of various health and social consequences (e.g., productivity problems, alcohol-impaired driving, and criminal-justice problems) was assessed among respondents during the 12 months preceding the survey. For example, respondents were asked *How many times in the past 12 months did you not get promoted because of your drinking?* The prevalence of these consequences was compared between binge drinkers and current drinkers who did not report binge drinking in the past 30 days.

The overall response rate for the 2005 survey was 51.8%, a rate based on the number of completed, usable interviews of personnel who were eligible to participate in the survey. Nonrespondents were people who either did not show up for the survey session during their scheduled time or did not return a completed survey. Fewer than 1% of respondents were removed from the final study population because of missing information on binge drinking, resulting in a final sample size of 16,037.

To examine the relationship between drinking pattern (binge versus nonbinge) and health and social consequences, summary Mantel–Haenszel ORs were computed, adjusting for potential confounders (i.e., age group and gender). All analyses occurred from May 2007 to August 2007 after the release of the public-use data files; they were run using SAS version 9.1 and SAS-callable SUDAAN version 9.0.1 software to take into account the complex weighting structure of the survey.

Results

Overall Active-Duty Military

In 2005, a total of 43.2% of ADMP reported at least one episode of binge drinking in the past 30 days. Of these respondents, 12.1% reported at least one episode of binge drinking in the previous month, and 31.1% reported two or more. Extrapolation to 1 year produced an estimate of 30 million episodes of binge drinking, or 29.7 episodes of binge drinking per active-duty person per year (Table 1). The highest numbers of

Table 1. Prevalence, total number, and per-capita episodes of binge drinking[a] among all U.S. active-duty personnel by selected characteristics, 2005

	Number of respondents	Weighted proportion of military population (n=1,004,879)	Binge-drinking prevalence, past 30 days (%)	Total (estimated) weighted number of binge episodes per year	Estimated per-capita binge-drinking episodes (episodes/person/year)
All respondents[b]	16,037	100.0	43.2	29,844,000	29.7
Service					
Army	3,629	31.9	51.8	12,196,000	38.0
Navy	4,595	26.8	40.1	7,791,000	28.9
Marine Corps	3,350	12.8	51.4	4,904,000	38.3
Air Force	4,463	28.5	32.9	4,954,000	17.3
Gender					
Male	12,048	85.3	46.6	28,021,000	32.7
Female	3,989	14.8	24.1	1,823,000	12.3
Age (years)					
17–20	1,290	14.1	44.4	5,038,000	35.5
21–25	4,277	32.6	59.5	14,990,000	45.7
26–34	4,283	30.2	38.9	6,521,000	21.5
≥35	6,187	23.1	25.2	3,296,000	14.2
Race/ethnicity					
White	9,800	64.4	46.5	20,599,000	31.8
Black	2,604	17.6	32.0	3,742,000	21.2
Hispanic	1,993	8.9	46.7	3,037,000	34.1
Other[c]	1,640	9.2	38.8	2,466,000	26.8
Education					
≤High school	4,279	34.0	53.6	14,736,000	43.1
Some college	6,975	44.0	42.1	11,941,000	27.0
College graduate	4,783	22.0	29.5	3,167,000	14.4
Pay grade/rank[d]					
Junior enlisted	2,582	24.0	50.5	9,522,000	39.4
NCO	6,322	49.5	46.6	16,260,000	32.7
Senior NCO	3,191	9.7	29.9	1,697,000	17.4
Warrant officer	399	1.0	26.1	183,698	17.7
Junior officer	1,437	9.4	38.3	1,629,000	17.2
Senior officer	2,106	6.3	19.2	551,776	8.7
Marital status					
Married	9,936	54.5	35.2	11,501,000	21.0
Not married	6,101	45.5	52.9	18,343,000	40.1
Dependents					
Children present	7,174	39.8	30.0	6,311,000	17.0
Children not present	4,331	30.5	49.6	10,375,000	36.6
No children	3,538	29.7	51.2	10,075,000	36.4
Region					
U.S. location	9,878	68.2	39.8	17,290,000	25.2
Overseas location	4,946	22.4	51.1	8,397,000	37.3
Onboard ship	1,213	9.4	49.4	4,157,000	44.0
Type of housing					
Single[e]	2,927	25.4	57.4	11,239,000	47.8
Military family	3,273	19.4	34.2	3,441,000	19.2
Rent/lease/own	8,595	53.9	38.3	11,523,000	23.1
Other[f]	213	1.4	38.2	461,080	34.4
Alcohol intake[g]					
None	3,735	24.1	—	—	—
Nonheavy	8,853	56.2	44.1	7,172,000	13.3
Heavy	2,724	19.8	94.6	21,333,000	112.6

Note: Columns will not add to 100%, based on weighted prevalence.

[a]Binge drinking is defined as consuming on a single occasion ≥5 drinks for men or ≥4 drinks for women.

[b]Sample sizes are weighted to the entire active-duty military population.

[c]*Other* includes Asian, Pacific Islander, and Native American.

[d]Ranks are as follows: junior enlisted, E1–E3; NCO, E4–E6; senior NCO, E7–E9; warrant officer, W1–W5; junior officer, O1–O3; senior officer, O4–O10.

[e]*Single* housing includes military barracks, dormitories, and bachelor quarters.

[f]*Other* housing includes living onboard ships, embassy, and quarters in theater.

[g]Nonheavy alcohol intake is defined as consuming an average of ≤2 drinks per day for men and ≤1 per day for women. Heavy alcohol intake is defined as consuming >2 drinks per day for men and >1 drink per day for women.

NCO, noncommissioned officer

per-capita episodes of binge drinking were estimated for personnel in the Marine Corps (38.3 per person per year) and Army (38.0 per person per year), and the lowest were estimated for the Air Force (17.3 per person per year).

Overall, youth and young adults aged 17–25 years accounted for 46.7% of all ADMP and 67.1% of all binge-drinking episodes (approximately 20 million; Table 1). Correspondingly, yearly per-capita episodes of binge drinking were highest in the younger age groups and declined with increasing age. Young adults aged 21–25 years had the highest number of per-capita episodes (45.7 per person per year), followed by those aged 17–20 years (35.5 per person per year). However, the number of per-capita episodes of binge drinking remained high for personnel aged 26–34 years (21.5 per person per year) and for those aged ≥35 years (14.2 per person per year).

During the study period, men—who accounted for 85.3% of all ADMP—were approximately twice as likely to engage in binge drinking as women (46.6% vs 24.1%), and they reported 93.9% of all binge-drinking episodes (Table 1). However, women—most of whom were of childbearing age (i.e., aged 18–44 years)—had an estimated 12 episodes per person per year. When evaluated by race/ethnicity, whites accounted for 69.0% of the total binge-drinking episodes, but Hispanics had the highest number of per-capita episodes (34.1 per person per year). Junior enlisted and noncommissioned officers (approximately 74% of all ADMP) had the highest prevalence of binge drinking (50.5% and 46.6%, respectively), which resulted in more than 86% of all binge-drinking episodes (approximately 26 million); they had the highest numbers of per-capita episodes (39.4 episodes and 32.7 episodes per person per year, respectively) among the various pay grades and ranks. The prevalence of binge drinking among personnel in the remaining pay grades and ranks ranged from 38.3% for junior officers to 19.2% for senior officers, while the number of per-capita episodes ranged from 17.2 per year for junior officers to 8.7 per year for senior officers. When evaluated by duty station, the highest prevalence of binge drinking was reported by those who were stationed overseas (51.1%), and the highest number of per-capita episodes was estimated for those stationed onboard ships (44 per person per year). When evaluated by type of housing, personnel living in single, on-base housing had the highest prevalence of binge drinking (57.4%) and the highest number of per-capita episodes (47.8 per person per year).

Of the almost 20% of ADMP who were classified as heavy drinkers based on average daily alcohol consumption, 94.6% reported one or more binge-drinking episode in the past 30 days, which extrapolates to approximately 21 million episodes, or 112.6 episodes per person per year (approximately two episodes per week; Table 1). Because of their high frequency of binge drinking, heavy drinkers accounted for 71.5% of all binge-drinking episodes in this population in 2005. Compared to nonheavy drinkers, heavy drinkers were more likely to be in the Army (41%); to be male (78%); aged 21–25 years (49%); non-Hispanic white (70%); a noncommissioned officer (53%); stationed in the U.S. (59%); and with a high school education or less (48%; data not shown).

Active-Duty Military Who Consumed Alcohol

Among the 76% of ADMP who were current drinkers, the prevalence of binge drinking was 56.6% (59.4% for men, 36.7% for women), and the number of per-capita episodes of binge drinking was estimated at 38.9 per person per year (41.8 for men and 18.8 for women; Table 2). The highest prevalence of binge drinking was among current drinkers living in single housing (75.4%). More than two thirds of personnel who were junior enlisted, single, or who had a high school education or less reported binge drinking in the last 30 days. By age, the highest number of per-capita episodes of binge drinking was estimated for drinkers aged 17–20 years (59.8 per person per year). The prevalence of binge drinking among male drinkers ranged from 78.5% for those aged 17–20 years to 35.0% for those aged ≥35 years. Among female drinkers, the prevalence ranged from 47.1% for those aged 17–20 years to 26.0% for those aged 26–34 years; in fact, the prevalence of binge drinking among active-duty women aged ≥35 years (26.2%) was slightly higher than it was for those aged 26–34 years. Although the prevalence and number of per-capita episodes of binge drinking were higher for men than for women, the subgroups at highest risk for binge drinking (e.g., youth and young adults, Hispanics, and those with a high school education or less) were generally similar for both male and female drinkers.

Health and Social Outcomes

Compared with current drinkers who did not binge-drink, binge drinkers were more likely to report not being promoted, getting into a fight and hitting someone, working below their normal level of performance, and drinking and driving (Figure 1). These behaviors and outcomes were increasingly prevalent with more-frequent binge drinking (Figure 1). In terms of adverse outcomes or high-risk behaviors that were explicitly attributed to alcohol consumption, 36.4% of current-drinking ADMP reported at least one outcome or behavior in the past year (e.g., driving after having had too much to drink; Table 3). Specifically, 18.4% of ADMP who drank reported one or more alcohol-attributable problems related to their job performance, 28.1% reported injury-related outcomes or risk behaviors, 2.2% reported problems related to interpersonal

Table 2. Prevalence and per-capita episodes (per person per year) of binge drinking[a] among current drinkers in the U.S. active-duty military, by gender and selected sociodemographic information, 2005

	Men (n=9472)[b]		Women (n=2725)[b]		Total (n=12,197)[b]	
Characteristics	**Prevalence (%)**	**Per-capita episodes**	**Prevalence (%)**	**Per-capita episodes**	**Prevalence (%)**	**Per-capita episodes**
All respondents	59.4	41.8	36.7	18.8	56.6	38.9
Service						
Army	68.5	51.1	43.1	24.8	65.5	48.1
Navy	55.0	40.4	32.6	18.2	52.1	37.6
Marine Corps	68.7	51.6	48.0	26.7	67.8	50.4
Air Force	47.4	25.8	33.0	13.2	44.9	23.6
Age (years)						
17–20	78.5	64.2	47.1	26.7	74.9	59.8
21–25	73.8	58.7	46.7	23.9	69.9	53.7
26–34	54.0	30.1	26.0	12.6	50.6	27.9
≥35	35.0	20.0	26.2	12.9	34.0	19.2
Race/ethnicity						
White	60.9	42.8	38.1	17.4	58.4	40.0
Black	50.7	33.6	31.6	21.3	47.1	31.3
Hispanic	64.2	48.7	42.5	19.0	61.5	45.0
Other[c]	57.6	40.2	34.7	21.0	54.3	37.4
Education						
≤High school	72.9	59.5	47.4	29.6	70.6	56.9
Some college	58.6	38.8	35.9	15.8	55.4	35.5
College graduate	39.3	19.0	29.0	14.9	37.7	18.3
Pay grade/rank[d]						
Junior enlisted	74.4	59.6	47.8	25.6	71.2	55.5
NCO	63.3	45.5	36.1	17.5	59.8	41.9
Senior NCO	42.2	24.5	23.4	14.4	40.6	23.6
Warrant officer	32.0	20.1	37.3	44.2	32.4	22.0
Junior officer	49.1	22.2	36.4	15.8	46.9	21.1
Senior officer	24.7	10.4	18.2	14.2	23.9	10.9
Marital status						
Married	48.6	29.3	27.1	13.4	46.5	27.8
Not married	73.1	57.4	43.5	22.6	68.3	51.8
Dependents						
Children present	42.7	24.7	23.8	9.9	40.4	22.9
Children not present	66.3	49.5	39.3	24.3	62.9	46.4
No children	71.2	52.1	47.1	22.5	67.6	47.7
Region						
U.S. location	55.7	36.1	33.8	16.6	52.6	33.4
Overseas location	67.7	50.5	45.5	24.2	65.4	47.8
Onboard ship	64.9	58.9	43.9	27.5	62.9	56.0
Type of housing						
Single[e]	77.8	65.9	51.8	32.3	75.4	62.7
Military family	48.6	27.7	26.6	10.1	46.9	26.1
Rent/lease/own	52.6	32.4	33.2	16.1	49.4	29.7
Other[f]	53.4	53.7	63.0	25.7	54.9	49.4
Alcohol intake[g]						
Nonheavy	47.0	14.6	24.8	5.2	44.1	13.3
Heavy	95.9	117.7	84.2	67.6	94.6	112.6

Note: Columns will not add to 100%, based on weighted prevalence.
[a]Binge drinking is defined as consuming on a single occasion ≥5 drinks for men or ≥4 drinks for women.
[b]Sample sizes are weighted to the entire active-duty military population.
[c]*Other* includes Asian, Pacific Islander, and Native American.
[d]Ranks are as follows: junior enlisted, E1–E3; NCO, E4–E6; senior NCO, E7–E9; warrant officer, W1–W5; junior officer, O1–O3; senior officer, O4–O10.
[e]*Single* housing includes military barracks, dormitories, and bachelor quarters.
[f]*Other* housing includes living onboard ships, embassy, and quarters in theater.
[g]Nonheavy alcohol intake is defined as consuming an average of ≤2 drinks per day for men and ≤1 per day for women. Heavy alcohol intake is defined as consuming >2 drinks per day for men and >1 drink per day for women.
NCO, noncommissioned officer

relationships, and 7.7% reported criminal-justice problems. The prevalence of these outcomes was even higher for binge drinkers; more than half (52.2%) reported one or more adverse outcomes or high-risk behaviors. In fact, even after adjustment for age and gender, binge drinkers were significantly more likely than nonbinge drinkers to report 21 of 22 studied alcohol-related consequences. For example, binge

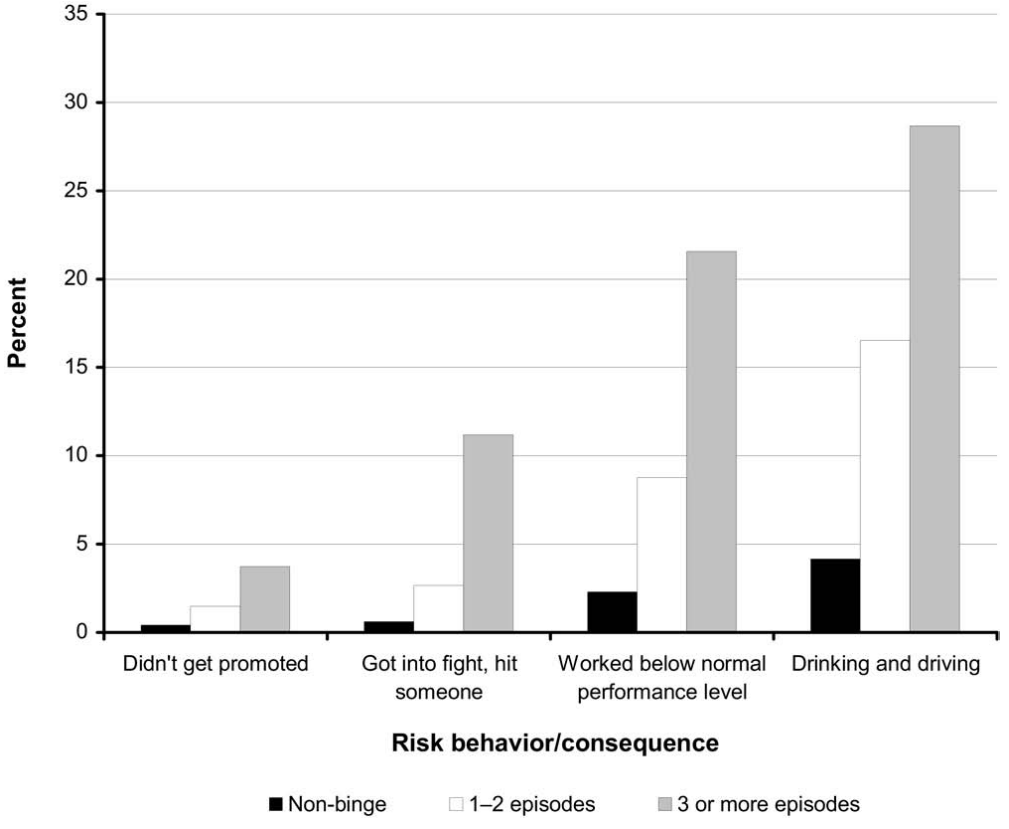


Figure 1. Alcohol-attributable risk behavior or consequences in the last 12 months among current drinkers in the active-duty military, by frequency of binge drinking, 2005

drinkers were almost six times as likely as nonbinge drinkers to report leaving work early or arriving late (OR=5.8; 95% CI=3.78, 8.93); nearly five times as likely to report drinking and driving (OR=4.9; 95% CI=3.68, 6.49); and more than five times as likely to report riding in a car with someone who had been drinking (OR=5.4; 95% CI=4.03, 7.17).

Discussion

This is the first in-depth study of binge-drinking episodes and alcohol-related problems among ADMP in all branches of the military. Almost half of all ADMP reported at least one episode of past-month binge drinking—an estimated 30 million episodes of binge drinking, or about 30 episodes of binge drinking per person per year. Especially high numbers of per-capita episodes of binge drinking were observed among Marines, male drinkers, youth and young adults, Hispanics, junior enlisted and noncommissioned officers, those stationed onboard ships, and those living in on-base housing. Approximately two thirds of binge-drinking episodes involved personnel aged 17–25 years, and more than 70% were reported by heavy drinkers, who reported an average of two binge-drinking episodes per person per week. Alcohol-related problems were reported by more than half of all binge drinkers, and binge drinkers were significantly more likely than nonbinge drinkers to report alcohol-related harms, including job-performance problems, alcohol-impaired driving, and alcohol-related criminal activity.

Studies examining binge drinking among U.S. adults estimate the prevalence to be about 14%,[4] while college students report a higher prevalence (44%).[31] The similarity of binge-drinking rates among college students and the military reflects some of the similarities between these two groups in demographic characteristics (e.g., the predominance of young, unmarried adults); living arrangements (e.g., dormitories or single residences); and environmental exposures (e.g., access to alcohol).[29] Consistent with this, and in contrast to studies of binge drinking in the civilian population,[4] approximately two thirds (67.1%) of the episodes of binge drinking in the military in 2005 involved youth and young adults aged <26 years. However, even among young adults, the age-specific prevalence of binge drinking in the military was higher than that reported among U.S. adults (e.g., 44.4% for active-duty members aged 17–20 years vs 26.1% for comparably aged civilians).[4,32] These findings suggest that interventions directed toward reducing youth access to alcohol, particularly among those who are underage and living in on-base housing or stationed onboard ships, could have a substantial impact on reducing binge drinking in the military population. At the same time, it is important to recognize that high levels of binge drinking were also reported by officers and senior enlisted personnel. Thus, it is important to reduce binge drinking among all ADMP, including those aged ≥26 years.

Although military men accounted for a higher percentage of the total number of binge-drinking episodes, military women, particularly those aged 17–25 years, also had high numbers of per-capita episodes of binge drinking—in fact, several times higher than those reported by similarly aged women in the civilian population.[4,33] This is particularly concerning because almost all female ADMP are of childbearing age (aged 18–44 years).[34] In addition, other studies have found that, as in the U.S. civilian population, more than half of pregnancies in the military are unintended.[35,36]

Table 3. Alcohol-attributable risk behavior or consequences in the last 12 months among current drinkers in the active-duty military, by binge-drinking[a] status, 2005

Category of alcohol-attributable risk behavior or consequence	Current drinkers[b] % (*n*=12,197)	Binge drinkers[b] % (*n*=6,030)	Nonbinge drinkers[b] % (*n*=6,167)	AOR[c] (95% CI) for binge drinkers
Any alcohol-attributable risk behavior or consequence	36.4	52.2	15.7	4.9 (4.12, 5.75)
Alcohol-attributable job-performance problem				
Any job-performance problems	18.4	28.9	4.6	6.5 (4.65, 9.15)
Worked below normal level of performance	11.4	18.0	2.8	6.3 (4.46, 8.98)
Late for work or left work early	7.1	11.3	1.6	5.8 (3.78, 8.93)
Did not come to work at all	2.0	3.3	0.4	6.5 (3.27, 12.85)
Did not get promoted	1.9	3.1	0.4	4.1 (2.45, 6.95)
Got a lower score on efficiency report or performance rating	1.9	3.1	0.3	4.3 (2.04, 9.13)
Drunk while working	4.3	7.2	0.6	7.43 (3.01, 18.3)
Called up during off-duty hours and reported to work drunk	3.6	6.1	0.4	12.6 (5.75, 27.54)
Drank while working, during lunch break, or during work break[d]	5.2	7.2	2.6	2.3 (1.52, 3.39)
Alcohol-attributable injury-related outcome or risk behavior				
Any injury outcome or risk behavior	28.1	40.0	12.5	4.0 (3.36, 4.68)
Caused an accident where someone else was hurt or property damaged	1.3	2.0	0.4	3.7 (1.94, 7.00)
Hurt in accident	1.6	2.5	0.3	5.4 (1.72, 17.00)
Drove after having had too much to drink	16.7	25.3	5.5	4.9 (3.68, 6.49)
Rode with someone who had too much to drink	18.0	27.7	5.5	5.4 (4.03, 7.17)
Drove or rode in a boat after having had too much to drink	4.4	6.9	1.1	5.0 (3.34, 7.53)
Operated machinery after having too much to drink	3.9	6.3	0.8	5.7 (3.27, 10.03)
Diagnosed with an STD[e]	3.7	4.1	3.3	1.0 (0.77, 1.39)
Alcohol-attributable interpersonal problems				
Any interpersonal problems	2.2	3.5	0.4	5.4 (3.07, 9.64)
Spouse of live-in partner threatened to leave me or left me	2.0	3.1	0.4	5.1 (2.86, 8.90)
Was asked to leave or did leave my spouse or live-in partner	1.2	1.9	0.2	8.9 (3.32, 24.02)
Alcohol-attributable criminal-justice problems				
Any criminal justice problems	7.7	12.5	1.4	6.2 (4.00, 9.72)
Got into a fight and hit someone (not family member)	5.2	8.8	0.6	10.2 (5.68, 18.30)
Received UCMJ punishment	2.7	4.3	0.5	4.9 (2.81, 8.63)
Arrested for DUI	1.8	2.7	0.5	3.6 (1.47, 8.81)
Arrested for drinking incident	1.6	2.7	0.3	5.8 (2.38, 14.34)
Spent time in jail, stockade, or brig	1.7	2.6	0.4	4.2 (2.10, 8.50)

[a]Binge drinking is defined as consuming on a single occasion ≥5 drinks for men or ≥4 drinks for women.
[b]Sample sizes are weighted to the entire active duty military population.
[c]Adjusted for age and gender
[d]Incident occurred in the last 30 days.
[e]Question was not asked in relation to their alcohol use.
DUI, driving under the influence; STD, sexually transmitted disease; UCMJ, Uniform Code of Military Justice

Unintended pregnancy is, in turn, associated with delayed pregnancy recognition, which increases the risk that a woman might unintentionally expose a developing fetus to high levels of alcohol if she binge-drinks during her pregnancy, thus increasing the risk of fetal alcohol spectrum disorder and fetal alcohol syndrome.[12,37] Therefore, in addition to reducing youth drinking, special consideration should be given to

preventing binge drinking among female ADMP of childbearing age.[38]

Personnel stationed onboard ships also reported a high prevalence of binge drinking (49%) and a high per-capita number of episodes of binge drinking (37 per person per year). In addition, high numbers of per-capita episodes of binge drinking and total binge drinking episodes were reported by ADMP living in single housing, which is frequently occupied by underage personnel (e.g., military barracks, dormitories, and bachelor quarters). This suggests that important reductions in binge drinking among military personnel could be achieved by enforcing the laws restricting the access of youth to alcohol and enforcing blood alcohol–concentration policies related to drinking and driving on military property.

The finding that more than 70% of all binge-drinking episodes involved ADMP who were heavy drinkers means that approximately one in five ADMP reported binge drinking an average of more than twice per week. This concentration of binge-drinking episodes among heavy drinkers contrasts with studies of binge-drinking episodes in the civilian population, where total episodes were fairly evenly divided between heavy and nonheavy drinkers.[4] These findings suggest that there is a substantial minority (20%) of ADMP who binge-drink frequently and thus put themselves and others at substantially increased risk for a wide range of health and social problems.[31,39,40] These findings further underscore the need to combine policy and environmental approaches for reducing binge drinking with clinical interventions that are designed to screen ADMP for alcohol misuse (i.e., binge drinking) and to provide those who screen positive with brief counseling, referral to specialized treatment, or both, depending on the severity of their alcohol problems.[41]

This study also highlights the potential impact of binge drinking by ADMP on job performance and force readiness. ADMP who reported binge drinking were consistently more likely than nonbinge drinkers to report a wide range of alcohol-attributable problems, including problems with job performance and drinking and driving, both of which were reported by more than one quarter of all binge drinkers. Binge drinkers were also substantially more likely than nonbinge drinkers to report being drunk while working and being called to work during off-duty hours and reporting to work drunk. While binge drinking is also known to be strongly associated with a wide range of health and social problems in the civilian population (e.g., interpersonal violence and sexually transmitted disease), this pattern of alcohol consumption poses special risks in the military setting. For example, the performance of pilots has been shown to be impaired for up to 14 hours after drinking at a level sufficient to achieve a blood alcohol concentration of 0.10 grams per deciliter (g/dL).[42] In addition, serious criminal behavior resulting from binge drinking among military personnel can bring widespread media attention that damages the effectiveness and credibility of the U.S. military as a whole.[43,44] Finally, the high levels of binge drinking among ADMP, particularly among those aged ≤25 years, increase the likelihood of alcohol-related harms and alcohol-use disorders (e.g., alcoholism) following military service.[45–47] The impact of binge drinking in the military on the future drinking behavior of veterans and their families is important, because 13.3% of U.S. adults report current or past military service (CDC Behavioral Risk Factor Surveillance System, unpublished raw data, 2005). Thus, reducing binge drinking among ADMP could have both short- and long-term benefits for both the military and the general population.

Several strengths mark this study, including the large sample size and the ability to assess both alcohol consumption and alcohol-related outcomes among ADMP. Another strength is the use of standardized questions on alcohol use, which are comparable to those used in other large surveys of risk factors. This study also has several imitations. First, binge drinking and related consequences are underreported on surveys; thus, the estimates of the prevalence and frequency of binge drinking, and of the prevalence of alcohol-related problems, were likely conservative.[48,49] Second, although the response rate for this survey (51.8%) is similar to that of other large, population-based surveys,[50] respondents to this survey may differ from nonrespondents. However, based on the characteristics that were used to weight the survey population (e.g., branch of service, race/ethnicity), respondents were representative of ADMP (R. Bray, RTI, personal communication, April 2008). Third, while this study examined several different types of alcohol-related consequences among military personnel, the survey did not ask about a number of important secondhand effects of alcohol use (e.g., being a victim of vandalism, sexual assault).[51,52]

While this study provides new insights into the problem of binge drinking in the military, the problem itself is not new, and has, in fact, been documented in previous surveys of the active-duty population going back more than 20 years.[53] Although previous analyses of the current survey have not assessed the frequency or per-capita episodes of binge drinking, the prevalence of binge drinking is similar to that found in 2005 and only slightly increased (43.2% vs 41.8%) compared to 2002. However, this does not mean that binge drinking is so much a fixture of military life that it is impossible to change. For example, the military has been quite successful in reducing smoking rates among ADMP using a comprehensive public health approach that has included smoking-cessation programs and smoking bans.[54,55] These interventions were successfully implemented even though smoking was historically quite common among ADMP. In contrast, most alcohol programs in the military have tended to focus exclusively on screening for and treating alcoholism, even

though other studies suggest that only a small minority of ADMP meet the diagnostic criteria for alcoholism.[56–59] Nonetheless, small, base-specific, and community programs to reduce underage and binge drinking on military bases have been implemented and appear to be showing promising results.[60,61]

In addition to these programs, potentially effective community-based interventions include increasing the price of alcoholic beverages, particularly on military bases; enforcing and retaining laws prohibiting the sale to or acquisition of alcoholic beverages for underage youth, particularly at alcohol outlets adjacent to military bases; working with communities to limit the density of alcohol outlets; and discouraging drink specials that promote binge drinking (www.thecommunityguide.org).[62,63] In addition, bases should offer alcohol-free social events and increase the availability of recreational activities that do not involve drinking. Further, because a large percentage of young service members live on base, it is important to establish and enforce rules restricting the use of alcohol in dormitories, in single housing, and onboard ships. Finally, although the U.S. Preventive Services Task Force has noted that routine screening for binge drinking in primary care and other treatment settings is effective in reducing these types of behaviors[64] more research on the effectiveness of these interventions in the military setting is required.

The findings and conclusion in this report are those of the authors and do not necessarily represent the official position of the CDC or the U.S. Department of Defense.

No financial disclosures were reported by the authors of this paper.

References

1. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the U.S., 2000. JAMA 2004;291:1238–45.
2. Wechsler H, Dowdall GW, Davenport A, Rimm EB. A gender-specific measure of binge drinking among college students. Am J Public Health 1995;85:982–5.
3. National Institute of Alcohol Abuse and Alcoholism. NIAAA council approves definition of binge drinking. Vol Publication No. 04-5346. NIAAA Newsletter 2004;3:3.
4. Naimi TS, Brewer RD, Mokdad AH, Denny C, Serdula M, Marks JS. Binge drinking among U.S. adults. JAMA 2003;289:70–5.
5. Mann RE, Zalcman RF, Rush BR, Smart RG, Rhodes AE. Alcohol factors in suicide mortality rates in Manitoba. Can J Psychiatry 2008;53:243–51.
6. Davis L, Uezato A, Newell JM, Frazier E. Major depression and comorbid substance use disorders. Curr Opin Psychiatry 2008;21:14–8.
7. Schaffer M, Jeglic EL, Stanley B. The relationship between suicidal behavior, ideation, and binge drinking among college students. Arch Suicide Res 2008;12:124–32.
8. USDHHS. 10th special report to the U.S. Congress on alcohol and health. Washington DC: USDHHS, 2000.
9. Fisher C, Hoffman K, Austin-Lance J, Kao T. The relationship between heavy alcohol use and work productivity loss in active duty military personnel: a secondary analysis of the 1995 Department of Defense Worldwide Survey. Mil Med 2000;165:355–61.
10. Smith GS, Branas CC, Miller TR. Fatal nontraffic injuries involving alcohol: a metaanalysis. Ann Emerg Med 1999;33:659–68.
11. English DR, Holman CDJ, Milne E, et al. The quantification of drug caused morbidity and mortality in Australia. Canberra: Commonwealth Department of Human Services and Health, 1995.
12. Naimi TS, Lipscomb LE, Brewer RD, Gilbert BC. Binge drinking in the preconception period and the risk of unintended pregnancy: implications for women and their children. Pediatrics 2003;111:1136–41.
13. Cook RL, Clark DB. Is there an association between alcohol consumption and sexually transmitted diseases? A systematic review. Sex Transm Dis 2005;32:156–64.
14. Bryant CD. Olive-drab drunks and GI junkies: alcohol and narcotic addiction in the U.S. military. In: Bryant CD, ed. Deviant behavior: occupational and organizational bases. Chicago: Rand McNally, 1974.
15. Holcomb JF. Alcohol and the armed forces. Alcohol Health Res World 1981;82:2–17.
16. Burt MR. Prevalence and consequences of alcohol use among U.S. military personnel, 1980. J Stud Alcohol 1982;43:1097–107.
17. Bray RM, Marsden ME, Peterson MR. Standardized comparisons of the use of alcohol, drugs, and cigarettes among military personnel and civilians. Am J Public Health 1991;81:865–9.
18. Bray RM, Marsden ME, Mazzuchi JF, Hartman RW. Prevention in the Military. In: Ammerman RT, ed. Prevention and societal impact of drug and alcohol abuse. Mahwah NJ: Lawrence Erlbaum Associates, 1999.
19. Zadoo V, Fengler S. The effects of alcohol and tobacco use on troop readiness. Mil Med 1993;158:480–4.
20. James JJ, Dargon D, Day RG. Serum vs breath alcohol levels and accidental injury: analysis among U.S. Army personnel in an emergency room setting. Mil Med 1984;149:369–74.
21. Lesher S, Lee Y. Acute pancreatitis in a military hospital. Mil Med 1989;154:559–64.
22. Bell NS, Amoroso PJ, Yore MM, Smith GS, Jones BH. Self-reported risk-taking behaviors and hospitalization for motor-vehicle injury among active duty Army personnel. Am J Prev Med 2000;18(3S):85–95.
23. Helmkamp J, Kennedy R. National mortality profile of active duty personnel in the U.S. Armed Forces: 1980–1993. Atlanta GA: CDC, 1996.
24. Stout RW, Parkinson MD, Wolfe WH. Alcohol-related mortality in the U.S. Air Force, 1990. Am J Prev Med 1993;9:220–3.
25. Ames GM, Cunradi CB, Moore RS. Military culture and drinking behavior among U.S. Navy careerists. J Stud Alcohol Drugs 2007;68:336–44.
26. Dall TM, Zhang Y, Chen YJ, et al. Cost associated with being overweight and with obesity, high alcohol consumption, and tobacco use within the military health system's TRICARE prime-enrolled population. Am J Health Promot 2007;22:120–39.
27. Bray RM, Kroutil LA, Marsden ME. Trends in alcohol, illicit drug, and cigarette use among U.S. military personnel: 1980–1992. Armed Forces Soc 1995;21:271–93.
28. Bray RM, Bae KH, Federman EB, Wheeless SC. Regional differences in alcohol use among U.S. military personnel. J Stud Alcohol 2005;66:229–38.
29. Bray RM, Hourani L, Rae K, et al. 2002 Department of defense survey of health related behaviors among military personnel. RTI/7841/006-FR. Research Triangle Park NC: Research Triangle Institute, 2003.
30. Dawson D. Methodological issues in measuring alcohol use. Alcohol Res Health 2003;27:18–29.
31. Wechsler H, Lee JE, Kuo M, Seibring M, Nelson TF, Lee H. Trends in college binge drinking during a period of increased prevention efforts. Findings from 4 Harvard School of Public Health college alcohol study surveys: 1993–2001. J Am Coll Health 2002;50:203–17.
32. Miller JW, Naimi TS, Brewer RD, Jones SE. Binge drinking and associated health risk behaviors among high school students. Pediatrics 2007;119:76–85.
33. Tsai J, Floyd RL, Bertrand J. Tracking binge drinking among U.S. childbearing-age women. Prev Med 2007;44:298–302.
34. U.S. General Accounting Office. Military personnel: active duty benefits reflect changing demographics, but opportunities exist to improve. GAO-02-935. Washington DC: The Office, 2002.
35. Clark JB, Holt VL, Miser F. Unintended pregnancy among female soldiers presenting for prenatal care at Madigan Army Medical Center. Mil Med 1998;163:444–8.
36. Robbins AS, Chao SY, Frost LZ, Fonseca VP. Unplanned pregnancy among active duty servicewomen, U.S. Air Force, 2001. Mil Med 2005;170:38–43.
37. Floyd RL, O'Connor MJ, Sokol RJ, Bertrand J, Cordero JF. Recognition and prevention of fetal alcohol syndrome. Obstet Gynecol 2005;106:1059–64.
38. Floyd RL, Sobell M, Velasquez MM, et al. Preventing alcohol-exposed pregnancies. Am J Prev Med 2007;32:1–10.
39. Wechsler H, Nelson TF. Binge drinking and the American college student: what's five drinks? Psychol Addict Behav 2001;15:287–91.
40. Wechsler H, Nelson TF. Relationship between level of consumption and harms in assessing drink cut-points for alcohol research. Alcohol Clin Exper Res 2006;30:922–7.

41. National Institute of Alcohol Abuse and Alcoholism. Helping patients who drink too much. Rockville MD: NIH, 2005.
42. Yesavage JA, Leirer VO. Hangover effects on aircraft pilots 14 hours after alcohol ingestion: a preliminary report. Am J Psychiatry 1986;143:1546–50.
43. Bell NS, Harford T, McCarroll JE, Senier L. Drinking and spouse abuse among U.S. army soldiers. Alcohol Clin Exp Res 2004;28:1890–7.
44. von Zielbaur P. For U.S. troops at war, liquor is spur to crime. The New York Times, March 13, 2007. www.nytimes.com/2007/03/13/world/middleeast/13alcohol.html?ex=1331438400&en=4feaded77662cc8e&ei=5088&partner=rssnyt&emc=rss.
45. Dawson D. Gender differences in the risk of alcohol dependence: U.S. Addiction 1996;91:1831–42.
46. Jennison KM. The short-term effects of unintended long-term consequences of binge drinking in college: a 10-year follow-up study. Am J Drug Alcohol Abuse 2004;30:659–84.
47. Vaillant GE. A 60-year follow-up of alcoholic men. Addiction 2003;98:1043–51.
48. Alanko T. An overview of techniques and problems in the measurement of alcohol consumption. Res Adv Alcohol Drug Probl 1984;8:209–26.
49. Lemmens P, Tan 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.
50. CDC. 2004 Behavioral Risk Factor Surveillance System: data quality report handbook. www.cdc.gov/brfss/technical_infodata/pdf/2004DQRHandbook.pdf.
51. Wechsler H, Moeykens B, Davenport A, Castillo S, Hansen J. The adverse impact of heavy episodic drinkers on other college students. J Stud Alcohol 1995;56:628–34.
52. Wechsler H, Lee JE, Hall J, Wagenaar AC, Lee H. Secondhand effects of student alcohol use reported by neighbors of colleges: the role of alcohol outlets. Soc Sci Med 2002;55:425–35.
53. Bray RM, Hourani L. Substance use trends among active duty military personnel: findings from the U.S. Department of Defense health related behavior surveys, 1980–2005. Addiction 2007;102:1092–101.
54. Klesges RC, Haddock CK, Lando H, Talcott GW. Efficacy of forced smoking cessation and an adjunctive behavioral treatment on long-term smoking rates. J Consult Clin Psychol 1999;67:952–8.
55. Woodruff SI, Conway TL, Edwards CC. Effect of an eight week smoking ban on women at U.S. Navy recruit training command. Tob Control 2000;9:40–6.
56. Polich JM, Orvis BR. Alcohol problems: patterns and prevalence in the U.S. Air Force. R-2308-AF. Santa Monica CA: RAND, 1979.
57. Federman EB, Bray RM, Kroutil LA. Relationships between substance use and recent deployments among women and men in the military. Mil Psychol 2000;12:205–20.
58. Stagliano RF, Richards JD, Kuehr W, Deal CE. Operation Desert Shield/Storm performance of soldiers enrolled in the Alcohol and Drug Abuse Prevention and Control Program. Mil Med 1995;160:631–5.
59. Kolb D, Coben P, Heckman NA. Patterns of drinking and AA attendance following alcohol rehabilitation. Mil Med 1981;146:200–4.
60. Voas RB, Johnson M, Lange J. Permission to cross the border: effective policy reduces high-risk drinking by marines. J Stud Alcohol 2002;63:645–8.
61. U.S. Air Force. Concept of operations (CONOPS): establishing a culture of responsible choices (CoRC). U.S. Department of Defense, 2006. www.afcrossroads.com/websites/corc_docs/CoRC_CONOPS_Feb_2006.pdf.
62. National Research Council and Institute of Medicine. Reducing underage drinking: a collective responsibility. In: Bonnie RJ, O'Connell ME, Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education, eds. Committee on developing a strategy to reduce and prevent underage drinking. Washington DC: The National Academies Press, 2004.
63. Babor T, Caetano R, Casswell S, et al. Alcohol: no ordinary commodity. A summary of the book. Addiction 2003;98:1343–50.
64. U.S. Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. Ann Intern Med 2004;140:554–6.

REFLECTION

A Foundation of Failure

Mike Stillman

Boston Medical Center, Brookline, MA, USA.

J Gen Intern Med 24(4):538–9
DOI: 10.1007/s11606-009-0921-0

At a recent meeting of our medical school's advisory college our Dean of Student Affairs asked 190 first-year students for their definition of professionalism. The majority of these young people, just months out of college, were perhaps unqualified to answer the question, yet enthusiasm overwhelmed inexperience, eager faces glowed with engagement, and willing hands shot up in response.

"It means always doing your best," one student yelled, as the Dean scribbled the answer on a white board.

"Being respectful to everyone at all times," contributed another. A murmur of approval was heard from the faculty.

"Putting patients' needs first," declared a third, and the congress roiled for the next fifteen minutes, bubbling with mostly redundant yet admirably idealist contributions.

For most physicians, "professionalism," an admittedly nebulous term, represents not simply the opposite of "amateurism" but rather the very best practice of medicine. And the students, though overlooking "beneficence" and "sustained commitment to learning" in their description, were right in insisting that it is predicated on proper behavior, hard work, and a stringent commitment to one's patients. Yet several years into practice, it is increasingly clear to me that professionalism is rooted in an intimacy with and a struggle against failure which inevitably makes demands of and weighs on physicians. It is a difficult and costly proposition, and none of us is spared its toll.

As the new student-doctors happily spun their list, optimistically anticipating that good intentions foretell a smooth career, cheerfully unaware of the difficulties awaiting them, my mind turned to two patients I had recently treated and how my role in their care had enhanced and supplied nuance to my own understanding of professionalism.

John, a thoroughly enjoyable bus driver, came to see me after an emergency room visit for pharyngitis. He had been evaluated the night before at a local hospital where his rapid strep test had been negative, and he had been told his problem was "viral." He had been sent home on prednisone and salt water gargles to help reduce his symptoms and told to follow-up with his primary care physician—for whom I was covering—the next day.

Received November 28, 2008
Accepted January 26, 2009
Published online February 14, 2009

I spent very little time with John—much less than I should have—and didn't closely interrogate his history. I took his temperature which was normal, I quickly palpated his neck and found he had no swelling or lymph node enlargement, and I peered into the back of his throat and saw minimal inflammation without pus. John told me he wasn't feeling much better than he had the night before, but as his symptoms hadn't progressed, I endorsed the care he had received and told him to call me if he didn't start to feel better within a day or two.

John's wife, a medical assistant, called me the next afternoon and told me his throat pain was so severe he couldn't even swallow a can of Ensure. She brought him back to my office and on second and closer inspection I noted that his voice had changed, he was having difficulties opening his mouth, and his submandibular tissues were tender. I admitted him to the hospital and arranged for immediate intravenous antibiotics, a CT scan of his neck, and a formal otolaryngology consultation, yet despite these interventions John required a neck debridement and placement of a temporary tracheostomy. He went on to a full recovery, but only after a harrowing and perhaps avoidable stay in the intensive care unit.

Steve, a 50-year-old nurse for whom I've cared for several years, had recently complained to me of an acute cough with low-grade fevers. A chest x-ray was negative but his symptoms persisted for several weeks despite trials of cough medicine, anti-histamines, oral and nasal steroids, and finally an antibiotic. After Steve had been coughing for six weeks I asked him to be tested for Bordetella, but several days after his blood draw, with results still pending, I received a frantic email from his wife. Two days prior, Steve had suddenly appeared "dusky." Upon evaluation in a local emergency room his D-Dimer had been elevated and a CT scan had shown a large central pulmonary embolus. Steve had been anti-coagulated and admitted to the intensive care unit, and a leg ultrasound had demonstrated a sizeable deep vein thrombosis. Even though Steve had never complained of shortness of breath or chest pain and had never been tachypneic or hypoxic, the intensivist believed that the chronic cough had been caused by a cascade of small emboli.

Professionalism certainly hinges on propriety, and in this way the students' intuition was sharp. Yet despite the innumerable pleasures we encounter in our work, it is only through an unrelenting, emotionally costly, and solitary struggle

against error and failure that we are challenged to become as thorough, conscientious, and empathetic as possible. This grim truth, often invisible to young trainees, emerges only with experience and sustained introspection yet lies at the very core of our practice. How else but through missing Ludwig's angina does a young physician re-learn the importance of digging into patients' histories, interrogating and respecting their physical findings, and attending as rapidly as possible to their needs? What better way than by losing sleep over a patient in whom you believe you've missed chronic pulmonary emboli to become vigilant against error? And what could be more humbling and humanizing than being present with patients to whom you've admitted errors or oversights, continuing to be active in their care, and working to maintain their trust in you?

My father, a superb physician, once characterized medicine as a "lone intellectual struggle against disease," and the longer I practice the more regard I have for the words "lone" and "struggle." Physicians who act and practice properly may occasionally be honored by colleagues and adored by patients, and we should allow ourselves to be renewed in those moments. John, it turns out, believes I saved his life and has actually referred family members to my practice, and Steve tried desperately to reassure me—even prior to the state lab calling to report his positive Bordetella test—that I had been appropriately thorough in my evaluation of his cough. Yet medicine remains a difficult and sometimes terrifying pursuit, one in which fear of failure and self-doubt are our close companions and which at times stretches the very best of us to our limits.

Patient Education and Counseling 75 (2009) 411–417

Contents lists available at ScienceDirect

Patient Education and Counseling

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

Health literacy and physical and psychological wellbeing in Japanese adults

Yasuharu Tokuda [a,*], Nobutaka Doba [b], James P. Butler [c], Michael K. Paasche-Orlow [d]

[a] Center for Clinical Epidemiology, St Luke's Life Science Institute, St Luke's International Hospital, Tokyo, Japan
[b] The Life Planning Center, Tokyo, Japan
[c] Department of Environmental Health, Harvard School of Public Health, Boston, USA
[d] Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, USA

ARTICLE INFO

Article history:
Received 29 August 2008
Received in revised form 21 March 2009
Accepted 26 March 2009

Keywords:
Health literacy
Wellbeing
Socioeconomic status
Health Risk Behavior
Chronic illnesses

ABSTRACT

Objective: To determine the prevalence of low health literacy and investigate the relationship between low health literacy and physical and psychological wellbeing in the Japanese general population.
Methods: A web-based cross-sectional survey was conducted in a national sample of Japanese adults. Health literacy was measured by self-report using the validated single-item screening question, "How confident are you filling out forms by yourself?" Wellbeing was measured with the physical and psychological domains of the World Health Organization Quality of Life Assessment-BREF. Effect sizes were computed by dividing the mean difference in scores by the standard deviation of the scores of all participants.
Results: In 1040 adult enrollees (mean age, 57-year-old; women, 52%), there were 161 (15.5%; 95% confidence interval [CI], 13.3–17.7%) with low health literacy. Individuals with low health literacy reported lower physical wellbeing (60.6 vs. 71.7, $p < 0.001$) and psychological wellbeing (59.7 vs. 68.3, $p < 0.001$) compared with those with adequate health literacy. After adjusting for sociodemographic characteristics, health risk behaviors and chronic conditions, these differences were still significant (physical wellbeing, $p < 0.001$; psychological wellbeing, $p < 0.001$). The effect sizes of the difference of scores were moderate for physical wellbeing (−0.55) and also for psychological wellbeing (−0.44).
Conclusion: The prevalence of self-reported low health literacy in Japanese adults is substantial and it is independently associated with poorer physical and mental wellbeing.
Practice implications: Efforts to monitor health literacy and to evaluate causal pathways to poor wellbeing should be encouraged in the Japanese population.

1. Introduction

There is a growing body of evidence supporting the impact of inadequate health literacy on the health of individuals [1]. Inadequate health literacy has been linked to lower use of preventive services [2], delayed care-seeking when symptomatic [3], poor understanding of one's medical condition [4], low adherence to medical instructions [5], poor self-care [6], higher healthcare costs [7] and increased mortality [8,9]. In a recent report, 48% of U.S. adults lack the reading and numeracy skills to fully understand and act on health information [1]. Similarly, substantial portions of European populations have also been shown to have inadequate health literacy, despite the small number of publications related to the impact of inadequate health literacy in European countries [10].

To date, little research has been conducted on the prevalence of inadequate health literacy in other parts of the world, such as Asian countries including Japan. Japan in particular is well-known for having a high standard of educational attainment, including a high proportion that obtain university degrees [11]. Educational attainment, along with race/ethnicity and age, has been shown to be the leading demographic predictors of health literacy in the U.S. [12,13]. Whereas 15% of U.S. adults do not have a high school diploma and 19% completed at least a 4-year university degree [14], only 8% of Japanese adults lack a high school diploma and fully 34% have completed at least a 4-year university degree [15]. Thus, there might be a lower prevalence of inadequate health literacy in Japan, although no research has been conducted to determine the prevalence of inadequate health literacy among Japanese adults.

English is a phonographic language in which phonemes, which do not intrinsically represent any particular meaning, are brought together to represent words. On the other hand, Japanese is a

* Corresponding author at: Center for Clinical Epidemiology, St Luke's Life Science Institute, St Luke's International Hospital, 9-1 Akashi-cho, Chuo City, Tokyo 104-8560, Japan. Tel.: +81 3 5550 2426; fax: +81 3 5550 2426.
E-mail address: tokuyasu@orange.ocn.ne.jp (Y. Tokuda).

0738-3991/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.pec.2009.03.031

logographic language which uses a mixture of Chinese grapheme characters (*Kanji*) and two syllabary character systems (*Hiragana* and *Katakana*), to depict concepts, i.e., the images have intrinsic and somewhat flexible meaning. A phonographic script primarily represents words as units of sound. A logographic script primarily represents words as visual images [16]. Further, in Japanese, basic characters are often combined to communicate complex ideas. Consequently, most people can easily conceive the ideas close to the correct meaning of written words and sentences without prior knowledge.

Historically, the Japanese began adopting written language in the third century A.C.E. with Kanji through the Korean peninsula from China and then originally developed the Hiragana and Katakana in the ninth century. Currently, the Japanese Ministry of Education designates a list of 1006 Kanji (*Kyoiku Kanji* or Education Kanji) as the learning objective for all elementary school children. These and an additional 939 Kanji are designated as the learning objective for all junior high school students (*Joyo Kanji* or Ordinary Kanji). Logographic and syllabic characteristics of Japanese may be one cause of the high written literacy rate in Japan. In fact, 99.8% of junior high school graduates have mastered the Hiragana and Katakana [15,17].

Due to the linguistic differences between English and Japanese, tools that have been developed in English language to directly measure health literacy cannot be simply translated. While simple, standardized assessment tools for readability and functional health literacy should be developed in Japanese, a recently developed surrogate measure for inadequate health literacy can be employed [18–20]. This one-item screening question was validated in recent studies as a measure of health literacy [18–20]; areas under the receiver operating characteristics curve for this single question were 0.84 based on the rapid estimate of adult literacy in medicine (REALM) [19] and 0.80 for the test of functional health literacy in adults (STOFHLA) [18]. Chew et al. have also shown that additional questions did not significantly increase the accuracy in detecting inadequate health literacy [19].

Despite accumulating evidence on health issues related to health literacy in the U.S. and European countries [6,21,22], a recent study indicated no association between health literacy and health status in ethnic minorities in the U.S. [23]. To determine the association between functional health literacy and physical and mental health status in Latinos and African Americans, Guerra et al conducted a cross-sectional study that used the STOFHLA and SF-12 in a sample of about 1300 Medicaid and/or Medicare Latino and African American adult patients at community clinics in Philadelphia and found that health literacy was not significantly associated with physical or mental health status thus questioning the generalizability to a sample of ethnic minorities of the perceived link between inadequate health literacy and poor health status [23]. Thus, research on the potential link between health literacy and health status is needed for people living outside the U.S. or Europe. Thus, the objective of this study was to estimate the prevalence of inadequate health literacy by examining self-reported low health literacy and to investigate the relationship between low health literacy and health status in the Japanese general population.

2. Methods

2.1. Study participants

The data for this study was collected from responses to a national cross-sectional on-line survey conducted from July 3 to July 8, 2008. No personal identifying information was collected (such as name or address) and institutional review board approval was obtained from the National Institute of Japanese Language. All areas in Japan were stratified into 10 regions, including 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 in the Yahoo JAPAN Co. (Tokyo, Japan), through probability sampling proportionate to age and gender, similar to the national census data of population distributions for 30–90-year-old in 2007. Inclusion criteria were Japanese adults aged 30-year-old or older. Many Japanese of age of 20s go to colleges or other schools of higher education. Thus, for measuring the final educational attainment in people, we recruited people aged 30 or older. Exclusion criteria were any types of healthcare workers, such as physicians, nurses, hospital workers, or public health workers. No gifts or payments were given for participating in the survey.

2.2. Data collection

The survey gathered anonymous data for demographics and socioeconomics as well as responses to the questionnaire for health-related quality-of-life (HRQOL) and health literacy. Demographic data included age, gender, annual income, education and occupation. The cutoff points for annual income of 2, 4, 6 and 8 million Japanese Yen (JY) were used to generate five income categories (*Note*: the average exchange rate to 1 U.S. dollar in July 2008 was about 100 JY). Although poverty benefits were provided in 2008 by the government to those with an annual income of less than about 1.5 million JY for a single person and less than about 2 million JY for a couple in 2008 [24], the National Tax Agency regards an income of 2 million JY as the cutoff level for low-wage workers and reports the distribution of income using cutoffs of 2, 4, 6 and 8 million JY. For educational attainment, five categories were used (did not graduate high school, high school graduate, vocational school, short-term college, and university graduate/Masters/PhD). For occupational status, five categorical levels were used, including persons working full time, homemaker, those working part time, retired, and those 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, or depression) [22].

The HRQOL was assessed using the physical and psychological wellbeing domains (a total of 13 items) of the shortened Japanese version of the World Health Organization Quality of Life-BREF (WHOQOL-BREF). The WHOQOL-BREF raw scores were transformed into scores from 0 to 100 with the lowest score of zero and the highest score of 100. The WHOQOL-BREF transformed domain scores have demonstrated good discriminant validity, content validity, and test–retest reliability [25]. The instrument is considered a valid and reliable measure for assessing HRQOL in different populations [25].

For measuring health literacy, we asked the single item screening question: "How confident are you filling out forms by yourself?" with five Likert responses of "not at all", "a little bit", "somewhat", "extremely", or "quite a bit". The threshold optimizing both sensitivity (83%) and specificity (82%) for this single item as compared to the REALM criteria was at the response of "somewhat" or less [19]. Thus, we used this single item with this threshold for dichotomizing participants into adequate or low levels of self-reported health literacy.

Two additional items for measuring self-reported health literacy were also included in this study for confirming a concurrent validity of the aforementioned item. In these two items, we asked: "How often do you have problems learning about your medical condition because of difficulty understanding written information?" ("problems learning") and "How often do you have someone help you read hospital materials?" ("help read")

with five Likert responses of "never", "occasionally", "sometimes", "often", or "always". These items have also shown good validity profiles for measuring health literacy in comparison to the STOFHLA and the REALM despite relatively lower accuracy compared to the item of "confident with forms" [19,20].

2.3. Statistical analysis

Spearman's correlation coefficients were computed to confirm the concurrent validity of the items for measuring self-reported health literacy. Descriptive statistics for sociodemographic variables and current and past smoking, current alcohol use, and chronic conditions were calculated and presented as counts with proportions for all participants. The differences of proportions in these variables between participants with adequate health literacy and those with low health literacy were estimated using Chi-square test or trend test where appropriate.

Differences of the mean scores in the two wellbeing domains between the groups by self-reported health literacy level were estimated with unadjusted linear regression models. Furthermore, adjusted linear regression models for the physical and psychological wellbeing were constructed, including age, gender, income, education, occupation, smoking status, current alcohol use, chronic conditions, and health literacy. Because of possible over-adjustment by including education in a regression model focused on health literacy [22], adjusted models without education were also constructed. Model fit of multivariate models was assessed with the adjusted *R*-square.

Since the WHOQOL-BREF does not have the well-defined values for minimal clinically important differences, we used the distribution-based approach for estimating the effect size. Thus, to examine the clinical significance of the differences between the groups, we computed effect sizes by dividing the mean difference in scores by the SD for all participants [26]. For interpretation of effect sizes, we followed the recent criteria of <0.3 as small, 0.3–0.8 as moderate and >0.8 as large effect sizes [27–29]. Statistical analyses were performed using STATA 10.0 (College Station, TX). Two-tailed p-values <0.05 were considered statistically significant.

3. Results

Of 2500 subjects randomly selected from the on-line panel, 1074 participated in the study (a response rate, 43.0%). Among these, data for 1040 persons were available for our analysis and were considered as the final sample. Table 1 shows sociodemographic characteristics of all participants. The mean age was 57-year-old (range, 30–90) and 52% were women.

We found 161 participants (15.5%; 95% confidence interval [CI], 13.3–17.7%) to have self-reported low health literacy in this study, based on the item "confident with forms" as the criterion. The three self-reported health literacy items had a significant and moderate positive correlation. Spearman's correlation coefficients for the item "confident with forms" were 0.384 ($p < 0.001$) to the item "problems learning" and 0.331 ($p < 0.001$) to the item "help read".

Table 1 also presents the distributions of sociodemographic characteristics by self-reported health literacy group (low vs. adequate). Educational attainment was significantly different between the two groups ($p = 0.002$ for trend). Participants with

Table 1
Sociodemographics of study participants by health literacy level[a].

Characteristic	All participants		Adequate health literacy (1)		Low health literacy (2)		*p*-Value (1) vs. (2)
	n	%	*n*	%	*n*	%	
N	1040	100%	879	100%	161	100%	
Age (years)							0.403
30–39	183	17.6%	157	17.9%	26	16.1%	
40–49	160	15.4%	135	15.4%	25	15.5%	
50–59	224	21.5%	190	21.6%	34	21.1%	
60–69	238	22.9%	202	23.0%	36	22.4%	
70–79	181	17.4%	153	17.4%	28	17.4%	
80 or older	54	5.2%	42	4.8%	12	7.5%	
Gender							0.397
Men	497	47.8%	425	48.4%	72	44.7%	
Women	543	52.2%	454	51.6%	89	55.3%	
Income (Japanese Yen)							0.072
<2 million	92	8.8%	70	8.0%	22	13.7%	
2–3.99 million	264	25.4%	219	24.9%	45	28.0%	
4–5.99 million	290	27.9%	252	28.7%	38	23.6%	
6–7.99 million	160	15.4%	138	15.7%	22	13.7%	
8 million or more	234	22.5%	200	22.8%	34	21.1%	
Education							0.002
<Grade 12	51	4.9%	34	3.9%	17	10.6%	
High school graduate	379	36.4%	316	35.9%	63	39.1%	
Vocational school	107	10.3%	91	10.4%	16	9.9%	
Short-term college	139	13.4%	116	13.2%	23	14.3%	
University graduate/Master/PhD	364	35.0%	322	36.6%	42	26.1%	
Working status							0.161
Working full time	445	42.8%	382	43.5%	63	39.1%	
Homemaker	273	26.3%	227	25.8%	46	28.6%	
Working part time	91	8.8%	76	8.6%	15	9.3%	
Retired	135	13.0%	117	13.3%	18	11.2%	
Currently not working	96	9.2%	77	8.8%	19	11.8%	

Low (not at all, a little bit, somewhat) and adequate (extremely, quite a bit).

[a] Health literacy level was based on the question of "confident with forms": how confident are you filling out forms by yourself?

Table 2
Health behavior and chronic conditions of study participants by health literacy level[a].

Characteristic	All participants		Adequate health literacy (1)		Low health literacy (2)		*p*-Value (1) vs. (2)
	n	%	*n*	%	*n*	%	
N	1040	100%	879	100%	161	100%	
Smoking							0.336
Current	200	19.2%	162	18.4%	38	23.6%	
Former	247	23.8%	214	24.3%	33	20.5%	
Never	593	57.0%	503	57.2%	90	55.9%	
Current alcohol use							0.993
None to light	588	56.5%	494	56.2%	94	58.4%	
Moderate	407	39.1%	350	39.8%	57	35.4%	
Heavy	45	4.3%	35	4.0%	10	6.2%	
Chronic condition							
Cancer	38	3.7%	33	3.8%	5	3.1%	0.687
Cardiovascular disease	21	2.0%	18	2.0%	3	1.9%	0.878
Hypertension	221	21.3%	184	20.9%	37	23.0%	0.559
Diabetes	55	5.3%	44	5.0%	11	6.8%	0.341
Arthritis	45	4.3%	34	3.9%	11	6.8%	0.089
Asthma or COPD	29	2.8%	23	2.6%	6	3.7%	0.432
Depression	33	3.2%	24	2.7%	9	5.6%	0.057

Low (not at all, a little bit, somewhat) and adequate (extremely, quite a bit). COPD = chronic obstructive pulmonary disease.
[a] Health literacy level was based on the question of "confident with forms": how confident are you filling out forms by yourself?

low health literacy represented a higher proportion of those with a lower education attainment, while participants with adequate health literacy represented a higher proportion of those with a higher education attainment. Table 2 shows the distributions of health behavior and chronic conditions of the study participants by self-reported health literacy level. There were no significant differences of smoking status, alcohol use and prevalence of chronic conditions between the two groups.

The mean scores were 70.0 (S.D., 17.6) for physical wellbeing and 66.9 (S.D., 17.1) for psychological wellbeing. Fig. 1 presents the distributions of physical wellbeing by health literacy level. Individuals with low health literacy reported significantly lower mean physical wellbeing scores compared with those with adequate health literacy (60.6 vs. 71.7, respectively) (Table 3). The unadjusted difference was −11.1 with 95% CI of −14.0 to −8.3. The effect size of the unadjusted difference was 0.63, which represents an effect size of moderate magnitude.

Fig. 1 also depicts the distributions of the psychological wellbeing by health literacy level. Individuals with low health

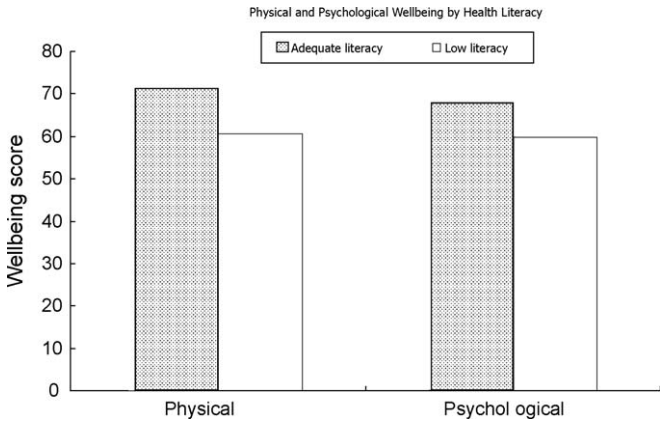


Fig. 1. Physical and psychological wellbeing by health literacy level. A bar indicates the mean value of scores by health literacy level.

Table 3
Unadjusted and adjusted linear regressions for physical and psychological wellbeing.

Measure	Unadjusted model	Full adjusted model[a]	Adjusted model without education[b]
Physical wellbeing, mean ± SD			
Adequate health lietracy	71.7 ± 17.0	–	–
Low health lietracy	60.6 ± 17.9	–	–
Difference of physical wellbeing (95% CI)	−11.1 (−14.0 to −8.3)	−9.6 (−12.4 to −6.9)	−9.8 (−12.5 to −7.1)
p-Value	<0.001	<0.001	<0.001
Adjusted *R*-square	0.052	0.170	0.172
Effect size	−0.63	−0.55	−0.56
Psychological wellbeing, mean ± SD			
Adequate health lietracy	68.3 ± 16.9	–	–
Low health lietracy	59.7 ± 16.6	–	–
Difference of psychological wellbeing (95% CI)	−8.5 (−11.4 to −5.7)	−7.5 (−10.3 to −4.7)	−7.6 (−10.4 to −4.9)
p-Value	<0.001	<0.001	<0.001
Adjusted *R*-square	0.032	0.103	0.105
Effect size	−0.50	−0.44	−0.44

SD = standard deviation: CI = confidence interval.
[a] Adjusted for age, gender, income, education, occupation, smoking status, alcohol use, and chronic conditions.
[b] Adjusted for age, gender, income, occupation, smoking status, alcohol use, and chronic conditions.

literacy reported significantly lower mean psychological wellbeing scores compared with those with adequate health literacy (59.7 vs. 68.3, respectively) (Table 3). The unadjusted difference was −8.5 with 95% CI of −11.4 to −5.7. The effect size of the unadjusted difference was 0.50, which represents an effect size of moderate magnitude.

Table 3 presents the results of unadjusted and adjusted linear regression models for physical and psychological wellbeing scores. Self-reported low health literacy was significantly associated with a lower physical wellbeing in the full-adjusted model (difference, −9.6). Similarly, self-reported low health literacy was significantly associated with a lower psychological wellbeing in the full-adjusted model (difference, −7.5). Based on the full-adjusted models, the effect sizes of the difference in wellbeing by health literacy levels were −0.55 (moderate) for physical wellbeing and −0.44 (moderate) for psychological wellbeing. Exclusion of education from the models did not significantly influence the relationships between health literacy and physical or psychological wellbeing.

4. Discussion and conclusion

4.1. Discussion

To our knowledge, this is the first study to explore the prevalence of self-reported low health literacy in Japan. Our results indicate that 15.5% of Japanese adults may have low health literacy. This is a substantial portion of the population and should prompt planning for further epidemiologic studies and possible interventions in Japan. By comparison, while 15.5% is lower than what has been reported in other countries such as the U.S. [1], this figure is higher than a recent national sample of British adults (11.4%) [30], and a primary care population in Canada (9%) [31]. However, since the self reported measure was used in the current study and different measures were used in other countries, it may need to be careful when comparing these figures.

In addition, after controlling for age, gender, income, education, occupation, smoking, alcohol use and chronic conditions, Japanese individuals with low health literacy reported significantly lower physical and psychological wellbeing than those with adequate health literacy. The magnitudes of the differences of wellbeing between the groups were clinically important according to Cohen's criteria for interpreting effect sizes. Based on previous reports using the WHOQOL-BREF, the magnitude of the difference of physical wellbeing was comparable to that of having survived myocardial infarction (−11.3) [26], and a diagnosis of lung cancer (−15) [32]. The magnitude of the difference of psychological wellbeing was a half of that of having a diagnosis of bipolar disorder (−15) [33].

Comparison of the regression models with and without education as a covariate showed that both health literacy and education may, to a certain extent, operate in the same causal pathway for health status [22]. Specifically, in the model without education as a covariate, the magnitude of the association between health literacy and our health status measures increased by a small margin in comparison to our model with education as a covariate. Though this effect is small, it is likely an indication of over-adjustment [22].

The current findings, linking low health literacy and poor physical and psychological wellbeing in Japan is consistent with previous studies showing the independent association between low health literacy and poor self-rated health status, including physical and psychological health functioning [21,22,34–36]. Since recent studies also show that inadequate health literacy increases mortality in the U.S. elderly [8,9], prospective studies examining mortality are also needed in Japan. The lack of significant differences of health behavior and chronic conditions of the study participants by health literacy level was consistent with the prior report [34].

There are several potential causal pathways linking low health literacy to poor physical and psychological wellbeing [22,37]. First, individuals with low health literacy are less likely to access and utilize healthcare services, such as recommended vaccination and cancer screening programs [2,38,39]. Second, low health literacy is related to poor health knowledge, including important knowledge about prevention and chronic diseases [4,40,41] and this lack of knowledge can lead to lower adherence to medical instructions [5,6,42,43]. Third, communication between individuals with low health literacy and health care providers within medical encounters may be ineffective, since physicians often do not communicate at a level that is understood by patients with low health literacy [44]. Similarly, educational materials distributed in clinics and hospitals are mostly difficult to understand for those with low health literacy [45,46]. Fourth, the quality of self-care may be poorer among those with low health literacy at the time of both acute symptomatic episodes and chronic illnesses [47]. Consequently, all of these factors could contribute to poor wellbeing among individuals with low health literacy [36]. On the other hand, there are other studies that do not show these relationships, such as a study indicating higher adherence among HIV-infected patients with low health literacy [48]. However, none of these issues, regarding the potential mechanisms that link low health literacy to worse health outcomes have been evaluated in Japan.

There are several limitations in our study. First, because our study was based on cross-sectional data, a causal relationship between low health literacy and physical and mental wellbeing cannot be established. Second, we used only the single-item screener to measure self-reported health literacy in study participants. While the test sensitivity, specificity and reliability of this item are considered good compared to the STOFHLA and the REALM, which are currently the most commonly used measures of health literacy, the operating characteristics of this screening question have not been evaluated in Japanese. Japanese language versions of the STOFHLA and the REALM cannot be implemented because of the logographical nature of Japanese. Until a new test paradigm is developed and validated for health literacy in Japanese, the field will need to be initiated with indirect test measures, as we have done. Third, we assessed the presence of chronic conditions, based not on medical records, but on the self-reported questionnaire. However, a high level of agreement has been shown between self-reported survey results of chronic conditions and medical records [49], and there is little association between education and the validity of self-reported chronic conditions [50].

Fourth, while the National Surveys in the U.S. (NAAL) and other countries (IALS) were interviewer-facilitated household surveys, data collection through a web-based survey is likely to underestimate the prevalence of inadequate health literacy. Indeed, the educational attainment and income of our study participants is relatively higher than that of the general population in Japan (Grade <12 proportions: 5% in our study participants vs. 8% in the general population by the government report) [15]. An adjusted estimate for the prevalence of low health literacy would be increased to about 18%, if our sample had included 8% of people not finishing high school instead of 5% and if all of those had low health literacy (the worst case scenario). The appropriateness of a web-based survey for estimating the prevalence of health literacy may need to be evaluated further to estimate the magnitude of this effect.

Fifth, it has recently been shown that literacy influences the validity of Likert scale measures [51]. This work, however, was conducted with non-readers, who were not included in our self-

administered survey and were not at all included in the International Adult Literacy Survey, the 1992 U.S. National Adult Literacy Survey, and the 2003 U.S. National Assessment of Adult Literacy (NAAL). Indeed, in the NAAL, 3% of all potential subjects were excluded due to being unable to participate in the survey and this portion is typically excluded from all discussion of the NAAL results. It is unclear what portion of the Japanese population would have been similarly excluded if the current study had been conducted as an in-person survey. Finally, sixth, low literacy might directly affect a person's ability to complete a wellbeing questionnaire. Lower HRQOL in low literacy might have resulted from this group's inability to accurately report their wellbeing compared to those with adequate literacy. This issue could have been overcome by administering the WHOQOL via interview.

4.2. Conclusion

In conclusion, this study explores the prevalence of self-reported low health literacy in the general Japanese population for the first time and provides the first evidence on the relationship between low health literacy and poor physical and psychological wellbeing in Japan. Though this method has been validated, the estimate we provide is based on a single-item self-report measure for health literacy. Appropriate epidemiological tools for direct measurement of health literacy should be developed in Japanese. In addition, to develop effective public health interventions, future studies are needed to examine how people with low health literacy in Japan recognize and respond to health problems and the health care system.

4.3. Practice implications

Japanese clinicians and public health officials may need to pay attention to our results which indicate a substantial proportion of self-reported low health literacy among Japanese adults. Although our estimate was based on a single self-report screening item, our Internet-based data collection likely led to an underestimation of this problem. Japanese clinicians and public health officials may be able to simplify their patient education materials and improve their communication techniques. In particular, further examination of patient comprehension is warranted. These kinds of changes are likely to help all patients and be particularly beneficial to patients with low health literacy.

Conflict of interest

We declare no actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations within 3 years of beginning the submitted work that could inappropriately influence, or be perceived to influence, our work.

Acknowledgements

We thank all staff of the National Institute of Japanese Language and Life Planning Center for their support of our research.

Role of funding: This study was supported by a Research Grant from the Pfizer Health Research Funds. The funding source had no role in data collection and analysis of this study.

References

[1] Nielsen-Bohlman L. Institute of medicine. Health literacy: a prescription to end confusion. National Academy Press; 2004.
[2] 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.
[3] Bennett CL, Ferreira MR, Davis TC, Kaplan J, Weinberger M, Kuzel T, Seday MA, Sartor O. Relation between literacy, race, and stage of presentation among low-income patients with prostate cancer. J Clin Oncol 1998;16:3101–4.
[4] Gazmararian JA, Williams MV, Peel J, Baker DW. Health literacy and knowledge of chronic disease. Patient Educ Couns 2003;51:267–75.
[5] 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.
[6] Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, Palacios J, Sullivan GD, Bindman AB. Association of health literacy with diabetes outcomes. J Amer Med Assoc 2002;288:475–82.
[7] Weiss BD, Palmer R. Relationship between health care costs and very low literacy skills in a medically needy and indigent Medicaid population. J Am Board Fam Pract 2004;17:44–7.
[8] Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA, Huang J. Health literacy and mortality among elderly persons. Arch Intern Med 2007;167:1503–9.
[9] Sudore RL, Yaffe K, Satterfield S, Harris TB, Mehta KM, Simonsick EM, Newman AB, Rosano C, Rooks R, Rubin SM, Ayonayon HN, Schillinger D. Limited literacy and mortality in the elderly: the health, aging, and body composition study. J Gen Intern Med 2006;21:806–12.
[10] Kondilis BK, Kiriaze IJ, Athanasoulia AP, Falagas ME. Mapping health literacy research in the European Union: a bibliometric analysis. PLoS ONE 2008;3:e2519.
[11] Inoguchi T. Values and life styles in urban Asia. Tokyo: Institute of Oriental Culture; 2005.
[12] Hanchate AD, Ash AS, Gazmararian JA, Wolf MS, Paasche-Orlow MK. The demographic assessment for health literacy (DAHL): a new tool for estimating associations between health literacy and outcomes in national surveys. J Gen Intern Med 2008;23:1561–6.
[13] 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.
[14] Kutner M, Greenberg E, Baer J. A first look at the literacy of America's adults in the 21st century: national assessment of adult literacy. Washington, DC: National Center for Education Statistics; 2005.
[15] In: Tokyo; 2002. STATISTICAL ABSTRACT. Ministry of Education, Culture, Sports, Science and Technology.
[16] Taylor I, Olson DR. Scripts and literacy: reading and learning to read alphabets, syllabaries, and characters. Kluwer Academic Publishers; 1995.
[17] Wagner DA, Venezky RL, Street BV. Literacy: an international handbook. Boulder, CO: Westview Press; 1999.
[18] Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004;36:588–94.
[19] Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, Bradley KA, Nugent SM, Baines AD, Vanryn M. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008;23:561–6.
[20] 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.
[21] Baker DW, Parker RM, Williams MV, Clark WS, Nurss J. The relationship of patient reading ability to self-reported health and use of health services. Am J Public Health 1997;87:1027–30.
[22] Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional health status among older adults. Arch Intern Med 2005;165:1946–52.
[23] Guerra CE, Shea JA. Health literacy and perceived health status in Latinos and African Americans. Ethn Dis 2007;17:305–12.
[24] Shirahase S. Japanese income inequality by household types in comparative perspective: Maxwell school of citizenship and public affairs. New York: Syracuse University; 2001.
[25] Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004;13:299–310.
[26] Norekval TM, Wahl AK, Fridlund B, Nordrehaug JE, Wentzel-Larsen T, Hanestad BR. Quality of life in female myocardial infarction survivors: a comparative study with a randomly selected general female population cohort. Health Qual Life Outcomes 2007;5:58.
[27] 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.
[28] Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102–9.
[29] Contopoulos-Ioannidis DG, Karvouni A, Kouri I, Ioannidis JP. Reporting and interpretation of SF-36 outcomes in randomised trials: systematic review. BMJ 2009;338:a3006.
[30] von Wagner C, Knight K, Steptoe A, Wardle J. Functional health literacy and health-promoting behaviour in a national sample of British adults. J Epidemiol Community Health 2007;61:1086–90.
[31] Smith JL, Haggerty J. Literacy in primary care populations: is it a problem? Can J Public Health 2003;94:408–12.
[32] Mohan A, Mohan C, Bhutani M, Pathak AK, Pal H, Das C, Guleria R. Quality of life in newly diagnosed patients with lung cancer in a developing country: is it important? Eur J Cancer Care (Engl) 2006;15:293–8.

[33] Brissos S, Dias VV, Carita AI, Martinez-Aran A. Quality of life in bipolar type I disorder and schizophrenia in remission: clinical and neurocognitive correlates. Psychiatr Res 2008;160:55–62.
[34] Lincoln A, Paasche-Orlow MK, Cheng DM, Lloyd-Travaglini C, Caruso C, Saitz R, Samet JH. Impact of health literacy on depressive symptoms and mental health-related: quality of life among adults with addiction. J Gen Intern Med 2006;21:818–22.
[35] Dewalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med 2004;19:1228–39.
[36] Cho YI, Lee SY, Arozullah AM, Crittenden KS. Effects of health literacy on health status and health service utilization amongst the elderly. Soc Sci Med 2008;66:1809–16.
[37] Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31:S19–26.
[38] Sudore RL, Mehta KM, Simonsick EM, Harris TB, Newman AB, Satterfield S, Rosano C, Rooks RN, Rubin SM, Ayonayon HN, Yaffe K. Limited literacy in older people and disparities in health and healthcare access. J Am Geriatr Soc 2006;54:770–6.
[39] 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.
[40] 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.
[41] 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.
[42] 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.
[43] 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.
[44] Lindau ST, Tomori C, McCarville MA, Bennett CL. Improving rates of cervical cancer screening and Pap smear follow-up for low-income women with limited health literacy. Cancer Invest 2001;19:316–23.
[45] Hearth-Holmes M, Murphy PW, Davis TC, Nandy I, Elder CG, Broadwell LH, Wolf RE. Literacy in patients with a chronic disease: systemic lupus erythematosus and the reading level of patient education materials. J Rheumatol 1997;24:2335–9.
[46] Davis TC, Bocchini Jr JA, Fredrickson D, Arnold C, Mayeaux EJ, Murphy PW, Jackson RH, Hanna N, Paterson M. Parent comprehension of polio vaccine information pamphlets. Pediatrics 1996;97:804–10.
[47] Paasche-Orlow MK, Riekert KA, Bilderback A, Chanmugam A, Hill P, Rand CS, Brancati FL, Krishnan JA. Tailored education may reduce health literacy disparities in asthma self-management. Am J Respir Crit Care Med 2005;172:980–6.
[48] Paasche-Orlow MK, Cheng DM, Palepu A, Meli S, Faber V, Samet JH. Health literacy, antiretroviral adherence, and HIV-RNA suppression: a longitudinal perspective. J Gen Intern Med 2006;21:835–40.
[49] Kriegsman DM, Penninx BW, van Eijk JT, Boeke AJ, Deeg DJ. Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. J Clin Epidemiol 1996;49:1407–17.
[50] 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;52:123–7.
[51] Chachamovich E, Fleck MP, Power M. Literacy affected ability to adequately discriminate among categories in multipoint Likert scales. J Clin Epidemiol 2009;62:37–46.

Dig Dis Sci (2009) 54:2694–2698
DOI 10.1007/s10620-008-0659-2

ORIGINAL ARTICLE

Sociodemographic Trends in National Ambulatory Care Visits for Hepatitis C Virus Infection

Judith I. Tsui · Judith Maselli · Ralph Gonzales

Received: 10 September 2008 / Accepted: 28 November 2008 / Published online: 23 December 2008

Abstract Poor and non-white patients are disproportionately infected with the hepatitis C virus (HCV). The objective of this research is to determine sociodemographic patterns of HCV-related ambulatory care visits over time. Data from the National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Ambulatory Medical Care Survey-Outpatient (NHAMCS-OPD) for the years 1997–2005 were analyzed in 3-year intervals. Demographic and other variables were compared for each period, and multivariable logistic regression was performed to examine whether the likelihood of a visit being HCV-related (versus non-HCV) was independently associated with (1) race and/or (2) Medicaid status over time. The total number of HCV-related ambulatory visits more than doubled from 3,583,585 during the years 1997–1999 to 8,027,166 during 2003–2005. During this time, the proportion of non-whites and Medicaid recipients presenting for HCV-related visits approximately doubled (non-whites: 16% vs. 33%, $P = 0.04$; Medicaid recipients: 10% vs. 25%, $P = 0.07$). In 2003–2005, HCV-related visits were more than twice as likely to occur among non-white patients vs. white patients (OR = 2.49; 95% CI: 1.60–3.86) and patients on Medicaid vs. non-Medicaid (3.49; 1.79–6.80). Our results show that HCV-associated ambulatory care visits are increasing, with a greater proportion of visits occurring among non-white patients and Medicaid recipients.

Keywords Hepatitis C · Health disparities · Ambulatory care · Health services

J. I. Tsui (✉)
Division of General Internal Medicine, Department of Medicine, Boston University School of Medicine/Boston Medical Center, 801 Massachusetts Ave, 2nd floor, Boston, MA 02118, USA
e-mail: judith.tsui@bmc.org

J. Maselli
Department of Biostatistics, University of California, San Francisco, USA

R. Gonzales
Division of General Internal Medicine, Department of Medicine, University of California, San Francisco, CA, USA

Introduction

More than 3 million Americans are now estimated to be infected with the hepatitis C virus (HCV) [1]. The incidence of HCV peaked in the 1980s (prior to blood product screening) and is now relatively low except among certain high-risk groups such as injection drug users [2, 3] As a consequence, most affected individuals are 40–50 years of age and have likely been infected for decades. Over time, these individuals are at risk for developing cirrhosis and liver cancer, which has prompted concern about a potential surge in HCV-associated morbidity and mortality in the coming years [4]. Providing ambulatory care for HCV-infected individuals so that they can be evaluated for treatment should be of paramount importance.

To date, there is little information on the patterns of ambulatory health care usage among HCV-infected persons. The prevalence of HCV is disproportionately higher among minorities, illicit drug users, and individuals of low socioeconomic status [1]. These are vulnerable populations that may face numerous barriers to adequate health care. Although HCV-related healthcare utilization has been reported to be increasing overall [5, 6], it is unclear whether these vulnerable groups are receiving care for their HCV. This study was conducted to analyze whether the national pattern of HCV-related ambulatory care visits

differed by age, gender, race, and insurance status over time, and to examine the proportion of visits that involved a prescription for anti-HCV therapy.

Methods

Data Sources and Study Design

Data from the National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Ambulatory Medical Care Survey-Outpatient (NHAMCS-OPD) for years 1997–2005 were utilized. These surveys, conducted annually by the National Center for Health Statistics and the Centers for Disease Control, capture nationally representative samples of visits to ambulatory clinics (NAMCS) and hospital-based clinics (NHAMCS-OPD). A detailed description of NAMCS and NHAMCS methodology is available through the National Center for Health Statistics [7]. In brief, trained interviewers provide materials and instruction to physicians, who then record information on patient visits during the reporting period. This study included information only on adult visits (age 18 years or older). Because of relatively small numbers of HCV-associated visits, we combined annual survey data in 3-year intervals (1997–1999, 2000–2002, and 2003–2005). Analysis of this publicly available data set was exempted from institutional review board review by the University of California, San Francisco.

Variables

An HCV-related visit was defined as one in which any of the three principal diagnosis fields contained the following International Classification of Diseases, Ninth Revision (ICD-9 codes): 070.41, 070.44, 070.51, 070.54, 070.70, 070.71, or V02.62. Demographic covariates that were examined included age, sex, race (white vs. non-white), and insurance status (private, Medicare, Medicaid, or other). Additional covariates examined were whether or not the visit was conducted with the patient's primary care provider (as determined by the provider), whether the visit also contained a diagnosis code for complications from HCV defined as cirrhosis, ascites, esophageal varices or hepatocellular cancer (ICD-9 codes: 571.5, 571.6, 155.0, 789.5, 456.1, 456.2, or 571.2), and whether the visit involved prescription of anti-HCV medications (standard and pegylated interferon alpha-2a and 2b, and/or ribavirin).

Analyses

All analyses took into account the complex survey design using weights, strata, and primary sampling unit design variables when calculating estimates. Comparisons of the proportion of visits representing certain patient demographics and prescription of anti-HCV therapy for each 3-year interval were compared using a Chi-square test. Multivariable logistic regression was performed to examine whether the likelihood of a visit being HCV-related (versus unrelated to HCV) was independently associated with race and Medicaid status, and whether there was an interaction between time and those covariates (i.e., whether the effects of race and Medicaid status varied over time). A P-value threshold of <0.05 was used for all statistical testing, including tests for interaction. All analyses were conducted using SAS version 9.1.3 (SAS Institute, Cary, NC) and SUDAAN, version 9.0.3 (RTI International, Research Triangle Park, NC).

Results

Using this data source, it is estimated that national HCV-related ambulatory visits increased from 3,583,585 (95% CI: 2,305,477 to 4,861,693) during 1997–1999 to 8,027,166 (3,714,378 to 12,339,954) during 2003–2005 (Fig. 1). These weighted estimates were based on an absolute number of 276 visits, 468 visits, and 583 visits for the years 1997–1999, 2000–2002, and 2003–2005, respectively (estimates are not stable for fewer than 30 visits) [7]. HCV-related visits constituted a larger percentage of total ambulatory visits over time: they comprised 0.17% of total visits in 1997–1999, 0.26% in 2000–2002, and 0.33% in 2003–2005 (Chi-square P-value $= 0.04$; trend P-value $= 0.1$).

Over the time period examined, the proportion of non-whites and Medicaid recipients presenting for HCV-related visits approximately doubled (Table 1). A relatively small number of visits involved a prescription for anti-HCV medications (<10%), and the proportion did not appear to be changing over time. Likewise, a minority of HCV-related visits also involved complications such as cirrhosis

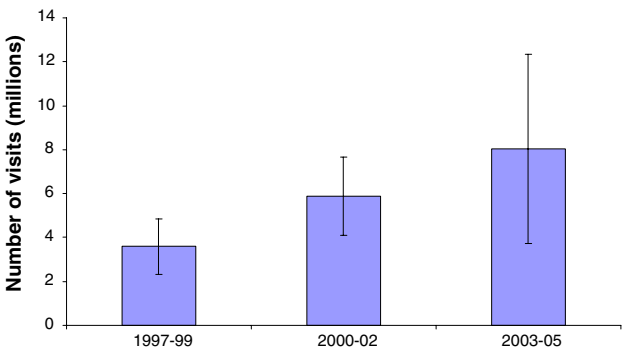


Fig. 1 Estimated number of HCV-related ambulatory visits in the U.S.

Table 1 Characteristics of HCV-related ambulatory visits (NAMCS/NHAMCS-OPD combined)

	1997–1999 $N = 3,583,585$ *N* (%)	2000–2002 $N = 5,875,678$ *N* (%)	2003–2005 $N = 8,027,166$ *N* (%)	*P*-value
Gender				
Female	1,594,071 (44)	2,020,134 (34)	2,702,298 (34)	0.46
Male	1,989,514 (56)	3,855,544 (66)	5,324,868 (66)	
Age				
18–39	841,005 (23)	1,210,659 (21)	1,184,421 (15)	0.83
40–59	2,302,117 (64)	3,775,097 (64)	5,459,966 (68)	
≥60	440,463 (12)	889,922 (15)	1,382,779 (17)	
Race				
White	3,015,711 (84)	4,708,309 (80)	5,339,075 (67)	0.04
Non-white	567,874 (16)	1,167,369 (20)	2,688,091 (33)	
Insurance				
Private	2,169,151 (60)	3,029,322 (52)	3,993,254 (50)	0.21
Medicare	655,850 (18)	773,304 (13)	911,405 (11)	
Medicaid	342,788 (10)	1,048,284 (18)	2,015,782 (25)	
Other[a]	415,796 (12)	1,024,768 (17)	1,106,725 (14)	
Medicaid (vs. non-Medicaid)	342,788 (10)	1,048,284 (18)	2,015,782 (25)	0.07
Diagnosis of HCV complications[b]	199,785 (6)[c]	314,129 (5)	508,510 (6)	0.92
Prescription of anti-HCV meds	247,138 (7)[c]	565,892 (9.6)	503,428 (6.3)	0.76
Primary care provider visit[d]	1,502,684 (45)	2,400,056 (43)	3,466,357 (45)	0.97

[a] Includes self-pay/charity, workers comp, other, unknown and missing

[b] Cirrhosis, ascites, esophageal varices or hepatocellular carcinoma

[c] Less than 30 visits, estimate not stable

[d] 105 absolute visits missing this information = 929,107 weighted visits

and liver cancer, and there was no appreciable change over time. Almost half of all HCV-related visits occurred with the patient's primary care provider**.**

Results from the logistic regression, adjusting for age and sex, demonstrated that HCV-related visits were more likely to occur among non-whites and recipients of Medicaid over time (time–race interaction *P*-value = 0.02; time–Medicaid interaction = 0.04). In the most recent years (2003–2005), HCV-related visits were more than twice as likely to occur among non-white patients, and more than three times as likely to occur among patients on Medicaid (Table 2).

Table 2 Relative odds for visit being HCV-related (vs. non-HCV) associated with race and Medicaid status*

Covariate	1997–1999	2000–2002	2003–2005
White	Ref	Ref	Ref
Non-white	1.04 (0.58, 1.87)	1.43 (0.87, 2.34)	2.49 (1.60, 3.86)
Non-Medicaid	Ref	Ref	Ref
Medicaid	1.49 (0.80, 2.80)	3.54 (2.44, 5.14)	3.49 (1.79, 6.80)

* Results from the logistic regression were adjusted for age and sex

Conclusions

This study suggests that the number of HCV-related ambulatory care visits in the U.S. is rising, with an increasing percentage of visits occurring among non-white patients and recipients of Medicaid. Additionally, we observed that only a small percentage (<10%) of ambulatory HCV-related visits involved anti-HCV treatment, and that the proportion of visits involving treatment did not increase between 1997 and 2005.

There are some potential explanations for our finding that an increasing proportion of HCV-related ambulatory visits occurred among non-whites and Medicaid recipients. First, this may reflect the positive efforts to screen and bring to care individuals who are infected with HCV. National guidelines for HCV screening do not target any particular race or socioeconomic group [8], however, the prevalence of HCV is substantially higher among minorities and individuals of low socioeconomic status. Increasing proportions may reflect greater numbers of individuals from these vulnerable groups who become aware of their diagnosis and are able to access health care. As the incidence of acute HCV is extremely low (0.3 per

100,000) [3], it seems unlikely that a differential rate of new infections among non-whites and Medicaid recipients can fully explain our findings. Another possibility is that a differential rate of HCV-associated complications among whites and non-whites compels individuals to seek out care, as some prior research has suggested disparities between whites and blacks in liver disease outcomes [9]; however, the relatively small number of visits involving HCV-associated complications observed appears to argue against this being a major driving force behind our findings.

Regardless of cause, our findings have important public health implications. The findings support prior research showing an increasing contribution of HCV to national healthcare expenditures [5, 6], but add to the literature by showing that Medicaid is increasingly shouldering the costs of ambulatory healthcare for HCV. Politicians and policy-makers should be aware of the potential for increased Medicaid ambulatory care costs due to HCV. Further, if HCV-related complications such as cirrhosis and liver cancer (and costly treatments like liver transplant) increase in the future, as some researchers have predicted [4], there may be a substantial burden to the public healthcare system. It is nonetheless interesting to note that in this study the proportion of HCV-related ambulatory visits that were associated with complications did not increase over the study period. Models predicting HCV-associated burden have projected increases in HCV-related morbidity and mortality that peak around 2015 [10], so this study may have been conducted too early to detect upward trends in complications.

This study also found that only a small percentage of HCV-related ambulatory care visits (<10%) involved prescription of anti-HCV medications by the provider. Because the data contain no patient-specific information on treatment history, candidacy, and preferences, we cannot determine to what extent this proportion falls short of treatment guidelines. However, the seemingly low percentage of observed visits that involved treatment appears congruent with prior research showing substantial contraindications and barriers to treatment for many HCV-positive patients [11, 12]. In this study, almost half of the HCV-related visits occurred with a self-identified "primary care provider." If a significant proportion of HCV care is taking place in the offices of non-specialist primary care providers, training non-specialist providers to treat HCV may be one strategy to provide greater access to treatment for patients.

This study has several important limitations. Because the study is based on a sample of visits, rather than individuals, observations are restricted to the level of health care utilization. There was potential for misclassification, in particular not identifying all visits that involved HCV-related care. Because the survey only allowed for three diagnosis codes, patients with HCV who had multiple co-morbidities may have not have had their visit coded to reflect their HCV care. Finally, we had relatively small absolute numbers of HCV-related visits and even fewer visits that involved treatment; therefore we were unable to examine patterns of treatment among subgroups, such as non-whites and Medicaid recipients.

In summary, using a nationally representative survey, we found that HCV-related ambulatory care visits are increasing, and that more visits are occurring among non-white and Medicaid patients over time. Since the current incidence of HCV is low, this may reflect improved efforts to provide care for poor, non-white individuals who are chronically infected with HCV. Policy-makers should be aware that HCV-related ambulatory care visits are increasingly paid through Medicaid insurance, which may place a growing burden on the public health care system in the future.

Acknowledgments Dr. Tsui's efforts were supported by grant number KL2RR024130 from the NCRR, a component of NIH. The authors do not report any conflicts of interest.

References

1. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann Intern Med*. 2006;144(10):705–714.
2. Alter MJ. Epidemiology of hepatitis C. *Hepatology*. 1997;26 Suppl 1(3):62S–65S. doi:10.1002/hep.510260711.
3. Wasley A, Grytdal S, Gallagher K. Surveillance for acute viral hepatitis—United States, 2006. *MMWR Surveill Summ*. 2008;57(2):1–24.
4. Armstrong GL, Alter MJ, McQuillan GM, Margolis HS. The past incidence of hepatitis C virus infection: implications for the future burden of chronic liver disease in the United States. *Hepatology*. 2000;31(3):777–782. doi:10.1002/hep.510310332.
5. Grant WC, Jhaveri RR, McHutchison JG, Schulman KA, Kauf TL. Trends in health care resource use for hepatitis C virus infection in the United States. *Hepatology*. 2005;42(6):1406–1413. doi:10.1002/hep.20941.
6. Kim WR. The burden of hepatitis C in the United States. *Hepatology*. 2002;36 Suppl 1(5):S30–S34. doi:10.1053/jhep.2002.36791.
7. http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm.
8. Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. Centers for Disease Control and Prevention. *MMWR Recomm Rep*. 1998;47(RR-19):1–39.
9. Nguyen GC, Thuluvath PJ. Racial disparity in liver disease: biological, cultural, or socioeconomic factors. *Hepatology*. 2008;47(3):1058–1066. doi:10.1002/hep.22223.

10. Wong JB, McQuillan GM, McHutchison JG, Poynard T. Estimating future hepatitis C morbidity, mortality, and costs in the United States. *Am J Public Health*. 2000;90(10):1562–1569.
11. Mehta SH, Thomas DL, Sulkowski MS, Safaein M, Vlahov D, Strathdee SA. A framework for understanding factors that affect access and utilization of treatment for hepatitis C virus infection among HCV-mono-infected and HIV/HCV-co-infected injection drug users. *AIDS*. 2005;19(Suppl 3):S179–S189. doi:10.1097/01.aids.0000192088.72055.90.
12. Bini EJ, Brau N, Currie S, et al. Prospective multicenter study of eligibility for antiviral therapy among 4,084 U.S. veterans with chronic hepatitis C virus infection. *Am J Gastroenterol*. 2005; 100(8):1772–1779. doi:10.1111/j.1572-0241.2005.41860.x.

Hepatitis C Seropositivity and Kidney Function Decline Among Women With HIV: Data From the Women's Interagency HIV Study

Judith Tsui, MD, MPH,[1] Eric Vittinghoff, PhD,[2] Kathryn Anastos, MD,[3] Michael Augenbraun, MD,[4] Mary Young, MD,[5] Marek Nowicki, PhD,[6] Mardge H. Cohen, MD,[7] Marion G. Peters, MD,[2] Elizabeth T. Golub, PhD,[8] and Lynda Szczech, MD, MSCE[9]

Background: How coinfection with hepatitis C virus (HCV) impacts on the trajectory of kidney function in human immunodeficiency virus (HIV)-infected patients is unclear. This study examined the effect of HCV infection on kidney function over time in women infected with HIV.

Study Design: Retrospective observational cohort.

Setting & Participants: Study sample included participants from the Women's Interagency HIV Study who were HIV infected and had undergone HCV antibody testing and serum creatinine measurement at baseline.

Predictor: HCV seropositivity.

Outcomes & Measurement: Estimated glomerular filtration rate (eGFR) calculated from semi-annual serum creatinine measurements using the 4-variable Modification of Diet in Renal Diseases (MDRD) Study equation. Linear mixed models were used to evaluate the independent effect of HCV seropositivity on eGFR over time, adjusting for demographic factors, comorbid conditions, illicit drug use, measures of HIV disease status, use of medications, and interactions with baseline low eGFR ($<$60 mL/min/1.73 m^2).

Results: Of 2,684 HIV-infected women, 952 (35%) were found to be HCV seropositive. In 180 women with chronic kidney disease (CKD) at baseline (eGFR $<$ 60 mL/min/1.73 m^2), HCV seropositivity was independently associated with a fully adjusted net decrease in eGFR of approximately 5% per year (95% confidence interval, 3.2 to 7.2) relative to women who were seronegative. In contrast, HCV infection was not independently associated with a decrease in eGFR in women without low eGFR at baseline ($P < 0.001$ for interaction).

Limitations: The MDRD Study equation has not been validated as a measure of GFR in persons with HIV or HCV infection. Proteinuria was not included in the study analysis. Because the study is observational, effects of residual confounding cannot be excluded.

Conclusions: In HIV-infected women with CKD, coinfection with HCV is associated with a modest, but statistically significant, decrease in eGFR over time. More careful monitoring of kidney function may be warranted for HIV-infected patients with CKD who are also coinfected with HCV.

Am J Kidney Dis 54:43-50.

INDEX WORDS: Hepatitis C virus; human immunodeficiency virus (HIV); kidney diseases; women.

Infection with chronic hepatitis C virus (HCV) has been associated with various types of glomerulonephritis (in particular, membranoproliferative glomerulonephritis) in human immunodeficiency virus (HIV)-uninfected populations.[1] These diseases are difficult to treat and often result in poor outcomes.[2] Approximately 15% to 30% of HIV-infected individuals are also infected with HCV.[3] Treatment for HCV infection in HIV-infected individuals is problematic because of treatment toxicities and poor response rates.[4] As a result, patients coinfected with HIV and HCV may be at risk of HCV-related kidney disease.

Although research is limited, it appears that coinfection with HCV in HIV-infected populations may confer additional risk for adverse kidney-related outcomes. In the setting of HIV infection, HCV infection has been associated with proteinuria[5] and risk of developing acute renal failure,[6] as well as end-stage renal disease

From the [1]Boston University School of Medicine, Boston, MA; [2]University of California, San Francisco, CA; [3]Montefiore Medical Center and Albert Einstein College of Medicine, Bronx; [4]SUNY Downstate, Brooklyn, NY; [5]Georgetown University, Washington, DC; [6]University of Southern California, Los Angeles, CA; [7]CORE Center, Cook County Bureau of Health Services and Rush University, Chicago, IL; [8]Johns Hopkins University, Baltimore, MD; and [9]Duke University, Raleigh-Durham, NC.

Received November 27, 2008. Accepted in revised form January 14, 2009. Originally published online as doi: 10.1053/j.ajkd.2009.02.009 on April 27, 2009.

Address correspondence to Judith Tsui, MD, MPH, Boston University School of Medicine, 801 Massachusetts Ave, 2nd floor, Boston, MA 02118. E-mail: judith.tsui@bmc.edu

0272-6386/09/5401-0009$36.00/0
doi:10.1053/j.ajkd.2009.02.009

requiring renal replacement therapy.[7] However, the exact impact of HCV infection on kidney function trajectories over time in HIV-infected patients has not been fully characterized. One prior study of HIV-infected women found that creatinine clearance tended to be lower in women coinfected with HCV. However, results were not statistically significant, perhaps because of a relatively short follow-up.[8] Precisely how HCV infection impacts on the rate of kidney function decrease over time is important to clinicians and policymakers to anticipate the burden of chronic kidney disease (CKD) in HIV-infected patients.

The purpose of this study was to examine associations between HCV infection and kidney function over time, adjusting for potential confounders. HCV seropositivity was hypothesized to be independently associated with a greater decrease in kidney function over time in HIV-infected women.

METHODS

Study Participants

Women in this study were participants in the Women's Interagency HIV Study (WIHS), a multicenter prospective cohort study of the natural history, including treatment, of HIV infection. Full details of recruitment and baseline cohort characteristics have been described previously.[9,10] The WIHS enrolled women who were either infected with HIV (Western blot confirmed) or at risk of HIV infection between October 1994 and November 1995 and again between October 2001 and September 2002 from 6 clinical consortia in the United States: Chicago, IL; Los Angeles, CA; New York City (Bronx and Brooklyn), NY; San Francisco Bay Area, CA; and Washington, DC. This analysis included HIV-infected WIHS participants who had baseline HCV antibody screening test results and serum creatinine measurement. Participants were evaluated every 6 months by means of physical examination and questionnaires: data from follow-up visits through September 30, 2006, were included in the analysis. Informed consent was obtained from all participants in accordance with the US Department of Health and Human Services guidelines and the institutional review boards of participating institutions.

Study Variables

The outcome of interest was estimated glomerular filtration rate (eGFR), calculated using the 4-variable Modification of Diet in Renal Diseases (MDRD) Study equation (non–isotope dilution mass spectrometry traceable).[11,12] Although this equation was not developed in cohorts with HIV infection, it is commonly used in clinical practice and its use has been recommended in CKD screening guidelines for HIV-infected patients.[13] eGFR was used as a continuous variable and also dichotomized at a threshold of less than 60 mL/min/1.73 m^2 to define participants with baseline CKD based on low eGFR.[14] Because the distribution of eGFR was skewed and the MDRD Study equation is less accurate at greater values, the outcome was transformed by using the natural logarithmic transformation (logGFR). This normalized distribution and also served to downweight changes in eGFR that occurred in the lower versus upper ranges, which in effect "deemphasized" changes in the upper ranges of eGFR, which are less informative. With the outcome natural log transformed, regression coefficient estimates multiplied by 100 are approximately interpretable as percentage of change in average value of the outcome per unit increase in the predictor.[15]

The predictor of interest was baseline HCV serostatus, which was determined by using HCV antibody testing (Ortho-Clinical Diagnostic, Raritan, NJ). Demographic covariates used in the analysis were age, race (African American versus non–African American), income (annual income $\leq$ versus $\geq$\$12,000), and education (high school nongraduate versus graduate). Clinical (not HIV related) covariates included self-reported diagnosis of hypertension or diabetes, systolic and diastolic blood pressure, presence of hepatitis B surface antigen (HBsAg), liver enzyme levels (alanine and aspartate aminotransferase), recent (previous 6 months) illicit drug use, and injection drug use. HIV-related variables included CD4 cell count (cells/μL, analyzed in units of 100), log-transformed HIV viral load, diagnosis of acquired immunodeficiency syndrome (AIDS), and use of highly active antiretroviral therapy (HAART). Use of angiotensin-converting enzyme (ACE) inhibitors and potentially renal-toxic medications were evaluated, including adefovir, cidofovir, tenofovir, foscarnet, indinavir, acyclovir, gancyclovir, sulfamethoxazole/trimethoprim, amphotericin B, and pentamidine. Information about ACE-inhibitor use was based on an open-ended question to participants asking them to describe other non–HIV-related medications and review of pill bottles when patients brought them to study visits (as they were encouraged to do at later visits). Data for all variables, including medications, were collected every 6 months, with the exception of HCV antibody and hepatitis B surface antigen (baseline only).

Statistical Analysis

Demographic, clinical, and laboratory parameters at baseline were compared according to HCV serostatus by using t and χ^2 tests as appropriate. Multivariate logistic regression was used to estimate the relative odds of having eGFR less than 60 mL/min/1.73 m^2 at baseline according to HCV serostatus, adjusting for other covariates.

Linear mixed models with participant-specific random intercepts and slopes were used to estimate the relationship between HCV seropositivity and decrease in logGFR. These models take into account the correlation of outcome by subject and allow for differing numbers of observations across participants arising from missed visits and variable patterns of creatinine measurement. Normality of the residuals, as well as the linearity of covariate effects on logGFR, were examined by using graphical methods. To account for underlying secular trends in mean logGFR common to all participants, time trends were modeled by using linear, quadratic, and cubic terms. The additional effect of HCV

seropositivity on decrease in logGFR, net of any underlying trend, was modeled by using the interaction of time since study entry with baseline HCV serostatus; exploratory analyses showed no substantial departure from linearity in this effect. To determine whether the effect of HCV infection on decrease in logGFR was different in the subset of women with low eGFR (<60 mL/min/1.73 m^2) at baseline, we tested for a difference in HCV (and other covariate) effects by baseline eGFR status by including interaction terms for all covariates in a model that included only postbaseline logGFR values. Because the interaction with HCV infection was statistically significant, we subsequently estimated effects of HCV (and all other covariates) on decrease in postbaseline logGFR by using linear mixed models stratified by baseline eGFR less or greater than 60 mL/min/1.73 m^2. We evaluated for the significance for all covariate interactions with low baseline eGFR by using a Wald test to test for the equality of slope coefficients. We also performed sensitivity analysis of the final linear mixed models, substituting an interaction between HCV infection and baseline eGFR as a continuous variable.

To estimate the independent effect of HCV infection, we adjusted for age, race, poverty, diabetes, hypertension, measured blood pressure, HIV-related factors (AIDS, CD4 cell count, and HIV viral load), hepatitis B surface antigen, use of nephrotoxic medications (as defined previously) and ACE inhibitors, and illicit drug use, updating time-dependent covariates as appropriate. In addition, to fully address confounding by other influences on decrease in logGFR, we included interactions of time with race, diabetes, hypertension, illicit drug use, poverty, AIDS, and medications (HAART, renal-toxic medications, and ACE inhibitors). For diabetes, hypertension, and AIDS, the time-dependent interaction term was calculated as time since onset of the condition, whereas for illicit drug and medication use, it was calculated as current duration of use. For women currently free from a given exposure, the corresponding interaction term was set equal to zero. To create summary estimates of individual slopes, we also calculated rates of logGFR decrease for each participant by using fixed and random effects estimated by using the linear mixed model. This method borrows information across participants, efficiently shrinking slope estimates for those with relatively sparse or noisy logGFR values toward the average slope for other participants with similar covariate values.

Stata, version 9.0 (StataCorp, College Station, TX), was used for all analyses. $P = 0.05$ was considered statistically significant.

RESULTS

Of 2,791 HIV-positive women in WIHS, 2,702 (97%) had HCV serological results (Fig 1). Of those 2,702 women, 18 (0.7%) were missing baseline serum creatinine measurement and were excluded from the analysis, leaving a final study population of 2,684 women. Women who were missing HCV serological or baseline serum creatinine results (n = 107; 3.8% of the original 2,791) were slightly older (mean age, 37 ± 8 [SD] versus 35 ± 8 years; $P = 0.008$) and more likely to use injection drugs (47% versus 33%; $P = 0.002$) and less likely to be on HAART at baseline (8% versus 14%; $P = 0.02$). However, there were no significant differences in mean alanine aminotransferase or serum creatinine levels or the proportion with eGFR less than 60 mL/min/1.73 m^2 between women who were and were not excluded for missing data.

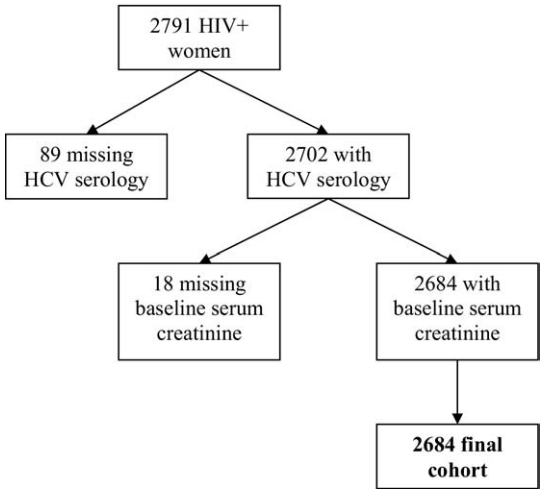


Figure 1. Flow chart of study population selection. Abbreviations: HCV, hepatitis C virus; HIV, human immunodeficiency virus.

Of 2,684 women in the final cohort, 945 (35%) were HCV seropositive. HIV/HCV-coinfected women were more likely to be older, African American, poor, and drug users at baseline and less likely to report being on HAART (Table 1). Of women with eGFR of 60 mL/min/1.73 m^2 or greater, those who were HCV seropositive were more likely to have a greater HIV viral load, have had an AIDS-defining illness, and have hypertension. Diabetes was not significantly more common in women with HCV infection.

At baseline, 180 (6.7%) women in the sample had eGFR less than 60 mL/min/1.73 m^2. At baseline, there was a greater prevalence of CKD in women who were HCV seropositive: 9.8% (93 of 945) versus 5% (87 of 1,739; $P < 0.01$). Before adjustment, women with HCV infection appeared to be twice as likely to have prevalent CKD based on eGFR (unadjusted odds ratio, 2.07; 95% confidence interval, 1.53 to 2.81; $P < 0.001$). After adjustment for age, the relative odds was attenuated to 1.47 (95% confidence

Table 1. Baseline Characteristics of Sample Population by eGFR and HCV Status

	eGFR ≥ 60 mL/min/1.73 m²			eGFR < 60 mL/min/1.73 m²		
	HCV Seronegative (n = 1,652)	HCV Seropositive (n = 852)	*P*	HCV Seronegative (n = 87)	HCV Seropositive (n = 93)	*P*
Age (y)	33 ± 8	39 ± 6	<0.001	39 ± 9	42 ± 7	0.008
African American	908 (55)	513 (60)	0.01	31 (36)	44 (47)	0.1
Non–high school graduate	580 (35)	374 (44)	<0.001	27 (31)	37 (40)	0.2
Income ≤ $12,000/y	868 (54)	595 (72)	<0.001	45 (52)	65 (71)	0.009
Injection drug use	79 (5)	707 (83)	<0.001	7 (8)	80 (86)	<0.001
Any drug use	424 (26)	437 (51)	<0.001	21 (24)	49 (53)	<0.01
Diabetes diagnosis	64 (4)	42 (5)	0.2	8 (9)	9 (10)	0.9
Hypertension diagnosis	166 (10)	180 (21)	<0.001	29 (33)	36 (39)	0.5
Systolic blood pressure (mm Hg)	114 ± 14	117 ± 17	<0.001	122 ± 21	122 ± 22	0.9
Diastolic blood pressure (mm Hg)	73 ± 10	76 ± 12	<0.001	78 ± 13	80 ± 13	0.2
Hepatitis B virus surface antigen positive	42 (3)	30 (4)	0.2	2 (2)	2 (2)	0.9
Aspartate aminotransferase (IU/L)	31 ± 32	58 ± 69	<0.001	34 ± 24	58 ± 56	<0.001
Alanine aminotransferase (IU/L)	29 ± 35	47 ± 51	<0.001	30 ± 30	41 ± 42	0.03
CD4 cell count (cells/μL)	421 ± 292	417 ± 332	0.8	344 ± 312	362 ± 276	0.7
log HIV viral load (copies/mL)	3.9 ± 1.2	4.2 ± 1.1	<0.001	4.5 ± 1.0	4.2 ± 1.1	0.09
AIDS	325 (20)	295 (35)	<0.001	28 (32)	38 (41)	0.2
Highly active antiretroviral therapy	330 (20)	51 (6)	<0.001	6 (7)	2 (2)	<0.01
Tenofovir use	21 (1)	3 (0.4)	0.03	0 (0)	0 (0)	—
Foscarnet use	2 (0.1)	0 (0)	0.3	0 (0)	0 (0)	—
Indinavir use	17 (1)	6 (0.7)	0.4	1 (1)	2 (2)	0.6
Acyclovir use	137 (8)	76 (9)	0.6	12 (14)	5 (5)	0.06
Gancyclovir use	3 (0.2)	1 (0.1)	0.7	0 (0)	0 (0)	—
Sulfamethoxazole/trimethoprim use	528 (32)	375 (44)	<0.001	41 (48)	54 (59)	0.1
Pentamidine use (intravenous)	12 (0.7)	8 (0.9)	0.6	2 (2)	0 (0)	0.1
Angiotensin-converting enzyme inhibitor use*	14 (0.7)	3 (0.4)	0.3	1 (1)	2 (3)	0.5

Note: Values expressed as mean ± SD or number (percent). Numbers and percentages may not sum perfectly because of missing data and rounding.

Abbreviations: AIDS, acquired immunodeficiency syndrome; eGFR, estimated glomerular filtration rate; HCV, hepatitis C virus; HIV, human immunodeficiency virus.

*Results based on visit 5 data because of no reports of angiotensin-converting enzyme inhibitor use at baseline visit.

interval, 1.07 to 2.01; $P = 0.02$), and after full adjustment for all covariates (age, African American ethnicity, education, low income, diabetes, hypertension, AIDS, CD4 cell count, log HIV viral load, HAART, use of renal-toxic medications, injection drug use, and any illegal drug use), the estimate was attenuated further and no longer significant (odds ratio, 1.35; 95% confidence interval, 0.93 to 1.97; $P = 0.1$).

Median follow-up was 4.8 years (first and third quartiles, 3.5 and 11 years) for women without CKD at baseline (based on eGFR) and 4.5 years (first and third quartiles, 1 and 11) for women with CKD. Median numbers of creatinine measurements were 9 (first and third quartiles, 4 and 16) for women without baseline CKD and 6 (first and third quartiles, 2 and 14) for women with CKD. There were no missing follow-up creatinine data for 2,429 (91%) women in the study, 100 (3%) were missing only 1 measurement, and 155 (4%) were missing 2 or more measurements. Linear mixed models allowed for differing numbers of observations across participants arising from missed visits.

Based on calculation of individual slopes from linear mixed models, the majority of HIV-infected women had either improvement or no change or only mildly decreased eGFR over time regardless of HCV status (Fig 2). However, women who were also HCV seropositive were more likely to experience a decrease in eGFR over time and greater rates of decrease.

In combined data, we found that the effect of HCV on net decrease in eGFR differed significantly by baseline eGFR status. Therefore, linear mixed-model analyses were stratified by eGFR less than 60 mL/min/1.73 m^2. In women with CKD at baseline, HCV seropositivity was statis-

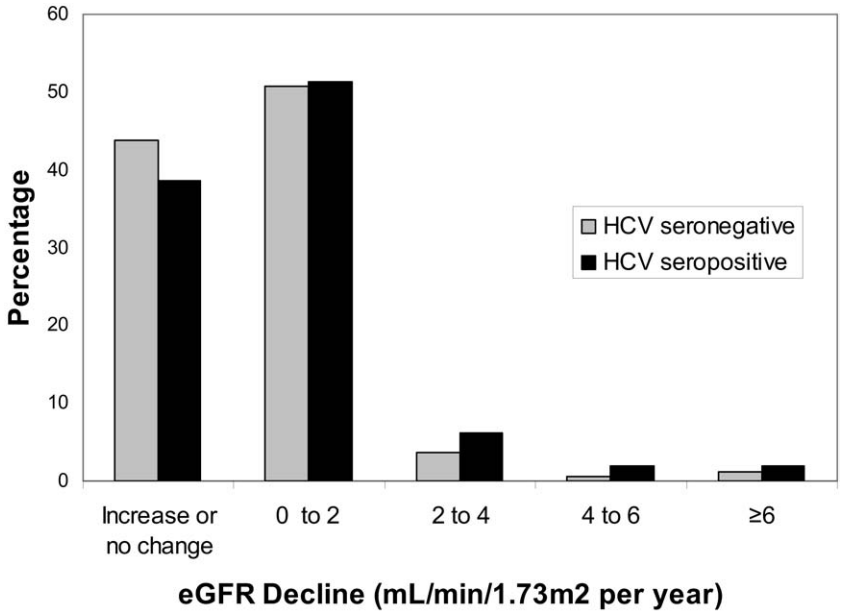


Figure 2. Distribution of rates of estimated glomerular filtration rate (eGFR) decrease by hepatitis C virus (HCV) status (based on estimated individual slopes).

tically significantly associated with a net decrease in eGFR of 5.6% per year after adjustment for other covariates (Table 2). This effect was greater than the effect observed for hypertension and slightly less than the effect for diabetes. In contrast, for women with baseline eGFR of 60 mL/min/1.73 m^2 or greater, HCV infection did not appear to have a significant effect on change in eGFR over time. Results from the sensitivity analysis using an interaction term for HCV and eGFR as a continuous variable (as opposed to dichotomous) also were significant.

DISCUSSION

In this study of HIV-infected women, HCV seropositivity was associated with a slightly lower eGFR over time in women who had eGFR less than 60 mL/min/1.73 m^2 at baseline. In contrast, it did not appear to be associated with a lower eGFR over time in women with baseline eGFR of 60 mL/min/1.73 m^2 or greater. The association between HCV seropositivity and decrease in renal function was statistically significant even after adjusting for demographic factors, illicit drug use, diabetes, hypertension, parameters of HIV disease, and medication use (HAART, nephrotoxic medications, and ACE inhibitors). This is the first study to our knowledge to find an association between HCV seropositivity and longitudinal eGFR in HIV-infected women with CKD.

There are several possible explanations for the association between HCV infection and renal function decrease. Renal function decrease could be caused by HCV-induced glomerular disease. Studies support an association between HCV infection and various types of glomerulonephritis (particularly membranoproliferative glomerulonephritis) and cryoglobulinemia.[1,16-20] Alternatively, HCV infection could be accelerating renal disease associated with HIV, diabetes, and hypertension. In non–HIV-infected populations, HCV infection has been associated with a more rapid decrease in renal function in patients with diabetes.[21] Studies have linked HCV infection to atherosclerosis and atherosclerotic diseases in both HIV- and non–HIV-infected populations.[22-25] It is unlikely that the decrease in renal function could be related to hepatorenal syndrome in the setting of HCV-induced cirrhosis: only 1% of participants reported having cirrhosis during later years of the survey (the question was not asked at baseline). Finally, given the observational nature of the study, it is still possible that the findings could be caused by residual confounding.

Our finding that HCV was associated with eGFR decrease in only women with eGFR less than 60 mL/min/1.73 m^2 is surprising, but also consistent with the prior literature. A large study of veteran health care users found that HCV seropositivity was associated with increased risk

Table 2. Longitudinal Differences in eGFR Associated With HCV and Other Covariates: Results of Fully Adjusted Linear Mixed Models

	eGFR ≥ 60 mL/min/1.73 m²			eGFR < 60 mL/min/1.73 m²			
	Change/y (%)	95% Confidence Interval	*P*	Change/y (%)	95% Confidence Interval	*P*	Interaction *P**
HCV seropositive	−0.6	−1.3 to 0.1	0.08	−5.2	−3.2 to −7.2	<0.001	<0.001
Age at cohort entry	0.01	−0.04 to 0.05	0.8	−0.1	−0.3 to 0.01	0.06	0.07
African American	−1	−1.7 to −0.4	0.002	0.1	−2.4 to 2.6	0.9	0.4
Non–high school graduate	0.5	−0.2 to 1.1	0.2	−2.2	−4.9 to 2.6	0.1	0.06
Income < $12,000/y	0.5	−0.3 to 1.2	0.2	−1.1	−1.6 to 3.8	0.4	0.6
Diabetes diagnosis	−1.6	−2.7 to −0.5	0.004	−7.0	−10.4 to −3.5	<0.001	0.004
Hypertension diagnosis	−1	−1.8 to −0.3	0.008	−4.0	−6.3 to −1.7	0.001	0.01
Systolic blood pressure†	0.3	−0.02 to 0.8	0.2	0.1	−1.8 to 2.0	0.9	0.8
Diastolic blood pressure†	−0.5	−1.2 to 0.3	0.2	−0.2	−3.0 to 2.6	0.9	0.9
AIDS	−1	−1.7 to −0.3	0.003	2.4	0.1 to 4.7	0.04	0.004
CD4 cell count‡	−0.2	−0.5 to 0.1	0.2	0.9	−0.4 to 2.2	0.2	0.1
Log HIV viral load	0.5	−0.2 to 1.1	0.2	1.1	−1.6 to 3.8	0.4	0.7
Highly active antiretroviral therapy	1.7	0.9 to 2.5	<0.001	−2.0	−4.9 to 1.0	0.2	0.02
Hepatitis B surface antigen positive	−0.3	−2.4 to 1.9	0.8	−15.1	−30.3 to 0.04	0.05	0.06
Use of renal toxic medications	−2	−2.8 to −1.2	<0.001	−3.6	−6.4 to −0.9	0.009	0.3
Angiotensin-converting enzyme inhibitors	1.1	−2.6 to 4.7	0.6	−0.6	−12.3 to 11.2	0.9	0.8
Injection drug use	0.6	−1.4 to 2.6	0.6	9.1	1.4 to 16.8	0.02	0.4
Any drug use	−0.7	−1.5 to 0.1	0.1	−0.9	−3.7 to 2.0	0.5	0.9

Note: Trend covariates (interactions with time) are shown, with the exception of CD4 count, log HIV viral load, and systolic and diastolic blood pressure, which are treated as simple time-dependent covariates. All associations shown are net of underlying secular trends.

Abbreviations: AIDS, acquired immunodeficiency syndrome; eGFR, estimated glomerular filtration rate; HCV, hepatitis C virus; HIV, human immunodeficiency virus.

*Wald test for equality of slopes.

†Analyzed in units of 10 mm Hg.

‡Analyzed in units of 100 cells/μL.

of developing end-stage renal disease, but was not associated with prevalent CKD (defined as eGFR < 60 mL/min/1.73 m^2).[26] The investigators hypothesized that patients with HCV infection who reach CKD may experience a more rapid decrease to end-stage renal disease (and renal replacement therapy) or death, and therefore fewer numbers are observed to have eGFR in the CKD range at any single point in time. This study appears to support this hypothesis by showing that in HIV-infected women who had low eGFR at baseline, those who were HCV seropositive had significantly greater decreases in eGFR over time compared with those who were seronegative.

A major limitation of this study is the use of the MDRD Study equation to estimate GFR. This equation has not been independently validated as a measure of GFR in persons with HIV or HCV infection. However, the MDRD Study equation is incorporated widely into clinical care and has been recommended in guidelines for screening for CKD in HIV-infected populations.[13] Because our study was based on all HIV-infected women, principal results could be influenced only if the MDRD Study equation were selectively inaccurate in participants with HCV infection, which is possible given the muscle wasting associated with chronic liver disease. A number of studies comparing serum creatinine level with direct GFR measurement in cirrhotic patients have shown that creatinine level may overestimate true creatinine clearance (ie, appear normal in the setting of decreased GFR).[27,28] This should in theory bias our findings in the opposite direction. Regardless, research is needed to determine the accuracy of GFR-estimating equations in the setting of HIV and HCV infection.

An additional limitation is that HCV antibody status was used, rather than HCV RNA testing. However, prior research has shown that the majority of HIV-infected individuals with a positive screening HCV antibody test result will have chronic hepatitis C by means of RNA testing,[29] and seronegative HCV infection is relatively rare.[30] We did not adjust for current use of anti-HCV therapy; however, prior analyses of this cohort have shown that relatively few WIHS participants who tested positive for HCV antibody reported ever receiving treatment for HCV infection.[31] Ascertainment of non–HIV-related medication use (such as ACE inhibitors) likely was incomplete; however, misclassification should be nondifferential with regard to HCV status. We did not include measures of diabetic control, such as blood glucose or hemoglobin A_{1c} levels, in the analysis; however, diabetes was diagnosed in a relatively small percentage of patients. Finally, a major limitation of our analysis is that it did not include proteinuria because data were not routinely collected on our entire sample. Therefore, we cannot make broader inferences about the prevalence and incidence of true CKD.

In summary, this study found in HIV-infected women with CKD (based on eGFR $<$ 60 mL/min/1.73 m^2) that HCV seropositivity was associated with greater decreases in eGFR over time, and this association was independent of comorbidities, substance abuse, and use of renal-toxic medications. More research is needed to confirm these findings and explore potential mechanisms underlying this association. Clinicians should be aware that HIV-infected individuals with CKD may warrant more careful monitoring of their renal function over time if they are coinfected with HCV.

ACKNOWLEDGEMENTS

The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Support: WIHS is funded by the National Institute of Allergy and Infectious Diseases (UO1-AI-35004, UO1-AI-31834, UO1-AI-34994, UO1-A1-34989, UO1-AI-34993, and UO1-AI-42590) and the National Institute of Child Health and Human Development (UO1-HD-32632). The study is cofunded by the National Cancer Institute, National Institute on Drug Abuse, and National Institute on Deafness and Other Communication Disorders. Funding also is provided by the National Center for Research Resources through Clinical and Translational Science Award UL RR024131 to the UCSF Clinical and Translational Science Institute and Grant KL2RR024130 to Dr Tsui.

Financial Disclosure: None.

REFERENCES

1. Meyers CM, Seeff LB, Stehman-Breen CO, Hoofnagle JH: Hepatitis C and renal disease: An update. Am J Kidney Dis 42:631-657, 2003

2. Ramos-Casals M, Trejo O, Garcia-Carrasco M, Font J: Therapeutic management of extrahepatic manifestations in patients with chronic hepatitis C virus infection. Rheumatology (Oxford) 42:818-828, 2003

3. Sulkowski MS, Thomas DL: Hepatitis C in the HIV-infected person. Ann Intern Med 138:197-207, 2003

4. Sulkowski MS, Benhamou Y: Therapeutic issues in HIV/HCV-coinfected patients. J Viral Hepat 14:371-386, 2007

5. Szczech LA, Gange SJ, van der Horst C, et al: Predictors of proteinuria and renal failure among women with HIV infection. Kidney Int 61:195-202, 2002

6. Franceschini N, Napravnik S, Eron JJ Jr, Szczech LA, Finn WF: Incidence and etiology of acute renal failure among ambulatory HIV-infected patients. Kidney Int 67: 1526-1531, 2005

7. Atta MG, Gallant JE, Rahman MH, et al: Antiretroviral therapy in the treatment of HIV-associated nephropathy. Nephrol Dial Transplant 21:2809-2813, 2006

8. Gardner LI, Holmberg SD, Williamson JM, et al. Development of proteinuria or elevated serum creatinine and mortality in HIV-infected women. J Acquir Immune Defic Syndr 32:203-209, 2003

9. Bacon MC, von Wyl V, Alden C, et al: The Women's Interagency HIV Study: An observational cohort brings clinical sciences to the bench. Clin Diagn Lab Immunol 12:1013-1019, 2005

10. Barkan SE, Melnick SL, Preston-Martin S, et al: The Women's Interagency HIV Study. WIHS Collaborative Study Group. Epidemiology 9:117-125, 1998

11. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D: A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med 130:461-470, 1999

12. Levey AS, Greene T, Kusek JW, Beck GJ: A simplified equation to predict glomerular filtration rate from serum creatinine. J Am Soc Nephrol 11:155A, 2000 (abstr A0828)

13. Gupta SK, Eustace JA, Winston JA, et al: Guidelines for the management of chronic kidney disease in HIV-infected patients: Recommendations of the HIV Medicine Association of the Infectious Diseases Society of America. Clin Infect Dis 40:1559-1585, 2005

14. Levey AS, Coresh J, Balk E, et al: National Kidney Foundation practice guidelines for chronic kidney disease: Evaluation, classification, and stratification. Ann Intern Med 139:137-147, 2003

15. Vittinghoff E, Glidden D, Shiboski S, McCulloch C: Regression Methods in Biostatistics (ed 1). Springer-Verlag New York, LLC, 2005

16. Gumber SC, Chopra S: Hepatitis C: A multifaceted disease. Review of extrahepatic manifestations. Ann Intern Med 123:615-620, 1995

17. Johnson RJ, Gretch DR, Yamabe H, et al: Membranoproliferative glomerulonephritis associated with hepatitis C virus infection. N Engl J Med 328:465-470, 1993

18. McGuire BM, Julian BA, Bynon JS Jr, et al: Brief communication: Glomerulonephritis in patients with hepatitis C cirrhosis undergoing liver transplantation. Ann Intern Med 144:735-741, 2006

19. Stehman-Breen C, Willson R, Alpers CE, Gretch D, Johnson RJ: Hepatitis C virus-associated glomerulonephritis. Curr Opin Nephrol Hypertens 4:287-294, 1995

20. Yamabe H, Johnson RJ, Gretch DR, et al: Hepatitis C virus infection and membranoproliferative glomerulonephritis in Japan. J Am Soc Nephrol 6:220-223, 1995

21. Crook ED, Penumalee S, Gavini B, Filippova K: Hepatitis C is a predictor of poorer renal survival in diabetic patients. Diabetes Care 28:2187-2191, 2005

22. Freiberg MS, Cheng DM, Kraemer KL, Saitz R, Kuller LH, Samet JH: The association between hepatitis C infection and prevalent cardiovascular disease among HIV-infected individuals. AIDS 21:193-197, 2007

23. Fukui M, Kitagawa Y, Nakamura N, Yoshikawa T: Hepatitis C virus and atherosclerosis in patients with type 2 diabetes. JAMA 289:1245-1246, 2003

24. Ishizaka Y, Ishizaka N, Takahashi E, et al: Association between hepatitis C virus core protein and carotid atherosclerosis. Circ J 67:26-30, 2003

25. Vassalle C, Masini S, Bianchi F, Zucchelli GC: Evidence for association between hepatitis C virus seropositivity and coronary artery disease. Heart 90:565-566, 2004

26. Tsui JI, Vittinghoff E, Shlipak MG, et al: Association of hepatitis C seropositivity with increased risk for developing end-stage renal disease. Arch Intern Med 167:1271-1276, 2007

27. Caregaro L, Menon F, Angeli P, et al: Limitations of serum creatinine level and creatinine clearance as filtration markers in cirrhosis. Arch Intern Med 154:201-205, 1994

28. Skluzacek PA, Szewc RG, Nolan CR III, Riley DJ, Lee S, Pergola PE: Prediction of GFR in liver transplant candidates. Am J Kidney Dis 42:1169-1176, 2003

29. Bonacini M, Lin HJ, Hollinger FB: Effect of coexisting HIV-1 infection on the diagnosis and evaluation of hepatitis C virus. J Acquir Immune Defic Syndr 26:340-344, 2001

30. Chamie G, Bonacini M, Bangsberg DR, et al: Factors associated with seronegative chronic hepatitis C virus infection in HIV infection. Clin Infect Dis 44:577-583, 2007

31. Cohen MH, Grey D, Cook JA, et al: Awareness of hepatitis C infection among women with and at risk for HIV. J Gen Intern Med 22:1689-1694, 2007

Drug and Alcohol Dependence 105 (2009) 160–163

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

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

Short communication

Risk behaviors after hepatitis C virus seroconversion in young injection drug users in San Francisco

Judith I. Tsui[a,*], Eric Vittinghoff[b], Judith A. Hahn[c], Jennifer L. Evans[b], Peter J. Davidson[b], Kimberly Page[b]

[a] Department of Medicine, Boston University School of Medicine, Boston, MA, United States
[b] Department of Epidemiology and Biostatistics, University of California, San Francisco, United States
[c] Department of Medicine, San Francisco General Hospital, University of California, San Francisco, United States

ARTICLE INFO

Article history:
Received 17 March 2009
Received in revised form 18 May 2009
Accepted 26 May 2009
Available online 31 July 2009

Keywords:
Hepatitis C virus
Injection drug use
Screening

ABSTRACT

Background: The rationale for screening populations at risk for hepatitis C virus infection (HCV) includes the possibility of altering risk behaviors that impact disease progression and transmission. This study prospectively examined young injection drug users (IDU) to determine if behaviors changed after they were made aware of HCV seroconversion.
Methods: We estimated the effects of HCV seroconversion coupled with post-test counseling on risk behaviors (alcohol use, non-injection and injection drug use, lending and sharing injecting equipment, and having sex without a condom) and depression symptoms using conditional logistic regression, fitting odds-ratios for immediately after disclosure and 6 and 12 months later, and adjusting for secular effects.
Results: 112 participants met inclusion criteria, i.e. they were documented HCV seronegative at study onset and subsequently seroconverted during the follow-up period, with infection confirmed by HCV RNA testing. HCV seroconversion was independently associated with a decreased likelihood of consuming alcohol (OR = 0.52; 95% CI: 0.27–1.00, $p = 0.05$) and using non-injection drugs (OR = 0.40; 95% CI: 0.20–0.81, $p = 0.01$) immediately after disclosure, however, results were not sustained over time. There were significant ($p < 0.05$) declines in the use of alcohol, injection and non-injection drugs, and sharing equipment associated with time that were independent from the effect of seroconversion.
Conclusions: Making young IDU aware of their HCV seroconversion may have a modest effect on alcohol and non-injection drug use that is not sustained over time.

1. Background

The incidence of hepatitis C virus (HCV) remains high (16–42% per year) among young injection drug users (IDU) (Edlin and Carden, 2006; Hahn et al., 2002). The rationale for screening populations at risk for HCV includes the possibility of altering risk behaviors that impact disease progression and transmission, but limited research exist to support this hypothesis (Chou et al., 2004). Some studies suggest healthier alcohol use and injecting practices among individuals who are aware that they are HCV infected as opposed to those who are unaware (Kwiatkowski et al., 2002; McCusker, 2001; Nalpas et al., 2001; Tsui et al., 2007). However, a study of young IDU that examined drug use behaviors 6 months after disclosure of HCV test results failed to find any improvement (Ompad et al., 2002), and another cross-sectional study failed to find any association between awareness of HCV status and injecting behaviors (Cox et al., 2009). No studies have prospectively followed young IDU who HCV seroconvert to examine drug use and sexual behaviors before and after seroconversion.

This study sought to determine whether becoming HCV seropositive and receiving post-test counseling is associated with changes in drug use and sexual risk behaviors among young IDU. We examined whether awareness of seroconversion was associated with a reduction in alcohol, drug use, and sharing/lending of injecting equipment, and an increase in condom usage. In addition, in order to address the potential negative psychological consequences of being diagnosed with HCV, we also analyzed whether notification of seroconversion was associated with subsequent depressed symptoms.

2. Methods and materials

2.1. Study sample and design

This study used observational data from the UFO study, a longitudinal cohort of young injection drug users (<30 years old) in San Francisco who were followed with quarterly interviews and blood sample collection. Details of its study design and methods have been published previously (Hahn et al., 2002). For this study, we restricted our sample to participants who had a documented HCV seroconversion

* Corresponding author at: 801 Massachusetts Ave, Rm 2078, Boston, MA 02118, United States. Tel.: +1 617 414 6912; fax: +1 617 414 4676.
E-mail address: Judith.tsui@bmc.org (J.I. Tsui).

0376-8716/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.drugalcdep.2009.05.022

followed by disclosure/post-test counseling during the study period. Study participants were recruited from January 2000 to June 2007, followed prospectively until February 20, 2008.

2.2. Study outcomes

Outcome variables included alcohol use, injection drug use, sharing of injecting equipment and lending of syringes, non-injection drug use, and having sex without a condom. In addition, in order to examine the potential negative mental health effects of being told one was HCV positive, we examined symptoms of depression as measured by the 8-item version of the Center for Epidemiological Studies-Depression (CESD) Scale (Melchior et al., 1993; Radloff, 1977). A score of ≥7 was used to define significant depressive symptomatology. All behaviors were assessed for the previous three months except for alcohol and injection drug use which were assessed for the previous month. Non-injection drug use was defined as use of cocaine, crack or marijuana (the most commonly used non-injection drugs). Sharing ancillary injecting equipment included any sharing of cookers, cotton, or water and lending syringes was defined as letting someone else use the participant's used syringe. Alcohol use was defined as any use (abstinence versus non-abstinence). All outcomes were assessed from self-reported data collected at interviewer-administered structured interviews.

2.3. Study predictors

The primary predictor was disclosure of HCV seroconversion, followed by post-test counseling. Quarterly HCV testing included antibodies to HCV (anti-HCV) with enzyme immunoassays (EIA) (HCV EIA 2.0, Abbott Laboratories, Abbott Park, IL, or EIA-3, Ortho Clinical Diagnostics, Raritan NJ), as well as HCV RNA virus using transcription mediated amplification (TMA) technique (dHCV TMA assay component of the Procleix HIV-1/HCV assay, Gen-Probe Inc., San Diego, CA) to detect early HCV infection (Hahn et al., 2002). All screening HCV EIA results were confirmed with HCV RNA testing, and testing was done at study visits by study personnel. Disclosure of HCV seroconversion was documented and participants were provided post-test counseling. Counseling was provided by UFO Study counselors who were trained and certified in HIV test disclosure based on client-centered counseling policies of the California Department of Health Office of AIDS. For HCV testing disclosure, all UFO Study counselors received extensive training on interpretation of HCV test results, HCV natural history, and behaviors that impact disease progression and transmission. Pre- and post-test counseling was based on recommendations from the Centers for Disease Control (CDC, 1998), and information was provided regarding the need for a) preventing further harm to their liver (i.e. avoid alcohol), b) reducing risks for transmitting HCV to others (i.e. no lending/sharing of injecting equipment) and c) medical evaluation for liver disease and possible treatment. HIV prevention was emphasized to prevent co-infection. In addition to counseling to reduce risk of liver disease, HCV transmission, and co-infection with HIV, participants were offered vaccination for hepatitis A and B, and offered partner notification assistance.

Additional predictors, which were selected a priori included: age at baseline, gender, race, education, number of years of injection drug use, homelessness and incarceration within the previous three months.

2.4. Statistical analysis

Baseline characteristics of the sample were assessed using simple tabulations and calculation of means and medians. We estimated the effects of HCV seroconversion and disclosure/counseling on risk behaviors and depression using conditional logistic regression. The rationale for this approach was to avoid confounding by differences between seroconverters and other study participants. In this type of matched case-control analysis, each participant serves as his/her own control, and behaviors before and after seroconversion are compared. The effect of HCV seroconversion on the log-odds of each outcome was modeled by a "jump" at the notification visit, followed by a linear trend across subsequent visits. We then computed the fitted odds-ratio for the effect of seroconversion at 6 and 12 months later. To control for confounding of the effect of HCV seroconversion by other covariates, we adjusted for other time-dependent covariates (use of alcohol and drugs, recent homelessness, and incarceration). We controlled for secular effects (changes over time) by including time since study entry, using a linear spline if fit was improved over a simple linear trend. Fixed subject-specific covariates (age, sex, race, etc.), which represent between- rather than within-subject differences, had no influence on the conditional parameter estimates. Because of the small sample size, we used restrictive model selection criteria, retaining covariates if they had a p-value < 0.15 or their inclusion resulted in a >3% change in the seroconversion effect estimate. We checked for departures from linearity of trend in the seroconversion effect across subsequent post-conversion visits, and for collinearity between predictors. All statistical analyses were conducted using Stata version 10.0 (College Station, TX, USA).

Table 1
Baseline characteristics of the sample of young IDU with HCV seroconversion (n = 112).

	Number (%)[a] or mean (SD)
Age	22 (±3)
Female	38 (34%)
Non-white	24 (21%)
High School Graduate	53 (48%)
HIV Positive	4 (4%)
Primary Injecting Drug	
Heroin	61 (59%)
Speed	29 (28%)
Other	13 (13%)
Non-injection Drug Use within Past 3 Months[b]	90 (87%)
Homelessness within Past 3 Months	73 (71%)
Incarceration within Past 3 Months	33 (32%)
Drank Alcohol Past Month	80 (78%)
Days Drank Past Month	11 (±11)

[a] Numbers and percentages may not sum perfectly due to missing data.
[b] Use of marijuana, crack or other cocaine.

3. Results

From the 1223 young IDU that were screened, 555 individuals were enrolled and 403 were prospectively followed. Of those 403, 112 participants met inclusion criteria and were included in this study, i.e. they were documented HCV seronegative at study onset, subsequently seroconverted during the follow-up period. Participants who were included in the analysis tended to be slightly younger than those who were excluded from the analysis (mean age 22 (±3) versus 23 (±3), p-value < 0.01), otherwise there were no significant differences in any of the other variables examined. Participants included in this study were predominantly Caucasian males who injected heroin (Table 1). Most participants acknowledged using non-injection drugs (87%) and drinking alcohol (78%) at baseline. A substantial percentage of participants reported being recently (prior 3 months) homeless (71%) or incarcerated (32%). The median follow-up was 1.8 years (IQR: 1.2–3.3). Among the 11s participants there were a total of 757 visits during which risk behaviors were assessed: 440 visits occurred prior to seroconversion and 317 occurred afterwards. The median number of follow-up visits pre-seroconversion was 3 (IQR: 2–4), the median number post-seroconversion was 3 (IQR: 2–5).

Table 2
Adjusted relative odds for behaviors/depression associated with awareness of HCV seroconversion in 112 young IDU who seroconverted using conditional logistic regression[a].

	Immediately After Seroconversion[b]			6 months After Seroconversion			12 months After Seroconversion		
	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Past Month Alcohol Use	0.52	0.27–1.00	0.05	0.67	0.36–1.26	0.21	0.85	0.43–1.69	0.65
Past Month Injection Drug Use	0.84	0.35–2.05	0.7	0.85	0.36–1.98	0.71	0.86	0.33–2.20	0.75
Past 3 Month Non-injection Drug Use	0.4	0.20–0.81	0.01	0.48	0.23–1.00	0.05	0.57	0.25–1.32	0.19
Past 3 Month Lending of Syringes	0.80	0.29–2.25	0.68	0.49	0.21–1.16	0.10	0.3	0.08–1.09	0.07
Past 3 Month Sharing of Injecting Equipment	0.61	0.22–1.71	0.35	0.6	0.23–1.58	0.3	0.59	0.15–2.30	0.45
Past 3 Month Sex without Condom	1.65	0.77–3.58	0.2	1.57	0.72–3.40	0.26	1.48	0.63–3.48	0.37
Current Depression	0.76	0.23–2.53	0.65	0.78	0.28–2.16	0.63	0.8	0.19–3.29	0.76

[a] Adjusted for secular trends plus drug use, recent incarceration and homelessness; fixed covariates (age, sex, race, etc.), which represent between- rather than within-subject differences, have no influence in the conditional logistic model.
[b] OR for behavior immediately after seroconversion; model assumes change at seroconversion followed by linear trend.

In our analysis using conditional logistic models, we found that HCV disclosure after seroconversion was independently associated with immediate declines in use of alcohol and non-injection drugs (Table 2). However, the reductions became smaller over time (trend p-value for alcohol = 0.02, for non-injection drug use = 0.13). In contrast, while we found little evidence for an immediate decline in lending syringes due to seroconversion, this was the only behavior that consistently diminished over time in association with seroconversion and approached statistical significance at one year; however, the time trend was not statistically (p-value = 0.24). HCV seroconversion was not associated with changes in injection drug use, sharing of equipment or condom use, nor was it associated with depression. Most behaviors declined over time, independent of the HCV seroconversion effect: we observed statistically significant ($p < 0.05$) declines in the use of alcohol, injection as well as non-injection drugs, and sharing of ancillary injecting equipment, independent of the HCV seroconversion effect (data not shown). Our analysis of the correlation of covariates showed that while time since study entry was moderately strongly correlated with both having seroconverted (0.46) and time since seroconversion (0.62), this did not reach the level of collinearity to prevent us from examining their effects simultaneously. Finally, adjusting HIV seroconversion (of which there were only 3 known cases) did not substantially impact results.

4. Discussion

In this study of young IDU, we found that HCV seroconversion was associated with a decreased likelihood of consuming alcohol and using non-injection drugs immediately after disclosure of results and post-test counseling. However, improvements in behaviors were not sustained at 6 months and 12 months. There was no statistically significant change in injection drug use and injecting behaviors after seroconversion and post-test counseling, though there was a non-significant trend toward decreased lending of syringes. Finally, there was no indication that depression symptoms were increased after becoming aware of their HCV infected status.

The finding that injection drug use and injecting behaviors were not significantly affected by HCV seroconversion and post-counseling is similar to a prior study of young IDU in Baltimore that looked at behaviors 6 months after HCV testing and found that those who had tested positive had no change in injecting behaviors (Ompad et al., 2002). Our study, in contrast, did show mild improvements in reported alcohol use immediately after becoming aware of HCV seroconversion, though these improvements were not sustained at 6 months. These results suggest that screening and providing post-test counseling for HCV in young IDU is insufficient for changing long-term behaviors. Evidence from the HIV prevention literature supports the supposition that testing and education alone are insufficient to change behaviors (Calsyn et al., 1992), and that more targeted behavioral interventions are needed to generate sustained reductions in high-risk behaviors in IDU.

There were limitations to this study. It is important to note that the data were collected as part of a study whose primary goal was to detect and assess rates and correlates HCV seroconversion, and was not designed to study the impact of HCV screening and post-test counseling on behaviors. However, in lieu of a randomized controlled study of HCV screening in IDU which has never been conducted, this type of secondary data analysis can provide some insights. The relatively modest sample size and follow-up time were limitations to the study power. Furthermore, the moderately strong correlations of time since study entry with seroconversion as well as time since seroconversion did reduce power to assess these effects, so that our negative finding must be interpreted with caution. Our study was based on self-reported behaviors, and therefore may reflect socially desirable responses rather than actual behaviors, in particular since study personnel who assessed risk behaviors also disclosed test results and provided counseling.

In summary, this observational study of young IDU found modest improvements in reported alcohol and non-injection drug use immediately after disclosure of HCV seroconversion and receipt of post-test counseling, however, those improvements were not sustained over time. On the other hand, we found no evidence that patients became more depressed after learning that they had newly acquired HCV. While these results do not demonstrate that HCV testing and counseling have a major influence on risk behaviors on individuals who seroconvert, they do lend some evidence against substantial harm. More studies are needed to identify and implement effective interventions to reduce high-risk behaviors in young IDU who become HCV infected.

Role of the funding source

Funding for this study was supported by the U.S. National Institutes for Health (NIH) 2 R01 DA016017-03A1. Additional support was received from: NIH grants - 5 U19 AI40034-13 (Dr. Page, Ms. Evans), K01 DA023365 (Dr. Hahn). Funding for Dr. Tsui was provided through Grant Number KL2RR024130 from the National Center for Research Resources. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The NIH 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.

Contributions

Author Kimberly Page was responsible for the parent (UFO) study design. Authors Judith Tsui, Judith Hahn and Jennifer Evans were responsible for the study design for the secondary analysis for this paper. Authors Judith Tsui and Eric Vittinghoff were responsible for the statistical analysis. All authors (Kimberly Page, Judith Tsui, Eric Vittinghoff, Judith Hahn, Peter Davidson and Jennifer Evans) were responsible for the interpretation of results. Authors Judith Tsui and Eric Vittinghoff wrote the first draft of the manuscript. Authors Kim Shafer, Judith Hahn, Jennifer Evans and Peter Davidson all participated in writing later versions of the manuscript. All authors contributed to and have approved the final manuscript.

Conflicts of interest

All authors (Tsui, Vittinghoff, Hahn, Evans, Davidson and Page) declare that they have no conflicts of interest.

Acknowledgments

The ongoing participation of all the UFO Study participants contributes to the success of this research and to public health practice. We also wish to recognize the important contributions of Dr. Paula Lum to the health and welfare of study participants, as well as significant support from the San Francisco Department of Public Health (SFDPH), including the HIV Prevention Branch, Communicable Disease Branch, Dr. Joshua Bamberger with SFDPH Housing and Urban Health, the Tenderloin AIDS Resource Center (TARC), SOS, and the Homeless Youth Alliance (HIYA).

References

Calsyn, D.A., Saxon, A.J., Freeman Jr., G., Whittaker, S., 1992. Ineffectiveness of AIDS education and HIV antibody testing in reducing high-risk behaviors among injection drug users. Am. J. Public Health 82, 573–575.

CDC, 1998. Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. Centers for Disease Control and Prevention. MMWR Recomm. Rep. 47, 1–39.

Chou, R., Clark, E.C., Helfand, M., 2004. Screening for hepatitis C virus infection: a review of the evidence for the U.S. Preventive Services Task Force. Ann. Intern. Med. 140, 465–479.

Cox, J., Morissette, C., De, P., Tremblay, C., Allard, R., Graves, L., Stephenson, R., Roy, E., 2009. Access to sterile injecting equipment is more important than awareness of HCV status for injection risk behaviors among drug users. Subst. Use Misuse 44, 548–568.

Edlin, B.R., Carden, M.R., 2006. Injection drug users: the overlooked core of the hepatitis C epidemic. Clin. Infect. Dis. 42, 673–676.

Hahn, J.A., Page-Shafer, K., Lum, P.J., Bourgois, P., Stein, E., Evans, J.L., Busch, M.P., Tobler, L.H., Phelps, B., Moss, A.R., 2002. Hepatitis C virus seroconversion among young injection drug users: relationships and risks. J. Infect. Dis. 186, 1558–1564.

Kwiatkowski, C.F., Fortuin Corsi, K., Booth, R.E., 2002. The association between knowledge of hepatitis C virus status and risk behaviors in injection drug users. Addiction 97, 1289–1294.

McCusker, M., 2001. Influence of hepatitis C status on alcohol consumption in opiate users in treatment. Addiction 96, 1007–1014.

Melchior, L.A., Huba, G.H., Brown, V.B., Reback, C.J., 1993. A short depression index for women. Educ. Psychol. Meas. 53, 1117–1125.

Nalpas, B., Martin, S., Fontaine, H., Fabbro-Peray, P., Brechot, C., Pol, S., 2001. Impact of medical recommendations on alcohol consumption in HCV positive patients. J. Hepatol. 35, 312–313.

Ompad, D.C., Fuller, C.M., Vlahov, D., Thomas, D., Strathdee, S.A., 2002. Lack of behavior change after disclosure of hepatitis C virus infection among young injection drug users in Baltimore, Maryland. Clin. Infect. Dis. 35, 783–788.

Radloff, L., 1977. The CES-D scale: a self-report depression scale for research in the general population. Appl. Psychol. Meas. 1, 385–401.

Tsui, J.I., Saitz, R., Cheng, D.M., Nunes, D., Libman, H., Alperen, J.K., Samet, J.H., 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.

Association of Hepatitis C Virus Seropositivity With Inflammatory Markers and Heart Failure in Persons With Coronary Heart Disease: Data From the Heart and Soul Study

JUDITH I. TSUI, MD, MPH,[1] MARY A. WHOOLEY, MD,[2,3,4] ALEXANDER MONTO, MD,[3,5] KAREN SEAL, MD, MPH,[2,3] PHYLLIS C. TIEN, MD,[3,6] AND MICHAEL SHLIPAK, MD, MPH[2,3,4]

Boston, Massachusetts; San Francisco, California

ABSTRACT

Background: How hepatitis C virus (HCV) affects coronary heart disease (CHD) risk factors and outcomes is largely unknown.

Methods and Results: Among a cohort of patients with stable CHD, we examined the association between HCV seropositivity and levels of inflammatory markers (C-reactive protein [CRP], fibrinogen, interleukin-6, and tumor necrosis factor [TNF]-α) and risk for the following outcomes: death, cardiovascular (CV) events, and heart failure events. A total of 84 (8.6%) participants were found to be seropositive for HCV. HCV-seropositive patients were found to have significantly lower adjusted mean levels of CRP (2.6 vs. 4.4; $P < .01$) and fibrinogen (340 vs. 398; $P < .01$), but higher levels of TNF-α (7.1 vs. 4.8; $P < .01$). Age-adjusted rates for HCV seropositive vs. seronegative were as follows: death 93 vs. 42/1,000 p-y ($P < .01$), CV events 62 vs. 40 ($P = .13$), and heart failure 76 vs. 29 ($P < .01$). After adjustment for demographic and clinical factors, HCV remained significantly associated with an increased risk for heart failure events (HR = 2.13; 95% CI: 1.19–3.80).

Conclusions: In this cohort with CHD, HCV seropositive participants had higher rates of death, CV events, and heart failure hospitalizations during follow-up. After adjustment for CV risk factors, HCV seropositivity remained independently associated with risk for heart failure events. (*J Cardiac Fail 2009;15:451–456*)

Key Words: Hepatitis C virus, inflammatory markers, heart failure.

Approximately 4 million Americans are estimated to have been infected with hepatitis C virus (HCV) and the majority of those infected are currently in their 4th and 5th decades of life.[1] As individuals with HCV begin to age, they will inevitably face common comorbidities such as cardiovascular diseases. It is unknown how infection with HCV affects coronary heart disease (CHD) progression and outcomes.

Infectious etiologies have been hypothesized to contribute to the inflammatory cascade leading to atherosclerosis.[2] Some studies have found cross-sectional associations between HCV and cardiomyopathies,[3,4] coronary atherosclerosis,[5–7] carotid artery plaque,[8,9] and increased pulse wave velocity,[10] although not all studies support these findings.[11–13] A recent epidemiologic study showed that HCV-seropositive blood donors had higher rates of

From the [1]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine/Boston Medical Center; [2]Divisions of General Internal Medicine, Department of Medicine, University of California, San Francisco, CA; [3]Department of Medicine, Veterans Affairs Medical Center, San Francisco, CA; [4]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA; [5]Gastroenterology and Hepatology, University of California, San Francisco, CA and [6]Infectious Diseases, Department of Medicine, University of California, San Francisco, CA.

Manuscript received May 13, 2008; revised manuscript received October 22, 2008; revised manuscript accepted December 4, 2008.

Reprint requests: Judith I. Tsui, MD, MPH, General Internal Medicine, Boston Medical Center, 801 Massachusetts Ave, Boston, MA 02118. Tel: (617) 414-5912; Fax: (617) 414-4676. E-mail: judith.tsui@gmail.com

All decisions regarding this manuscript were made by a Guest Editor.

This work was supported by grants from the Department of Veterans Affairs (Epidemiology Merit Review Program); the National Heart, Lung and Blood Institute (R01 HL079235); the Robert Wood Johnson Foundation (Generalist Physician Faculty Scholars Program); the American Federation for Aging Research (Paul Beeson Faculty Scholars in Aging Research Program); and the Ischemia Research and Education Foundation. This publication was also made possible by Grant Number KL2RR024130 from the NCRR, a component of the National Institutes of Health. None of these funding sources had any role in the collection of data, interpretation of results, or preparation of this manuscript. The contents are solely the responsibility of the authors and do not necessarily represent the official view of the organizations listed above. The authors have no conflicts of interest or financial disclosures to report.

Conflict of interest: none.

1071-9164/$ - see front matter

doi:10.1016/j.cardfail.2008.12.003

cardiovascular mortality compared to uninfected donors.[16] Furthermore, HCV infection is recognized to cause chronic immune stimulation, leading to an inflammatory response and cytokine production.[17] These altered cytokine profiles observed in the setting of chronic HCV could potentially lead to adverse cardiovascular outcomes.[18]

Based on these prior findings, we conducted this study to explore the relationship between HCV seropositivity and inflammatory markers and clinical outcomes among individuals with established CHD. We hypothesized that HCV-seropositive participants might be distinguished by different levels of inflammatory markers and increased risk for subsequent CHD events.

Methods

Study Participants

The Heart and Soul study is an ongoing prospective cohort study designed to determine how psychosocial factors influence disease progression in persons with coronary disease. Methods have been described in depth previously.[19,20] In brief, administrative databases were used to identify outpatients with documented CHD at two Department of Veterans Affairs Medical Centers (San Francisco and Palo Alto), 1 university medical center (University of California, San Francisco), and 9 public health clinics in the Community Health Network of San Francisco. Patients were eligible to participate if they had known CHD documented by at least 1 of the following: a history of myocardial infarction (MI), angiographic evidence of ≥50% stenosis in one or more coronary vessels, prior evidence of inducible ischemia by treadmill or nuclear testing, or a history of coronary revascularization. Between September 2000 and December 2002, 1024 participants were enrolled and attended a baseline study appointment that included a medical history interview, a physical examination, and a comprehensive health status questionnaire, as well as blood draw for serum and plasma samples. This study cohort was restricted to 981 participants with HCV antibody test results. HCV serostatus was determined from documentation of a prior positive HCV antibody test in the patient's medical record (n = 27), or if no evidence of prior testing existed, from testing baseline serum samples (n = 964) using a 3rd-generation anti-HCV immunoassay (Johnson & Johnson's Vitros Anti-HCV Assay). Results were reported as positive, negative, or indeterminate as recommended by Centers for Disease Control and Prevention guidelines.[21] Indeterminate results were excluded from the analysis (n = 10).

Measurement of Inflammatory Markers

Participants were instructed to fast for 12 hours (except for medications), not to take aspirin for 1 week, and not to smoke for 5 hours before their study appointment. Venous blood samples were obtained, and plasma and serum samples were stored at −70°C until the time of the assay. Laboratory technicians who assayed the inflammatory markers were blinded to clinical characteristics. We used the Roche Integra high-sensitivity assay to measure C-reactive protein (CRP) in 229 participants and (because of a change in lab protocol) the Beckman Extended Range high-sensitivity CRP assay to measure CRP in the remaining 756 participants. Serum fibrinogen levels were determined by the Clauss assay.[22] We used the R&D Systems (Minneapolis, MN) Quantikine HS IL-6 immunoassay to determine the concentration of interleukin (IL)-6. We used the Human Serum Adipokine Panel B LINCOplex Kit (Linco Research, Inc., St. Charles, MO) to measure tumor necrosis factor (TNF)-α.

Longitudinal Outcome Measures

We examined the following outcomes: all-cause mortality; cardiovascular (CV) events defined as CHD death, MI, or stroke; and heart failure (HF). We conducted annual telephone follow-up interviews with participants (or their proxy) and asked about death or hospitalization for "heart trouble." For any reported event, medical records, electrocardiograms, death certificates, and coroner's reports were retrieved and reviewed by 2 independent and blinded adjudicators. If the adjudicators agreed on the outcome classification, their classification was binding. In the event of disagreement, consultation from a third blinded adjudicator was performed. All-cause mortality was determined by review of death certificates. Nonfatal MI was defined using the American Heart Association diagnostic criteria.[23] A CHD death was defined as a death occurring during the same hospitalization in which an acute MI was documented or a death occurring within 1 hour of the onset of terminal symptoms not explained by other etiologies. Stroke was defined as a new neurologic deficit not known to be secondary to brain trauma, tumor, infection, or other cause. An HF event was based on Framingham criteria and defined as hospitalization for a clinical syndrome involving at least 2 of the following: paroxysmal nocturnal dyspnea, orthopnea, elevated jugular venous pressure, pulmonary rales, third heart sound, and cardiomegaly or pulmonary edema on chest radiography.[24]

Other Covariates

We examined the following characteristics as potential confounding variables: age, gender, race (white vs. non-white), education status (high school vs. non–high school graduate), body mass index (BMI), being physically active, current smoking, regular alcohol use (>4 drinks per week), recent (past year) illicit drug use, diabetes, hypertension, HIV, total cholesterol, high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), triglycerides, systolic and diastolic blood pressure, and use of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors (statins), aspirin, β-blockers, angiotensin-converting enzyme (ACE) inhibitors, or angiotensin II receptor blockers (ARBs). Data on age, race/ethnicity, education, physical activity level, smoking status, alcohol use, illicit drug use, and medical comorbidities were determined by participant self-report. Study personnel recorded use of all medications at baseline study visit. Medication categories were categorized using Epocrates Rx (San Mateo, CA). Serum cholesterol was measured from fasting venous blood samples.

Statistical Analysis

We examined differences in demographic and clinical variables between HCV seropositive and seronegative participants using *t*-tests for continuous variables and chi-squared tests for categorical variables. We compared the adjusted means of inflammatory markers and cholesterol of HCV-seropositive and HCV-seronegative participants, using linear regression to adjust for age, sex, race, smoking, BMI, diabetes, hypertension, HIV, drug use, physical activity, aspirin, statin, β-blocker and ACE inhibitor/ARB use. These covariates were chosen on the basis of literature review and significant associations with HCV status observed in univariate analysis. All inflammatory marker values, except for

fibrinogen, were transformed for analysis using natural logarithm because of highly skewed data, and were back transformed for presentation of results. For cholesterol, we experimented with log transforming to account for more mild skewness in the data, but found that results were similar; we chose therefore to use non-transformed variables in the analyses.

To compare the risk of outcomes among anti–HCV-positive and negative participants, we compared age-adjusted incidence rates using Poisson regression and constructed Cox proportional hazards models for each outcome: all-cause death, CV events, and HF events. For our Cox proportional hazards models, we adjusted for age, sex, race, smoking, BMI, HIV, diabetes, hypertension, physical activity, illicit drug use, statin use, aspirin use, β-blocker use, ACE inhibitor/ARB use, total cholesterol, HDL, log CRP, fibrinogen, log IL-6, and log TNF-α. We adjusted sequentially for demographic variables, clinical variables, and inflammatory markers. To avoid overfitting the model, we used a stepwise backward selection procedure for covariates, retaining all demographic variables in each model. Other covariates were retained in the adjusted models if they had either a $P < .2$ for the outcome, or if their inclusion in the model caused the parameter estimate for HCV to change by more than 5%. Cox models were created for each outcome; additionally, to differentiate whether HCV was associated with new cases of HF vs. HF exacerbations, we excluded participants with preexisting diagnoses of HF and reran models examining HF events. Cox models were checked for violation of the proportional hazards assumption by assessing log-minus-log survival plots for patterns of nonproportionality and performing the Schoenfeld test. All statistical analyses were conducted using Stata 8.2 (Stata Corporation, College Station, TX).

Results

Of the 981 participants with CHD, 84 (8.6%) were seropositive for HCV. HCV-seropositive participants were younger, had lower BMI, and were more likely to be current smokers and to have recently used illicit drugs (Table 1). HCV-seropositive participants were also more likely to be HIV positive, although the proportion was still relatively low (n = 8 or 10%). There were significant differences in the receipt of CHD treatments: HCV-seropositive participants were less likely to be taking statins, aspirin, β-blockers, ACE inhibitors, or ARBs than seronegative participants. There was no difference in the prevalence of diabetes or in measured blood pressure or resting left ventricular ejection fraction between the groups at baseline.

At baseline, patients with HCV had lower levels of fibrinogen and higher levels of TNF-α (Table 2). They also tended to have lower levels of HDL, but this finding was not statistically significant. After adjustment for age and other covariates, including statin use (which was substantially lower in HCV seropositive patients), we found that participants with HCV had significantly lower levels of all lipid measures. There were also differences in levels of inflammatory markers: adjusted mean levels of CRP and fibrinogen were lower, whereas TNF-α levels were significantly higher for HCV seropositive participants.

Data on follow-up outcomes were available for 970 participants (11 lost to follow-up), and the mean follow-up was 4.1 years (range, 0.1–6.1 years). There were 182 deaths (161 HCV−, 21 HCV+), 151 CV events (137 HCV−, 14 HCV+), and 119 HF hospitalizations (103 HCV−, 16 HCV+) in the follow-up period. Age-adjusted incidence rates were higher among HCV-seropositive participants for all outcomes (Fig 1). Specific rates for HCV-seropositive vs. seronegative patients were as follows: for death, 93 vs. 42 ($P < .01$); for CV events 62 vs. 40 ($P = .13$); for HF hospitalizations 76 vs. 29 ($P < .01$).

To assess whether HCV seropositivity was associated with risk for clinical outcomes independent of other risk factors, we performed Cox-proportional hazards models, adjusting for age, clinical CVD risk factors, and inflammatory markers in a sequential fashion. Adjusting for age, sex, and race, we observed that HCV seropositivity was associated with a greater than 2-fold risk for death and HF hospitalizations, as well as an 80% increased risk for CV events (Table 3). After adjusting for other clinical variables, HCV remained associated with a 50% increase in risk of death and CV events, although the associations were no longer significant. The association of HCV with HF, however, remained 2-fold and significant. Further adjustment for

Table 1. Sample Characteristics by Hepatitis C Antibody Status

	HCV Ab Negative (n = 897)	HCV Ab Positive (n = 84)	
	Mean ± SD or n (%)†	Mean ± SD or %†	*P* Value*
Age (y)	67 ± 11	59 ± 11	<.01
Female	161 (18%)	17 (20%)	.61
Non-white	350 (39%)	39 (46%)	.15
High school graduate	783 (88%)	71(85%)	.42
BMI (kg/m^2)	29 ± 5	26 ± 6	.01
Physically active	567 (64%)	49 (58%)	.34
Current smoker	147 (16%)	46 (57%)	<.01
Regular alcohol use	264 (30%)	21 (25%)	.36
Recent illicit drug use	51 (6%)	24 (29%)	<.01
Diabetes mellitus	230 (26%)	23 (27%)	.75
Hypertension	634 (71%)	52 (62%)	.08
HIV	17 (2%)	8 (10%)	<.01
Statin use	605 (67%)	21 (25%)	<.01
Aspirin use	699 (78%)	57 (68%)	.04
β-blocker use	522 (58%)	37 (44%)	.01
ACE inhibitor/ARB use	470 (52%)	30 (36%)	<.01
Systolic BP	133 (±20)	134 (±23)	.52
Diastolic BP	74 (±11)	76 (±11)	.15
Resting LV ejection fraction	0.62 (±0.1)	0.62 (±0.1)	.89
Low platelets (< 130 × 103 μL)	30 (3%)	7 (8%)	.02

ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; BMI, body mass index; BP, blood pressure; HCV, hepatitis C virus; HIV, human immunodeficiency virus; LV, left ventricular; SD, standard deviation.

*Student's *t*-test or chi-square test.

†Numbers/percentages may not sum due to missing data; <1% data missing for all covariates.

Table 2. Mean Levels of Inflammatory Markers and Cholesterol by HCV Serostatus*

	Unadjusted Results			Adjusted Results		
	HCV Seronegative	HCV Seropositive		HCV Seronegative	HCV Seropositive	
	Mean (95%CI)	Mean (95%CI)	*P* value	Mean (95% CI)	Mean (95% CI)	*P* value
Total cholesterol (mg/dL)	178 (175–181)	171 (160–181)	.15	180 (177–183)	153 (144–162)	<.01
LDL (mg/dL)	104 (102 to 107)	103 (95–111)	.8	106 (103–108)	91 (84–99)	<.01
HDL (mg/dL)	46 (45–47)	43 (40–46)	.08	46 (45–47)	43 (40–46)	.05
Triglycerides (mg/dL)	142 (133–151)	125 (106–144)	.12	144 (136–152)	105 (72–130)	.01
CRP (mg/L)	4.6 (4.1–5.2)	4.0 (2.5–5.4)	.41	4.4 (4.1–4.8)	2.6 (1.9–3.5)	<.01
Fibrinogen (mg/dL)	397 (391–403)	343 (324–362)	<.01	398 (391–403)	340 (324–362)	<.01
IL-6 (pg/mL)	3.3 (3.1–3.5)	3.7 (3.1–4.3)	.16	3.2 (3.0–3.3)	3.6 (3.0–4.2)	.13
TNF-α (pg/mL)	4.4 (4.2–4.7)	7.4 (5.6–9.2)	<.01	4.8 (4.6–5.1)	7.1 (5.8–8.7)	<.01

CRP, C-reactive protein; HCV, hepatitis C virus; HDL, high-density lipoprotein; IL, interleukin; LDL, low-density lipoprotein; TNF, tumor necrosis factor.
*Adjusted for age, sex, race, smoking, body mass index, diabetes, hypertension, human immunodeficiency virus, drug use, physical activity, aspirin, statin, β-blocker, and angiotensin-converting enzyme inhibitor/angiotensin receptor blocker use.

inflammatory markers had little effect on point estimates and *P* values of results. After excluding participants with a preexisting diagnosis of HF, the association between HCV and HF hospitalizations remained significant (fully adjusted HR = 2.25; 95% CI: 1.02–4.97; *P* = .04).

Discussion

Among this cohort with CHD, we found HCV seropositivity to be associated with lower lipids, CRP and fibrinogen levels, and higher levels of IL-6 and TNF-α. Despite lower levels of LDL and CRP, HCV-seropositive participants experienced higher rates of death, CV events, and HF hospitalizations over time. After adjusting for risk factors, treatment differences, and inflammatory markers, HCV-seropositive participants still had a statistically significant 2-fold increase in risk of HF hospitalizations and a 50% elevated risk of death and CV events that did not reach statistical significance.

This is the first study to our knowledge to compare clinical outcomes by HCV serostatus in a cohort with CHD. Our finding that HCV is associated with HF hospitalizations is novel and warrants future investigation. A possible explanation for increased HF hospitalizations could be a higher incidence of cardiomyopathies or myocarditis among participants with HCV. Some studies, primarily from Japan, have noted a high prevalence of HCV infection among patients with cardiomyopathies and myocarditis,[3,4] although other studies have not confirmed this association.[14,15] A study of mice transgenic for the HCV core gene found increased development of cardiomyopathy, as evidenced by Doppler echocardiography and histologic changes.[25] Although HCV-seropositive participants in our study had similar baseline resting ejection fractions to seronegative participants, their ventricular function may have worsened over time. Another possibility is that HCV could be associated with progressive vascular changes because HCV seropositivity has been associated with accelerated coronary vasculopathy and decreased survival in heart transplant recipients.[26,27] Also, inflammatory markers IL-6 and TNF-α have been associated with HF severity and outcomes,[28–31] and our study did find higher levels of TNF-α in HCV-seropositive participants, as has been previously reported.[32–35] However, adjustment for inflammatory markers did not attenuate the association between HCV and HF in this study, and therefore they do not appear to mediate the association. Another explanation could be that HCV is a marker for other variables that we did not adequately control for such as adherence, quality of care, or social factors that could influence hospitalization. Finally, side effects of chronic HCV such as fatigue or depression could lead to worsened symptoms of HF leading to hospitalization.

Our finding that HCV-seropositive participants had significantly different levels of lipids from HCV-seronegative participants is consistent with most prior research. Several moderately sized clinical studies have also demonstrated that patients with HCV have lower total cholesterol,[36,37] HDL, and LDL[38] compared with healthy controls, and results have been confirmed in population-based studies

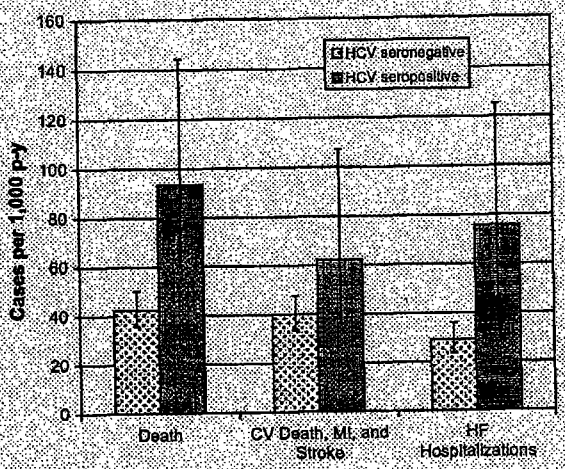


Fig. 1. Age-adjusted incidence of outcomes by hepatitis C virus status.

Table 3. Relative Hazards for Outcomes Associated with HCV Seropositivity

	Death		CV Death, MI, or Stroke		HF Hospitalizations	
	HR (95% CI)	*P* Value	HR (95%)	*P* Value	HR (95%)	*P* Value
Model 1: adjusted for demographic factors*	2.57 (1.61–4.11)	<.01	1.83 (1.04–3.23)	.04	2.80 (1.63–4.83)	<.01
Model 2: adjusted for the above plus significant clinical factors†	1.58 (0.95–2.63)	.08	1.54 (0.83–2.84)	.17	2.13 (1.19–3.80)	.01
Model 3: adjusted for the above plus significant inflammatory markers‡	1.62 (0.95–2.75)	.07	1.74 (0.92–3.32)	.09	2.05 (1.11–3.78)	.02

ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; BMI, body mass index; BMI, body mass index; CRP, C-reactive protein; CV, cardiovascular; HDL, high-density lipoprotein; HF, heart failure; HR, heart rate; HIV, human immunodeficiency virus; IL, interleukin; MI, myocardial infarction; TNF, tumor necrosis factor.

*Age, sex, and race.

†Retained covariates are as follows: 1) for death: smoking, drug abuse, HIV, BMI, diabetes, physical activity, statin, and ACE inhibitor/ARB use; 2) for CV outcomes: smoking, diabetes, hypertension, physical activity, statin use, ACE inhibitor/ARB use, total cholesterol, and HDL; 3) for HF hospitalizations: smoking, BMI, physical activity, diabetes, statin use, aspirin use, ACE inhibitor/ARB use.

‡Retained covariates are as follows: 1) for death: CRP and IL-6 2) for CV outcomes: CRP and IL-6 3) for HF hospitalizations: CRP, IL-6 and TNF-α.

as well.[39] The finding of lower adjusted CRP and fibrinogen levels is also supported by some prior research. One recent study of HIV-infected patients also found that HCV was associated with significantly lower CRP levels.[40] Another study of patients with chronic HCV demonstrated that CRP levels were lower than healthy controls and levels did not rise after treatment with interferon alpha-2b.[41] Because CRP is synthesized by the liver, there is also biological plausibility to this finding. Physicians may want to use caution when interpreting CRP levels among patients with HCV infection, as standard cutoff ranges may not apply to this special population. We found only one relatively small study that examined fibrinogen in persons with and without HCV and it did not find a significant difference.[38]

There are several important limitations of this study. This study used HCV antibody status rather than HCV RNA testing, which would more accurately characterize whether participants had chronic HCV. However, prior research has shown that the great majority (approximately 80%) of individuals with a positive HCV serology will have chronic infection.[1] Furthermore, incorrectly labeling HCV status should have the effect of biasing our results to the null, thus strengthening our positive findings. We did not have information HCV viral genotype or histologic data to determine stage of liver disease. We also did not have data on HCV treatment which is known to effect cholesterol levels and inflammatory markers; however, given that this was a population of predominantly older patients with preexisting heart disease, it seems unlikely that many participants would be actively undergoing HCV treatment at the time of this study. Finally, we examined HF hospitalizations, rather than true incident HF. However, when we excluded individuals who were diagnosed with HF at baseline, we found that HCV was still associated with HF hospitalizations (presumably representing new cases of HF), although we did not have echocardiography data during hospitalizations to confirm.

In conclusion, among this cohort with established CHD, we found that HCV-seropositive participants had higher rates of death, CV events, and HF hospitalizations over time, despite lower cholesterol and CRP levels. After adjustment for other factors, HCV seropositivity remained independently associated with risk for HF hospitalizations, and this was true when excluding participants with a preexisting diagnosis of HF. Levels of CRP, fibrinogen, IL-6, and TNF-α did not appear to mediate the association between HCV and HF, so other causal pathways must be evoked. Further research is needed to confirm these results and explore mechanisms through which HCV might impact risk for cardiovascular disease and disease outcomes.

References

1. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. Ann Intern Med 2006;144:705–14.
2. Danesh J, Collins R, Peto R. Chronic infections and coronary heart disease: is there a link? Lancet 1997;350:430–6.
3. Matsumori A, Matoba Y, Sasayama S. Dilated cardiomyopathy associated with hepatitis C virus infection. Circulation 1995;92:2519–25.
4. Matsumori A, Shimada T, Chapman NM, Tracy SM, Mason JW. Myocarditis and heart failure associated with hepatitis C virus infection. J Card Fail 2006;12:293–8.
5. Freiberg MS, Cheng DM, Kraemer KL, Saitz R, Kuller LH, Samet JH. The association between hepatitis C infection and prevalent cardiovascular disease among HIV-infected individuals. AIDS 2007;21:193–7.
6. Fukui M, Kitagawa Y, Nakamura N, Yoshikawa T. Hepatitis C virus and atherosclerosis in patients with type 2 diabetes. JAMA 2003; 289:1245–6.
7. Vassalle C, Masini S, Bianchi F, Zucchelli GC. Evidence for association between hepatitis C virus seropositivity and coronary artery disease. Heart 2004;90:565–6.
8. Ishizaka N, Ishizaka Y, Takahashi E, Tooda E, Hashimoto H, Nagai R, et al. Association between hepatitis C virus seropositivity, carotid-artery plaque, and intima-media thickening. Lancet 2002;359:133–5.
9. Ishizaka Y, Ishizaka N, Takahashi E, Unuma T, Tooda E, Hashimoto H, et al. Association between hepatitis C virus core protein and carotid atherosclerosis. Circ J 2003;67:26–30.
10. Tomiyama H, Arai T, Hirose K, Hori S, Yamamoto Y, Yamashina A. Hepatitis C virus seropositivity, but not hepatitis B virus carrier or seropositivity, associated with increased pulse wave velocity. Atherosclerosis 2003;166:401–3.

11. Volzke H, Schwahn C, Wolff B, Mentel R, Robinson DM, Kleine V, et al. Hepatitis B and C virus infection and the risk of atherosclerosis in a general population. Atherosclerosis 2004;174:99–103.
12. Momiyama Y, Ohmori R, Kato R, Taniguchi H, Nakamura H, Ohsuzu F. Lack of any association between persistent hepatitis B or C virus infection and coronary artery disease. Atherosclerosis 2005; 181:211–3.
13. Moriani M, Adachi K, Arima N, Takashima T, Miyaoka Y, Niigaki M, et al. A study of arteriosclerosis in healthy subjects with HBV and HCV infection. J Gastroenterol 2005;40:1049–53.
14. Dalekos GN, Achenbach K, Christodoulou D, Liapi GK, Zervou EK, Sideris DA, et al. Idiopathic dilated cardiomyopathy: lack of association with hepatitis C virus infection. Heart 1998;80:270–5.
15. Grumbach IM, Heermann K, Figulla HR. Low prevalence of hepatitis C virus antibodies and RNA in patients with myocarditis and dilated cardiomyopathy. Cardiology 1998;90:75–8.
16. Guiltinan AM, Kaidarova Z, Custer B, Orland J, Strollo A, Cyrus S, et al. Increased all-cause, liver, and cardiac mortality among hepatitis C virus-seropositive blood donors. Am J Epidemiol 2008;167: 743–50.
17. Jacobson Brown PM, Neuman MG. Immunopathogenesis of hepatitis C viral infection: Th1/Th2 responses and the role of cytokines. Clin Biochem 2001;34:167–71.
18. Hansson GK. Immune mechanisms in atherosclerosis. Arterioscler Thromb Vasc Biol 2001;21:1876–90.
19. Ruo B, Rumsfeld JS, Hlatky MA, Liu H, Browner WS, Whooley MA. Depressive symptoms and health-related quality of life: the Heart and Soul Study. JAMA 2003;290:215–21.
20. Bibbins-Domingo K, Gupta R, Na B, Wu AH, Schiller NB, Whooley MA. N-terminal fragment of the prohormone brain-type natriuretic peptide (NT-proBNP), cardiovascular events, and mortality in patients with stable coronary heart disease. JAMA 2007;297: 169–76.
21. Alter MJ, Kuhnert WL, Finelli L. Guidelines for laboratory testing and result reporting of antibody to hepatitis C virus. Centers for Disease Control and Prevention. MMWR Recomm Rep 2003;52:1–13. 15; quiz CE1-4.
22. Clauss A. Rapid physiological coagulation method in determination of fibrinogen. Acta Haematol 1957;17:237–46.
23. Luepker RV, Apple FS, Christenson RH, Crow RS, Fortmann SP, Goff D, et al. Case definitions for acute coronary heart disease in epidemiology and clinical research studies: a statement from the AHA Council on Epidemiology and Prevention; AHA Statistics Committee; World Heart Federation Council on Epidemiology and Prevention; the European Society of Cardiology Working Group on Epidemiology and Prevention; Centers for Disease Control and Prevention; and the National Heart, Lung, and Blood Institute. Circulation 2003;108:2543–9.
24. Redfield MM, Jacobsen SJ, Burnett JC Jr, Mahoney DW, Bailey KR, Rodeheffer RJ. Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. JAMA 2003;289:194–202.
25. Omura T, Yoshiyama M, Hayashi T, Nishiguchi S, Kaito M, Horiike S, et al. Core protein of hepatitis C virus induces cardiomyopathy. Circ Res 2005;96:148–50.
26. Haji SA, Starling RC, Avery RK, Mawhorter S, Tuzcu EM, Schoenhagen P, et al. Donor hepatitis C seropositivity is an independent risk factor for the development of accelerated coronary vasculopathy and predicts outcome after cardiac transplantation. J Heart Lung Transplant 2004;23:277–83.
27. Gasink LB, Blumberg EA, Localio AR, Desai SS, Israni AK, Lautenbach E. Hepatitis C virus seropositivity in organ donors and survival in heart transplant recipients. JAMA 2006;296:1843–50.
28. Deswal A, Petersen NJ, Feldman AM, Young JB, White BG, Mann DL. Cytokines and cytokine receptors in advanced heart failure: an analysis of the cytokine database from the Vesnarinone trial (VEST). Circulation 2001;103:2055–9.
29. Rauchhaus M, Doehner W, Francis DP, Davos C, Kemp M, Liebenthal C, et al. Plasma cytokine parameters and mortality in patients with chronic heart failure. Circulation 2000;102:3060–7.
30. Torre-Amione G, Kapadia S, Benedict C, Oral H, Young JB, Mann DL. Proinflammatory cytokine levels in patients with depressed left ventricular ejection fraction: a report from the Studies of Left Ventricular Dysfunction (SOLVD). J Am Coll Cardiol 1996;27:1201–6.
31. Tsutamoto T, Hisanaga T, Wada A, Maeda K, Ohnishi M, Fukai D, et al. Interleukin-6 spillover in the peripheral circulation increases with the severity of heart failure, and the high plasma level of interleukin-6 is an important prognostic predictor in patients with congestive heart failure. J Am Coll Cardiol 1998;31:391–8.
32. Nelson DR, Lim HL, Marousis CG, Pang JW, Davis GL, Shen L, et al. Activation of tumor necrosis factor-alpha system in chronic hepatitis C virus infection. Dig Dis Sci 1997;42:2487–94.
33. Blackard JT, Kang M, St Clair JB, Lin W, Kamegaya Y, Sherman KE, et al. Viral factors associated with cytokine expression during HCV/HIV co-infection. J Interferon Cytokine Res 2007;27:263–9.
34. Knobler H, Zhornicky T, Sandler A, Haran N, Ashur Y, Schattner A. Tumor necrosis factor-alpha-induced insulin resistance may mediate the hepatitis C virus-diabetes association. Am J Gastroenterol 2003; 98:2751–6.
35. Zylberberg H, Rimaniol AC, Pol S, Masson A, De Groote D, Berthelot P, et al. Soluble tumor necrosis factor receptors in chronic hepatitis C: a correlation with histological fibrosis and activity. J Hepatol 1999;30:185–91.
36. Serfaty L, Andreani T, Giral P, Carbonell N, Chazouilleres O, Poupon R. Hepatitis C virus induced hypobetalipoproteinemia: a possible mechanism for steatosis in chronic hepatitis C. J Hepatol 2001;34: 428–34.
37. Maggi G, Bottelli R, Gola D, Perricone G, Posca M, Zavaglia C, et al. Serum cholesterol and chronic hepatitis C. Ital J Gastroenterol 1996; 28:436–40.
38. Ferroni P, Mammarella A, Martini F, Paoletti V, Cardarello CM, Labbadia G, et al. Increased soluble P-selectin levels in hepatitis C virus-related chronic hepatitis: correlation with viral load. J Investig Med 2001;49:407–12.
39. Marzouk D, Sass J, Bakr I, El Hosseiny M, Abdel-Hamid M, Rekacewicz C, et al. Metabolic and cardiovascular risk profiles and hepatitis C virus infection in rural Egypt. Gut 2007;56:1105–10.
40. Floris-Moore M, Howard AA, Lo Y, Schoenbaum EE, Arnsten JH, Klein RS. Hepatitis C infection is associated with lower lipids and high-sensitivity C-reactive protein in HIV-infected men. AIDS Patient Care STDS 2007;21:479–91.
41. Kalabay L, Nemesanszky E, Csepregi A, Pusztay M, David K, Horvath G, et al. Paradoxical alteration of acute-phase protein levels in patients with chronic hepatitis C treated with IFN-alpha2b. Int Immunol 2004;16:51–4.

Using Video Images to Improve the Accuracy of Surrogate Decision-Making: A Randomized Controlled Trial

Angelo E. Volandes, MD, MPH, Susan L. Mitchell, MD, MPH, Muriel R. Gillick, MD, Yuchiao Chang, PhD, and Michael K. Paasche-Orlow, MD, MA, MPH

Introduction: When patients are unable to make important end-of-life decisions, doctors ask surrogate decision makers to provide insight into patients' preferences. Unfortunately, multiple studies have shown that surrogates' knowledge of patient preferences is poor. We hypothesized that a video decision tool would improve concordance between patients and their surrogates for end-of-life preferences.

Objective: To compare the concordance of preferences among elderly patients and their surrogates listening to only a verbal description of advanced dementia or viewing a video decision support tool of the disease after hearing the verbal description.

Methods: This was a randomized controlled trial of a convenience sample of community-dwelling elderly subjects (≥65 years) and their surrogates, and was conducted at 2 geriatric clinics affiliated with 2 academic medical centers in Boston. The study was conducted between September 1, 2007, and May 30, 2008. Random assignment of patient and surrogate dyads was to either a verbal narrative or a video decision support tool after the verbal narrative. End points were goals of care chosen by the patient and predicted goals of care by the surrogate. Goals of care included life-prolonging care (CPR, mechanical ventilation), limited care (hospitalization, antibiotics, but not CPR), and comfort care (only treatment to relieve symptoms). The primary outcome measure was the concordance rate of preferences between patients and their surrogates.

Results: A total of 14 pairs of patients and their surrogates were randomized to verbal narrative (n = 6) or video after verbal narrative (n = 8). Among the 6 patients receiving only the verbal narrative, 3 (50%) preferred comfort care, 1 (17%) chose limited care, and 2 (33%) desired life-prolonging care. Among the surrogates for these patients, only 2 correctly chose what their loved one would want if in a state of advanced dementia, yielding a concordance rate of 33%. Among the 8 patients receiving the video decision support tool, all 8 chose comfort care. Among the surrogates for these patients, all 8 correctly chose what their loved one would want if in a state of advanced dementia, yielding a concordance rate of 100%.

Conclusion: Patients and surrogates viewing a video decision support tool for advanced dementia are more likely to concur about the patient's end-of-life preferences than when solely listening to a verbal description of the disease. **(J Am Med Dir Assoc 2009; 10: 575–580)**

General Medicine Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA (A.E.V., Y.C.); John D. Stoeckle Center for Primary Care Innovation, Massachusetts General Hospital, Boston, MA (A.E.V.); Harvard Medical School, Boston, MA (A.E.V., S.L.M., M.R.G.); Hebrew SeniorLife, Boston, MA (S.L.M.); Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA (S.L.M.); Department of Ambulatory Care and Prevention, Harvard Medical School/Harvard Pilgrim, Boston, MA (M.R.G.); Department of Medicine, Boston University School of Medicine, Boston, MA (M.K.P.-O.).

Dr. Volandes was supported by a George Bennett Fellowship from the Foundation for Informed Medical Decision Making, a New Investigator Research Grant from the Alzheimer's Association, and a Center for Excellence Career Development Award from the Hartford Foundation. None of the Foundations participated in the collection, analysis, or interpretation of the data or in preparation, review, or approval of the manuscript.

Address correspondence to Angelo E. Volandes, MD, MPH, Massachusetts General Hospital, General Medicine Unit, 50 Staniford Street, 9th floor, Boston, MA 02114. E-mail: avolandes@partners.org

DOI:10.1016/j.jamda.2009.05.006

When patients are unable to make important end-of-life decisions, doctors ask surrogate decision makers to provide insight into patients' preferences.[1–3] Surrogates, who are often family members, are often instructed to use an idealized hierarchy of standards to guide decision making: patients' known wishes, substituted judgments, and patients' best interests.[4] Unfortunately, many surrogates do not know the wishes of the people they are supposed to represent.

The substituted judgment standard attempts to have surrogates imagine and predict what the patient would have wanted.[4] A large number of clinical research studies have convincingly concluded that surrogates are often inaccurate in predicting the medical preferences of patients when offered written or verbal hypothetical health states.[5–10] The overwhelming conclusion of these studies is that surrogates are

often no better than chance at predicting patients' preferences for future health states. Theories abound regarding why surrogates inaccurately predict patient preferences, including the role of family dynamics; the relative ease of overtreatment, also known as the "status quo bias"; and personal psychological stress regarding the burden of end-of-life decision making.[11–13]

One overlooked reason of why surrogates incorrectly predict patient preferences may be a lack of comprehension about hypothetical health states and likely treatment outcomes. Central to the process of predicting patient preferences is the surrogate's understanding of the underlying health state. Empirical studies often communicate information about future health states to both patients and surrogates with written or spoken words, and both parties must be willing to imagine often difficult and uncomfortable scenarios.

Video is a powerful and underused medium that can better communicate hypothetical health states[14–16] and may assist patients and surrogates in discussions of preferences. The medium of video engages and allows both patients and surrogates to envision future health states in a manner not captured with verbal communication. We hypothesized that video may enable patients and surrogates to better visualize and imagine a future health state, which would lead to more accurate surrogate predictions that were more concordant with patient preferences.

As part of a larger study on the use of video decision aids, we conducted a small randomized controlled trial of dyads of patients and designated surrogates to examine whether a video of a person with advanced dementia would lead surrogates and patients to have concordant responses to questions about end-of-life care. We hypothesized that patients and surrogates viewing a video decision-support tool for advanced dementia would be more likely to concur about the patient's end-of-life preferences than when solely listening to a verbal description of the disease.

METHODS

Participants

The protocol was approved by the institutional review board for both institutions and all subjects provided informed consent. This study was part of a larger study evaluating the use of video decision-support tools in advance care planning.[17] Elderly patients participating in the larger study whose surrogates were present during the clinic visit were asked to participate in the present study. Patients and their surrogates were recruited from a convenience sample at 2 urban geriatric clinics affiliated with 2 teaching hospitals in the greater Boston area. Recruitment occurred between September 1, 2007, and May 30, 2008. All scheduled English-speaking patients 65 years or older who presented to the clinic with their designated surrogate were given a flier by the clinic staff outlining the study. At the end of the clinic visit, patients and surrogates were asked by clinic staff if they were interested in participating in the study. If they indicated interest, the patient was brought to a private room alone and was initially interviewed for eligibility based on a Short Portable Mental Status Questionnaire (SPMSQ)[18] score of greater than or equal to 7 (scores <7 indicate moderate or severe cognitive impairment) and the ability to provide informed consent. After the patient completed the interview, the surrogate, who was not present during the patient's interview, was then brought into the private room and completed the interview separately. Inclusion criteria for the surrogate were ability to provide informed consent and being the patient's designated surrogate.

Study Design and Randomization

After obtaining informed consent from both the patient and the surrogate, all patient-surrogate dyads were randomized into 1 of 2 decision-making modalities: (1) listening to a verbal narrative describing advanced dementia (control group); or (2) listening to a verbal narrative followed by viewing a 2-minute video decision-support tool visually depicting a patient with advanced dementia (intervention group). Randomization was based on a computer-generated randomization scheme, and followed the randomization order of the larger study from which this subgroup was taken. Individual assignments were concealed in numbered envelopes until the pair was randomized. All data were collected in a quiet room in the clinic area by a trained member of the research team (A.E.V.) who followed a structured script. The patient and surrogate were interviewed separately and were unaware of each other's answers to the survey.

For both randomization groups, the interviewer read aloud a description of advanced dementia based on the Functional Assessment Staging (FAST)[19] stage 7a. The FAST criteria include 7 stages of dementia (1–7), with the later stages depicting more advanced disease. Stage 7 is further broken down into 6 substages (7a–7f). Stage 7a is generally considered the threshold for advanced dementia, and the threshold for advanced dementia used in our previous studies.[14–16] Advanced dementia was described as an incurable illness of the brain caused by many years of Alzheimer's disease or a series of strokes; its salient features are the inability to communicate understandably with others, inability to walk without assistance, and inability to feed oneself (see Appendix 1).

Patients randomized to the intervention group viewed the video decision support tool on a portable computer after listening to the same verbal narrative. The 2-minute video depicts the principal features of advanced dementia as described in the narrative. The video presents an 80-year-old female patient with advanced dementia together with her 2 daughters in the nursing home setting. The patient fails to respond to their attempts at conversation (inability to communicate). The patient is next shown being pushed in a wheelchair (inability to ambulate). Last, the patient is hand-fed pureed food (inability to feed oneself). Consent to film the patient with advanced dementia and to use the video for research purposes was obtained from the patient's designated health care proxy before filming.

The development of the video followed a systematic approach,[20] starting with a review of the dementia literature. We then used a panel of physicians with an iterative process of comments to review the design, content, and structure of

the video intervention. This panel included 5 geriatricians and 5 neurologists, all of whom specialize in the care of patients with dementia.

The video was filmed without the use of prompts or stage directions to convey a candid realism in the style known as cinema verite.[21] All filming and editing were done by the principal investigator (A.E.V.) following previously published filming criteria.[22] (The video is available at: www.ACPdecisions.com.)

Data Collection and Other Variables

The interviewer was not blinded to randomization group. Each patient was interviewed before and after receiving the verbal narrative alone or the narrative plus the video decision-support tool using structured questionnaires. The baseline structured interview (15 minutes) included the following components: demographic data and knowledge about advanced dementia. Sociodemographic data included age, race, gender, educational status, and marital status. Race was self-reported. Having a previous relationship with someone with advanced dementia was also obtained. Knowledge of advanced dementia was assessed using 5 true/false questions that asked patients and their surrogates whether advanced dementia is curable and if patients with advanced dementia are able to communicate with others, recognize family members, ambulate, and feed themselves. Thus, knowledge scores ranged from 0 to 5, with higher scores indicating better knowledge.

Immediately after receiving the verbal narrative alone or narrative plus video, a second structured in-person interview (15 minutes) was conducted that included the following components: knowledge of advanced dementia and preferences for goals of care; and for the intervention group, comfort using the video decision-support tool. The knowledge questions were identical to those asked in the baseline interview.

Each surrogate, who was not present during the interview with the patient and was unaware of the patient's answers, was asked an identical set of questions with the sole exception that each surrogate was asked to predict the preferences for the goals of care for their loved one using the substituted judgment criterion.

Preferences for goals of care were presented as 3 options: life-prolonging care, limited care, and comfort care (see Appendix 1). Examples of the kinds of care implied by each goal were verbally described to participants. The first option, life-prolonging care, was described as aiming to prolong life at any cost. It translates into all potentially indicated medical care that is available in a modern-day hospital, including cardiopulmonary resuscitation and treatment in the intensive care unit. The second option, limited care, was described as aiming to maintain physical functioning. It is consistent with treatments such as hospitalization, intravenous fluids, and antibiotics, but not with attempted cardiopulmonary resuscitation (CPR) and treatment in the intensive care unit (ICU). The third option, comfort care, was described as aiming to maximize comfort and to relieve pain. Treatments are focused on the relief of symptoms. It is compatible with oxygen and analgesics but not with intravenous (IV) therapies and hospitalization unless necessary to provide comfort. The aim is to relieve pain and to be kept as pain-free as possible. Comfort care does not include CPR, respirators, ICU care, and generally would not include IV therapy or hospitalization. Following these explanations, patients were asked their preferences in the event they developed advanced dementia. Surrogates were asked to predict which preference their loved ones would pick.

For those patients and surrogates randomized to the video intervention group, a 4-point Likert scale was used to assess perceived value of the video by asking them whether they had a better understanding of the disease after viewing the video, if they were comfortable watching the video, and if they would recommend the video to others. These questions were asked at the end of the oral survey. The survey is available on request.

Statistical Analysis

Patient-surrogate dyads were analyzed based on the decision-making modality to which they were randomized. The primary outcome measure for patients was their preferences for care if in a state of advanced dementia categorized as 3 options (life-prolonging, limited, or comfort). The primary outcome measure for surrogates was the preferences they felt their loved one would choose based on the substituted judgment criterion. Additional outcomes included change in knowledge scores for both patients and subjects before compared with after receiving the verbal narrative or video.

All subject characteristics and outcomes were described using proportions for categorical variables and means (SD) for continuous variables. Concordance rates for patient-surrogate dyads were the proportions of surrogates who chose the same preference as their loved one over the total number of dyads in the randomization group. Chi-square tests were used to compare the concordance rate and 2 sample t tests were used to compare change in knowledge scores from before to after the intervention between the 2 randomization groups. All reported P values are 2-sided, with $P < .05$ considered as statistically significant. Data were analyzed using SAS software, version 9.1 (SAS Institute Inc, Cary, NC).

RESULTS

Participant Flow

A total of 14 consecutive pairs of eligible patient-surrogate dyads were approached to participate in the study, all of whom agreed to be interviewed and were eligible to participate. Six patients were randomized to the control group, and 8 patients were randomized to the video-intervention group. Baseline characteristics of the patients and surrogates are shown in Table 1. Of the 14 designated surrogates, 7 were spouses of the patients, 5 were children of the patients, 1 was a sibling, and 1 was a friend. None had a previous relationship with someone with advanced dementia. Although the power to detect differences was small, there were no significant differences in either patients or surrogates in age, gender, or education.

Outcomes

Among the 6 patients receiving only the verbal narrative, 3 (50%) preferred comfort care, 1 (17%) chose limited care,

Table 1. *Characteristics of Community-dwelling Elderly Patients and Their Surrogates*

Characteristics	Elderly Persons (n=14)	Surrogates (n=14)
Age, mean (SD), y	83 (6.9)	67.5 (13.7)
Women, no. (%)	7 (50)	11 (78.6)
Race, no. (%)		
White	14 (100)	14 (100)
Education, no. (%)		
High school graduate or less	5 (35.7)	1 (7.1)
Some college or higher	9 (64.3)	13 (92.9)
Religion, no. (%)		
Catholic	5 (35.7)	6 (42.9)
Christian (non-Catholic)	6 (42.9)	5 (35.7)
Jewish	3 (21.4)	1 (7.1)
Other		2 (14.3)

and 2 (33%) desired life-prolonging care (Figure 1). Among the 6 surrogates receiving only the verbal narrative, only 2 predicted correctly what their loved one would want if in a state of advanced dementia, yielding a concordance rate of 33% (Figure 1). Of the 4 surrogates who misjudged the preferences of the patients, 3 surrogates chose less aggressive care compared with their loved ones.

Among the 8 patients receiving the video decision-support tool after the verbal narrative, all 8 chose comfort care. Among the 8 surrogates receiving the video decision-support tool as well, all 8 predicted correctly what their loved one would want if in a state of advanced dementia, yielding a concordance rate of 100% ($P = .015$ compared with the verbal narrative–alone group).

Knowledge scores increased for patients in both groups post intervention; however, the changes were higher in the narrative plus video group compared with those in the narrative-alone group (2.1 ± 1.6 versus 0.3 ± 1.6, respectively; $P = .068$). The change in knowledge scores was also higher for surrogates in the narrative plus video group compared with those in the narrative-alone group (2.5 ± 1.4 versus 2.0 ± 1.3, respectively; $P = .50$).

The video decision-support tool was very well accepted by patients and surrogates in the intervention group: 15 of 16 (94%) patients and surrogates found the video "very helpful" or "somewhat helpful"; 14 (88%) said they were "very comfortable" or "somewhat comfortable" viewing the video; and 15 (94%) said they would "definitely" or "probably" recommend the video to others. There were no adverse events in either group.

DISCUSSION

Patients and surrogates viewing a video decision-support tool for advanced dementia are more likely to concur about the patient's end-of-life preferences than when listening to a verbal description of the disease. Moreover, viewing the video decision-support tool was associated with a trend toward more knowledge of advanced dementia among both patients and their surrogates. The fact that participants' decisions with the video were more informed is consistent with prior research of ours that has also exhibited better knowledge, better certainty, and better stability of preferences for patients' end-of-life preferences after viewing a video decision-support tool.[14,17]

In the current project, surrogates' decisions were more likely to be concordant with patients' preferences. However, it is possible that our findings should be viewed in a different way. The instances of discordance between surrogates and patients all occurred among people who did not see the video and most were instances in which the patient wanted more aggressive medical care than suggested by the surrogates. As such, it is possible that our findings should be viewed primarily as a tool that improved decisions by patients and then, only as a consequence, similarly improved the rate of concordance between patients and surrogates.

Our study has several important limitations. First, the researcher surveying the patient-surrogate dyads was not blinded to the randomization assignment. This could have introduced bias into our findings. Prior randomized studies of interventions aimed at improving end-of-life decision

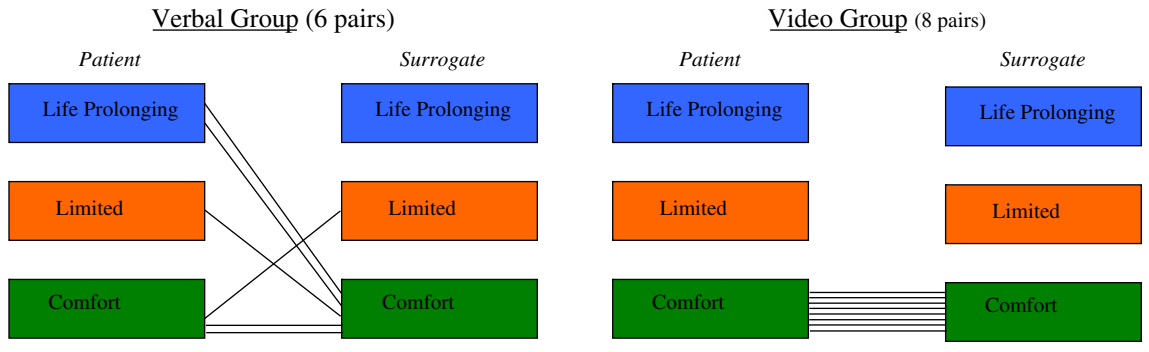


Fig. 1. *Patient and surrogate preferences by randomization group.*

making have seldom been blinded because limiting the number of interviewers eases the burden of addressing difficult and often painful subject matter.[23–25] We attempted to reduce the influence of this potential bias by using structured interviews and outcome measures. Second, video clips can be manipulated to favor a particular perspective. Although our carefully crafted video followed published filming criteria on using video in end-of-life discussions,[22] there may have been potential bias in the evaluation of the decision-support tool. It would be fruitful to study other video clips of similar patients. Third, the sample size was very small and did not permit analyses to evaluate the role of factors such as gender and race, which would have permitted relevant analyses to evaluate potential bias in the tool. Fourth, we asked subjects for their preferences in the context of a research study. The next step would be to investigate whether surrogates would correctly predict preferences in real time. Fifth, we asked subjects the same questions before and after the intervention. A more rigorous method of testing changes in knowledge and preferences would have been to use parallel forms of the questions to avoid biasing the results. Finally, our sample was drawn from the metro Boston area and did not include large minority groups such as African Americans, Latinos, and Asian Americans.

Including surrogates in the decision-making process has been an important yet complex advancement in modern medical care. To secure the delivery of high-quality end-of-life care that is concordant with patients' preferences, surrogates must be informed regarding their decision making. As has been shown in previous studies, using the substituted criterion is a complex task for surrogates, asking them to imagine what are often unimaginable health states for laypersons. Patient education using video decision-support tools can improve surrogate comprehension of disease states such as advanced dementia that are difficult to envision solely with words. Future work with surrogates may extend the use of video decision-support tools to other disease states such as advanced cancer. This study provides evidence that video decision-support tools enhance patients' and surrogates' decision making by ensuring that it is both more informed and concordant with subjects' wishes at the end of life.

REFERENCES

1. Buchanan AE, Brock DW. Deciding for others: The ethics of surrogate decision making. New York, NY: Cambridge University Press; 1989.
2. Brock DW. What is the moral authority of family members to act as surrogates for incompetent patients? Milbank Q 1996;74:599–618.
3. Arnold RM, Kellum J. Moral justifications for surrogate decision making in the intensive care unit: implications and limitations. Crit Care Med 2003;31:S347–S353.
4. Beauchamp TL, Childress JF. Principles of Biomedical Ethics. 6th ed. New York: Oxford University Press; 2009.
5. Teno JM, Lynn J, Phillips RS, et al. Do formal advance directives affect resuscitation decisions and the use of resources for seriously ill patients? SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. J Clin Ethics 1994;5:23–30.
6. Layde PM, Beam CA, Broste SK, et al. Surrogates' predictions of seriously ill patients' resuscitation preferences. Arch Fam Med 1995;4:518–523.
7. Gerety MB, Chiodo LK, Kanten DN, et al. Medical treatment preferences of nursing home residents: Relationship to function and concordance with surrogate decision-makers. J Am Geriatr Soc 1993;41:953–960.
8. Seckler AB, Meier DE, Mulvihill M, Paris BE. Substituted judgment: How accurate are proxy predictions? Ann Intern Med 1991;115:92–98.
9. Zweibel NR, Cassel CK. Treatment choices at the end of life: A comparison of decisions by older patients and their physician-selected proxies. Gerontologist 1989;29:615–621.
10. Suhl J, Simons P, Reedy T, Garrick T. Myth of substituted judgment. Surrogate decision making regarding life support is unreliable. Arch Intern Med 1994;154:90–96.
11. Sulmasy DP, Sood JR, Texiera K, et al. A prospective trial of a new policy eliminating signed consent for do not resuscitate orders. J Gen Intern Med Sep 11, 2006 [Epub ahead of print].
12. Tilden VP, Tolle SW, Nelson CA, Fields J. Family decision-making to withdraw life-sustaining treatments from hospitalized patients. Nurs Res 2001;50:105–115.
13. Anderson CJ. The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. Psychol Bull 2003;129:139–167.
14. Volandes AE, Lehmann LS, Cook EF, et al. Using video images of dementia in advance care planning. Arch Intern Med 2007;167:828–833.
15. Volandes AE, Ariza M, Abbo ED, Paasche-Orlow M. Overcoming educational barriers for advance care planning in latinos with video images. J Palliat Med 2008;11:700–706.
16. Volandes AE, Paasche-Orlow M, Gillick MR, et al. Health literacy not race predicts end-of-life care preferences. J Palliat Med 2008;11:754–762.
17. Volandes A, Paasche-Orlow M, Barry M, Mitchell SL. A video decision support tool for advance care planning in dementia: A randomized controlled trial. BMJ 2009;338:b2159.
18. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975;23:433–441.
19. Sclan SG, Reisberg B. Functional assessment staging (FAST) in Alzheimer's disease: Reliability, validity, and ordinality. Int Psychogeriatr 1992;4:55–69.
20. National Cancer Institute. Making health communication programs work: a planner's guide. June 26, 2008. Available at: http://www.cancer.gov/pinkbook/page1. Accessed May 1, 2009.
21. Grant BK, Sloniowski J. Documenting the Documentary: Close Readings of Documentary Film and Video. Detroit: Wayne State University Press; 1998.
22. Gillick MR, Volandes AE. The psychology of using and creating video decision aids for advance care planning. In: Lynch TE, editor. Psychology of Decision Making in Medicine and Health Care. New York, NY: Nova Science Publishers; 2007. p. 193–206.
23. Support. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators. JAMA 1995;274:1591–1598.
24. Hofmann JC, Wenger NS, Davis RB, et al. Patient preferences for communication with physicians about end-of-life decisions. SUPPORT Investigators. Study to Understand Prognoses and Preference for Outcomes and Risks of Treatment. Ann Intern Med 1997;127:1–12.
25. Covinsky KE, Fuller JD, Yaffe K, et al. Communication and decision-making in seriously ill patients: Findings of the SUPPORT project. The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. J Am Geriatr Soc 2000;48:S187–S193.

Appendix 1

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 with advanced dementia are not able to move around or walk, get out of bed independently, eat by oneself, or communicate understandably with others. People with advanced dementia often have difficulty chewing or swallowing, and require assistance with feeding oneself. Advanced dementia is an incurable disease and most commonly occurs after many years of Alzheimer's disease or as the result of strokes. People are not able to answer any questions or tell you about themselves."

Narrative describing the goals of care

"I am going to ask you a question about your preferences for medical care if you had a disease called advanced dementia. I will ask you what you prefer. You have 3 choices for medical care if you had this condition. I will first review these 3 choices with you. The 3 choices for medical care that I want you to think about for advanced dementia are Life-Prolonging Care, Limited Care, and Comfort Care.

Life-Prolonging Care

The goal of this category of care is to prolong life. There are no limits to care. This choice includes everything a modern hospital has to offer to maintain your life. Such procedures include: cardio-pulmonary resuscitation or CPR in which a doctor pushes on your chest when the heart stops and will often use electricity to shock the heart. Being placed on a breathing machine, also known as life support, in which a tube is placed down your throat into the lungs. And other medical procedures performed in the intensive care unit or ICU. The goal is to prolong life.

Limited Care

The goal of this category is to maintain physical and mental functions. Care will depend on your physical and mental functioning. Such care includes intravenous (IV) therapies like antibiotics and hospitalization. But does not include cardiopulmonary resuscitation/CPR and intensive care unit/ICU care. The goal is to maintain physical and mental functioning.

Comfort Care

The goal of this category is to maximize comfort. Only measures that comfort or relieve pain are performed. The aim is to relieve pain and to be kept as pain-free as possible. Comfort Care does not include cardiopulmonary resuscitation/CPR, respirators, intensive care unit/ICU care, and generally would not include intravenous (IV) therapy or hospitalization. The goal is maximizing comfort and relieving pain.

BMJ

RESEARCH

Video decision support tool for advance care planning in dementia: randomised controlled trial

Angelo E Volandes, instructor,[1] Michael K Paasche-Orlow, associate professor,[2] Michael J Barry, professor,[1] Muriel R Gillick, professor,[3] Kenneth L Minaker, professor,[1] Yuchiao Chang, senior statistician,[1] E Francis Cook, professor,[4] Elmer D Abbo, assistant professor,[5] Areej El-Jawahri, research fellow,[1] Susan L Mitchell, associate professor [6]

[1]General Medicine Unit, Department of Medicine, Massachusetts General Hospital, 50 Staniford Street, Boston, MA 02114, USA

[2]Department of Medicine, Boston University School of Medicine, 715 Albany Street, Boston, MA 02118, USA

[3]Department of Ambulatory Care and Prevention, Harvard Medical School/Harvard Pilgrim, 133 Brookline Ave, Boston, MA 02215, USA

[4]Department of Medicine, Brigham and Women's Hospital, 1620 Tremont Street, Boston, MA 02115, USA

[5]Department of Medicine, University of Chicago, 5801 South Ellis, Chicago, IL 60637, USA

[6]Hebrew SeniorLife, 1200 Centre Street, Boston, MA 02131, USA

Correspondence to: A E Volandes
avolandes@partners.org

Cite this as: *BMJ* 2009;338:b1964
doi:10.1136/bmj.b2159

ABSTRACT

Objective To evaluate the effect of a video decision support tool on the preferences for future medical care in older people if they develop advanced dementia, and the stability of those preferences after six weeks.

Design Randomised controlled trial conducted between 1 September 2007 and 30 May 2008.

Setting Four primary care clinics (two geriatric and two adult medicine) affiliated with three academic medical centres in Boston.

Participants Convenience sample of 200 older people (≥ 65 years) living in the community with previously scheduled appointments at one of the clinics. Mean age was 75 and 58% were women.

Intervention Verbal narrative alone (n=106) or with a video decision support tool (n=94).

Main outcome measures Preferred goal of care: life prolonging care (cardiopulmonary resuscitation, mechanical ventilation), limited care (admission to hospital, antibiotics, but not cardiopulmonary resuscitation), or comfort care (treatment only to relieve symptoms). Preferences after six weeks. The principal category for analysis was the difference in proportions of participants in each group who preferred comfort care.

Results Among participants receiving the verbal narrative alone, 68 (64%) chose comfort care, 20 (19%) chose limited care, 15 (14%) chose life prolonging care, and three (3%) were uncertain. In the video group, 81 (86%) chose comfort care, eight (9%) chose limited care, four (4%) chose life prolonging care, and one (1%) was uncertain (χ^2=13.0, df=3, P=0.003). Among all participants the factors associated with a greater likelihood of opting for comfort care were being a college graduate or higher, good or better health status, greater health literacy, white race, and randomisation to the video arm. In multivariable analysis, participants in the video group were more likely to prefer comfort care than those in the verbal group (adjusted odds ratio 3.9, 95% confidence interval 1.8 to 8.6). Participants were re-interviewed after six weeks. Among the 94/106 (89%) participants re-interviewed in the verbal group, 27 (29%) changed their preferences (κ=0.35). Among the 84/94 (89%) participants re-interviewed in the video group, five (6%) changed their preferences (κ=0.79) (P<0.001 for difference).

Conclusion Older people who view a video depiction of a patient with advanced dementia after hearing a verbal description of the condition are more likely to opt for comfort as their goal of care compared with those who solely listen to a verbal description. They also have more stable preferences over time.

Trial registration Clinicaltrials.gov NCT00704886.

INTRODUCTION

Respecting patients' preferences for treatment is a key component of high quality end of life care.[1-4] Traditionally, physicians help patients to engage in advance care planning for future health states by describing hypothetical situations such as advanced dementia and by exploring possible goals of care.[5 6] This traditional approach is limited because it is challenging to realistically envision hypothetical future disease states such as dementia from verbal descriptions,[7] descriptions are inconsistent among providers,[8-15] and the degree to which patients understand verbal descriptions of complex medical conditions depends on their level of health literacy.

Visual images can improve communication of complex health information[16-19] and inform decision making at the end of life.[20-22] In our previous investigations, a video decision support tool for advanced dementia seemed to improve communication and decision making for patients by helping them to visualise future health states.[20-22] However, there were significant shortcomings to these studies: they were conducted in healthy middle aged patients; they used a before and after study design that did not allow comparison of the video to the standard advance care planning approach of a verbal narrative; they did not measure knowledge of the disease to test whether understanding of the disease improved; and they did not follow patients' preferences over time.

To address these shortcomings, we conducted a randomised controlled trial of the video decision support tool among a diverse group of older patients to study the video with a higher level of rigour. We

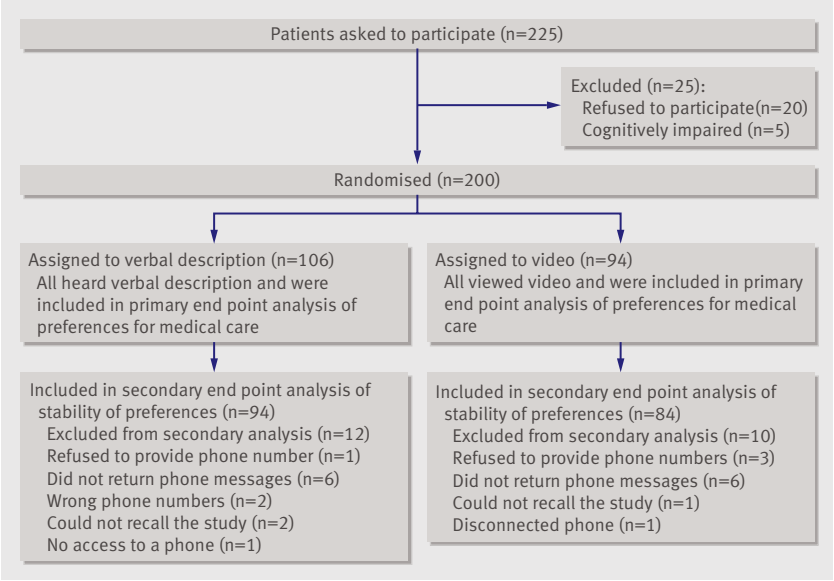


Fig 1 | Flow of participants through study

hypothesised that compared with participants randomised to a verbal description of advanced dementia, those viewing the video decision support tool after listening to a verbal description would have greater knowledge of advanced dementia, be more likely to opt for comfort oriented care that focuses on the relief of symptoms, and would be less likely to change their preferences over time. A secondary, exploratory hypothesis was that the goals of care would be predicted by health literacy.

Advanced dementia is an excellent model on which to test the hypothesis that visualising a hypothetical health state improves decision making. Advanced dementia is an ultimately fatal, progressive, neurological disease in which the median survival after the onset of symptoms is three to six years.[23 24] Patients with advanced dementia are at high risk of developing multiple yet predictable medical problems over the course of their illness, including aspiration pneumonia, pressure ulcers, and difficulty in swallowing. By virtue of their cognitive impairment, patients will seldom be able to participate in decisions about their care at the time problems develop. Healthy patients or patients in the early stages of dementia can, however, influence the treatment they will receive by exploring their goals of care with their physician. This entails deciding whether they would want specific interventions, such as cardiopulmonary resuscitation, intravenous antibiotics, or admission to hospital. This randomised controlled trial examined whether a video of a patient with advanced dementia could shape the choices made by people about the kind of care they would want in the future.

METHODS

Participants

Participants were recruited from a convenience sample of patients cared for at four primary care clinics located at three teaching hospitals in the greater Boston area. These comprised an urban geriatric clinic, a suburban geriatric clinic, an urban adult primary care clinic, and a suburban adult primary care clinic. Recruitment occurred between 1 September 2007 and 30 May 2008. Clinic staff gave all scheduled English speaking patients aged 65 or over a leaflet outlining the study after patients registered for their clinic visit, which was scheduled as part of their usual care. At the end of the visit, clinic staff asked patients if they were interested in participating in the study. If patients indicated interest, the research team initially interviewed them for eligibility. Eligibility criteria included ability to communicate in English, ability to provide informed consent, and absence of moderate or severe cognitive impairment based on a short portable mental status questionnaire (SPMSQ) score of ≥7 (scores <7 indicate moderate or severe cognitive impairment).[25]

Study design and randomisation

After we obtained informed consent, all patients who met the eligibility criteria were randomised into one of two groups: listening to a verbal narrative describing advanced dementia (control group) or listening to the same verbal narrative followed by watching a two minute video depicting a patient with advanced dementia (intervention group). We used simple randomisation based on a computer generated scheme. Individual assignments were concealed in numbered envelopes, half of which were made available to each interviewer. One randomisation list was generated for all four clinics. At the end of the trial, the randomisation order of participants was checked against the computer generated list. A trained member of the research team followed a structured script to collect data in a quiet room in the clinic area.

For both groups, the interviewer read aloud the verbal narrative describing advanced dementia (see appendix on bmj.com). This description was based on the functional assessment staging (FAST) stage 7a.[26] The FAST criteria include seven stages of dementia, with the later stages depicting more advanced disease. Stage 7 is further broken down into six substages (7a-7f). Stage 7a is generally considered the threshold for advanced dementia. The narrative states that advanced dementia is an incurable illness of the brain caused by many years of Alzheimer's disease or a series of strokes. Its salient features are the inability to communicate understandably with others, inability to walk without assistance, and inability to feed oneself.

Participants randomised to the intervention group viewed the video decision support tool on a portable computer after listening to the same verbal narrative. The two minute video depicts the principal features of advanced dementia as described in the narrative. The video presents an 80 year old female patient with advanced dementia together with her two daughters in the nursing home setting (www.bmj.com/video/care_preferences_dementia.dtl; also available at www.ACPdecisions.com). The patient fails to respond to their attempts at conversation (inability to communicate). The patient is next shown

being pushed in a wheelchair (inability to ambulate). Lastly, the patient is fed pureed food (inability to feed oneself). Before filming we obtained consent from her designated healthcare proxy to film the patient and to use the video for research purposes.

The development of the video followed a systematic approach,[27] starting with a review of the literature on dementia and advance care planning. We then used a panel of physicians with an iterative process of comments to review the design, content, and structure of the video intervention. This panel included five geriatricians and five neurologists, all of whom specialise in the care of patients with dementia.

The video was filmed without the use of prompts or stage directions to convey a candid realism.[28] The principal investigator (AEV) did all filming and editing, following previously published filming criteria.[29] The video is accompanied by the same narration that was used in the verbal description arm of the study.

Data collection and other variables

At all four study sites, two members of the research team (AEV and AEJ), who were not blinded to the randomisation group, used structured questionnaires to interview participants before and after they listened to the verbal narrative alone or listened to the narrative and watched the video. At the baseline structured interview (15 minutes) we collected demographic data and data on health status and knowledge about advanced dementia. Sociodemographic data included age, race (self reported), sex, educational status, and marital status. Health status was self rated on a Likert scale as excellent, very good, good, fair, or poor. Participants were also asked if they had had a diagnosis of dementia and whether they had known a person with advanced dementia. We assessed knowledge of advanced dementia with five true/false questions that asked whether advanced dementia is curable and if patients with advanced dementia are able to communicate with others, recognise family members, ambulate, and feed themselves. Knowledge scores therefore ranged from 0-5, with higher scores indicating better knowledge.

Participants underwent a second structured interview (15 minutes) immediately after the intervention. This included knowledge of advanced dementia, preferences for goals of care, health literacy, and, for the video group, comfort with the video decision support tool. The knowledge questions were identical to those asked in the baseline interview.

There were three options for preferences for goals of care: life prolonging care, limited care, and comfort care (see appendix on bmj.com). Researchers verbally described examples of the kinds of care implied by each goal. The first option, life prolonging care, was described as aiming to prolong life at any cost. It translates into all potentially indicated medical care that is available in a modern hospital, including cardiopulmonary resuscitation and treatment in the intensive care unit. The second option, limited care, was described as aiming to maintain physical functioning. It includes treatments such as admission to hospital, intravenous fluids, and antibiotics but not attempted cardiopulmonary resuscitation and treatment in the intensive care unit. The third option, comfort care, was described as aiming to maximise comfort and to relieve pain. Only measures that provide comfort are performed. It is compatible with oxygen and analgesics but not with intravenous treatments and admission to hospital unless necessary to provide comfort. After these explanations, participants were asked about their preferences for care if they developed advanced dementia. Participants who were unable to select a level of care were considered "uncertain."

We assessed health literacy using the rapid estimate of adult literacy in medicine tool (REALM).[30] This is a

Table 1 | Characteristics of older people living in the community randomised to verbal description and video decision support groups. Figures are numbers (percentages) of participants unless stated otherwise

Characteristics	Verbal (n=106)	Video (n=94)
Mean (SD) age (years)	75 (8)	75 (8)
Women	59 (56)	57 (61)
Race:		
Black/African-American	35 (33)	24 (26)
White	71 (67)	70 (74)
Health literacy*:		
≤6th grade (≤11 years)	19 (18)	16 (17)
7-8th grades (12-14 years)	15 (14)	9 (10)
≥9th grade (≥14 years)	72 (68)	69 (73)
Education:		
Elementary	5 (5)	6 (6)
Some high school	17 (16)	16 (17)
High school graduate	19 (18)	17 (18)
Some college	19 (18)	17 (18)
College graduate	16 (15)	14 (15)
Postgraduate or professional	29 (27)	24 (26)
Refused to say	1 (1)	0
Marital status:		
Married	43 (41)	42 (45)
Widowed	25 (24)	29 (31)
Divorced	21 (20)	13 (14)
Never married	17 (15)	10 (10)
Self reported health status:		
Excellent	15 (14)	9 (10)
Very good	23 (22)	37 (39)
Good	27 (26)	26 (28)
Fair	30 (28)	19 (20)
Poor	9 (8)	2 (2)
Refused to say	2 (2)	1 (1)
Diagnosis of dementia†:	12 (11)	6 (6)
Previous relationship with person with advanced dementia	11 (10)	18 (19)
Knowledge score before randomisation‡	2.3	2.1

*Assessed with rapid estimate of adult literacy in medicine (REALM).
†Participants were asked if they had diagnosis of dementia.
‡Knowledge score calculated by adding responses to five questions that test respondent's knowledge of advanced dementia. Each question has possible response of 1 (correct) or 0 (incorrect or unsure). Total knowledge score ranges from 0-5, with higher scores indicating greater knowledge.

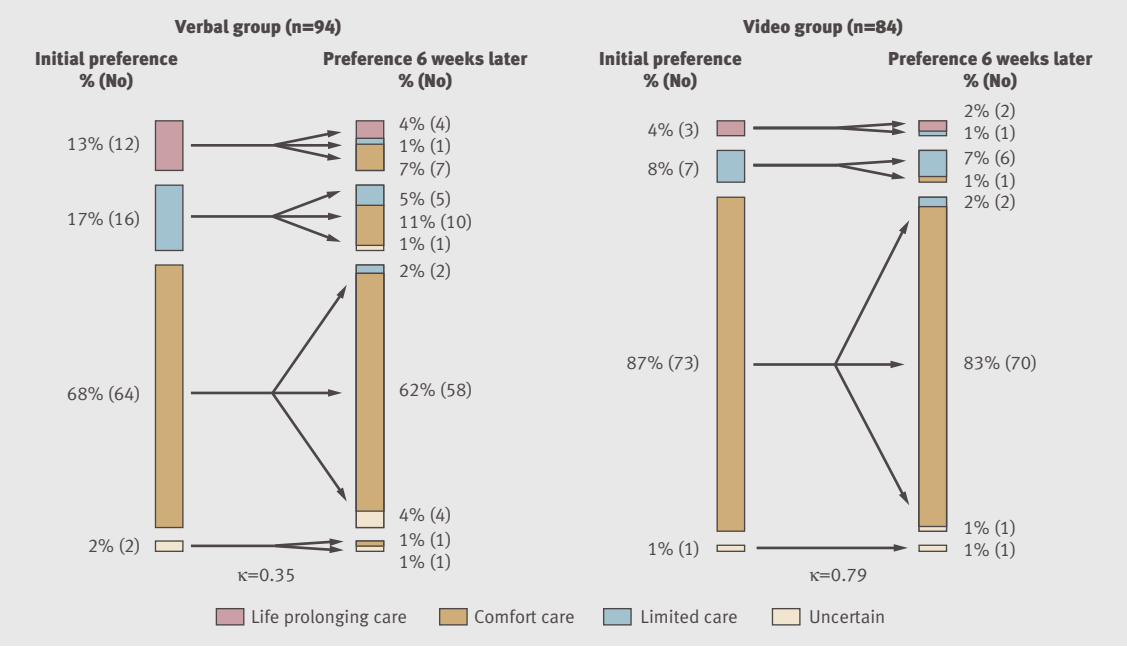


Fig 2 | **Initial preferences and stability of preferences after six weeks**

two to three minute English test of medically relevant vocabulary. It is a validated test of word pronunciation and has been shown to correlate well with tests that evaluate a range of literacy skills.[30] As others have done, we defined three categories for health literacy based on the REALM scores: 6th grade and below (up to age 11; score 0-45); 7-8th grade (ages 12-13; 45-60); and 9th grade and above (age 14 and over; 61-66).[31 32]

For those participants randomised to the video intervention group, we used a four point Likert scale to assess the perceived value of the video by asking participants whether they had a better understanding of the disease after viewing the video, if they were comfortable watching the video, if they would recommend the video to others, and whether they thought videos would be helpful for eliciting preferences for care in other diseases like cancer.

One interviewer (AEV) contacted participants by telephone six weeks after the initial interview to determine again what their preferences would be if they had advanced dementia in exactly the same manner as the initial interview. We chose a follow-up period of six weeks to ensure that an adequate amount of time elapsed from exposure to the intervention and to assess whether the video had an enduring effect.

Statistical analysis

Our analyses were based on the decision making group to which participants were randomised. The primary outcome measure was preferences for care if they developed advanced dementia categorised as four options (life prolonging, limited, comfort, or uncertain). Additional outcomes included change in knowledge scores before and after the intervention and the stability of preferences after six weeks.

All characteristics of participants and outcomes were described by using proportions for categorical variables and means (SD) for continuous variables. We used χ^2 tests to compare preferences for care (life prolonging, limited, comfort, or uncertain) between the two groups.

Two sample *t* tests compared change in knowledge scores before and after the intervention between the two groups. We used κ statistics to summarise the stability of preferences six weeks after the clinic interview for each group and compared the proportions who changed preferences with Pearson χ^2 exact test between the two groups.

The measure for the primary outcome analysis was the unadjusted difference in proportions of participants preferring comfort care between the two study groups. We conducted secondary analyses to identify factors associated with a preference for comfort care among all participants. Bivariate analyses determined the association between individual characteristics of participants (age, sex, race, education, marital status, health status, personal history of dementia, previous relationship with a person with advanced dementia, health literacy, and randomisation group) and a preference for comfort care with Fisher's exact test. Multivariable logistic regression analyses were used to identify factors independently associated with preferences for comfort care. Factors significant at 0.10 in the bivariate analyses were entered into a stepwise algorithm, retaining factors in the model that were significant at the 0.05 level. We used the variance inflation factor to diagnose colinearity among potential predictors.

All reported P values are two sided, with $P<0.05$ considered as significant. The study was designed to detect a 25% difference in the proportion of participants

choosing comfort care between the two groups, assuming the rate in the verbal group was 60%. With a target of 100 patients in each group, the power of the study was estimated to be >90%. Data were analysed with SAS software, version 9.1 (SAS Institute, Cary, NC).

RESULTS

Participant flow

We approached 225 consecutive and potentially eligible patients, of whom 205 (92%) agreed to be interviewed. Patients who declined did not differ significantly from the recruited participants in terms of age, sex, or race. The most common reason given for not participating was lack of time. Of the 205 recruited for the study, five were disqualified because their mental status questionnaire score was <7, resulting in a total of 200 study participants. Of these, 106 were randomised to the control group and 94 to the video intervention group (fig 1). Table 1 shows the baseline characteristics. Despite the randomisation process there were some baseline differences in the two groups, including diagnosis of dementia and previous relationship with someone with dementia,.

Outcomes

Among the 106 participants receiving only the verbal narrative, 68 (64%) chose comfort care, 20 (19%) chose limited care, 15 (14%) chose life prolonging care, and three (3%) were uncertain of their preferences. Among the 94 who also saw the video, 81 (86%) chose comfort care, eight (9%) chose limited care, four (4%) chose life prolonging care, and one (1%) was uncertain of her preferences (χ^2=13.0, df=3; P=0.003). Thus a significantly greater proportion of participants in the video

Table 2 | Unadjusted differences in proportions and multivariable analyses of associations with likelihood of choosing comfort care as primary goal of care

Characteristics	Frequency choosing comfort care	Difference in % choosing comfort care (95% CI)	Unadjusted P value	Odds ratio (95% CI)	
				Unadjusted	Adjusted
Age (years):					
<80†	104 (71)	12% (−0.3% to 25%)	0.10	2.0 (0.9 to 4.5)	—
≥80	45 (83)				
Sex:					
Female†	88 (76)	−3% (−16% to 9%)	0.62	0.8 (0.4 to 1.6)	—
Male	61 (73)				
Education:					
<College graduate†	79 (68)	15% (3% to 27%)	0.021	2.3 (1.2 to 4.6)	—
≥College graduate	69 (83)				
Marital status:					
Not married†	82 (71)	8% (−4% to 20%)	0.25	1.5 (0.8 to 2.9)	—
Married	67 (79)				
Health status‡:					
Fair or poor†	35 (58)	23% (9% to 37%)	0.001	3.0 (1.6 to 6.0)	—
Good or better	111 (81)				
Diagnosis of dementia:					
No†	136 (75)	−3% (−25% to 19%)	1.0	0.9 (0.3 to 2.5)	—
Yes	13 (72)				
Previous relationship with person with advanced dementia:					
No†	125 (73)	10% (−6% to 25%)	0.36	1.8 (0.6 to 4.9)	—
Yes	24 (83)				
Randomisation:					
Verbal†	68 (64)	22% (11% to 34%)	<0.001	3.5 (1.7 to 7.1)	3.9 (1.8 to 8.6)
Video	81 (86)				
Health literacy§:					
≤6th grade†	16 (46)				
7-8th grades	14 (58)	13% (−13% to 38%)		1.7 (0.6 to 4.7)	1.7 (0.54 to 5.3)
≥9th grade	119 (84)	39% (21% to 56%)	<0.001	6.4 (2.9 to 14.4)	4.1 (1.6 to 10.8)
Race:					
Black/ African-American†	30 (51)	34% (19% to 48%)	<0.001	5.2 (2.6 to 10.4)	2.9 (1.3 to 6.6)
White	119 (84)				

*For multivariable analysis, characteristics excluded from model if they were not related to outcome—that is, choosing comfort care—at P<0.05.
†Reference category.
‡Health status was one of excellent, very good, good, fair, or poor.
§Assessed with rapid estimate of adult literacy in medicine (REALM).

WHAT IS ALREADY KNOWN ON THIS TOPIC

Advance care planning is a complex process involving communication of future health states such as advanced dementia

Visual images might be helpful to improve decision making and communication of complex information regarding what type of medical care patients would want at the end of life

WHAT THIS STUDY ADDS

Video images of advanced dementia improved knowledge for patients choosing the type of medical care they would like if they developed advanced dementia

Patients who viewed a video decision support tool of advanced dementia after hearing a verbal description were more likely to choose a comfort oriented approach compared with patients solely listening to a verbal narrative of the disease

Patients using the video decision support tool had more stable preferences for end of life care over time

Video decision support tools might be most useful for patients with low health literacy

group opted for comfort care (difference 22%, 95% confidence interval 11% to 34%).

Mean knowledge scores (range 0-5) were significantly higher in the video group than in the control group (4.5 (SD 1.0) *v* 3.8 (SD 1.3), respectively; P<0.001). The mean increase in knowledge scores for the video group was 2.4 (2.1 to 2.7) and 1.5 (1.2 to 1.9) for the control group, which was significant (P<0.001).

Table 2 shows the unadjusted differences in proportions of participants and odds ratios preferring comfort care for each of the characteristics. The factors associated with a greater likelihood of preferring comfort care among all participants were being a college graduate or higher, good or better health status, greater health literacy, white race, and randomisation to the video group. The first four factors were highly correlated: those with higher degree of education were more likely to have better health status, greater health literacy, and more likely to be white; those with better health status were more likely to have greater health literacy and were more likely to be white; and those with greater health literacy were more likely to be white (all with P<0.05). The variance inflation factors, however, were all less than 2.5 when we tested these four factors in the regression model, which indicated weak evidence of multicollinearity. After inclusion of these variables in a multivariable logistic regression model, participants randomised to the intervention group had a greater likelihood of opting for comfort care (adjusted odds ratio 3.9, 1.8 to 8.6). Other factors independently associated with opting for comfort care included a health literacy level of greater than 9th grade (4.1, 1.6 to 10.8) and white race (2.9, 1.3 to 6.6) (table 2).

Six weeks after the initial clinic visit, we attempted to contact each participant by telephone. Among the 94 (89%) in the control group who could be contacted, 27 (29%) changed their preferences; the κ statistic for preference stability was 0.35 (0.15 to 0.54) (fig 2). Among the 84 (89%) participants contacted in the video group, five (6%) changed their preferences; the κ statistic for preference stability was 0.79 (0.62 to 0.98). After six weeks, the proportion of participants changing preferences was lower in the video group (P<0.001).

The video decision support tool was highly acceptable to participants: 83 of 94 (88%) found the video "very helpful" or "somewhat helpful"; 80 (85%) said they were "very comfortable" or "somewhat comfortable" viewing the video; 89 (95%) said they would "definitely" or "probably" recommend the video to others; and 78 (83%) thought that using videos for other diseases (such as cancer) would be "very helpful" or "somewhat helpful." There were no adverse events in either group.

DISCUSSION

Principal findings

When presented with the possibility of developing advanced dementia, older patients living in the community are more likely to choose comfort as the primary goal of care after viewing a video of a patient with the disease and listening to a verbal description rather than just hearing a verbal description of advanced dementia. Moreover, viewing the video improved knowledge of advanced dementia and enhanced stability of preferences for treatment over time compared with hearing only the verbal narrative. Finally, health literacy seems to be associated with end of life preferences among older patients.

Comparison with other studies

To the best of our knowledge, this study represents the first randomised controlled trial in a group of older patients of a video decision support tool for decision making at the end of life. In our previous before and after investigation of the advanced dementia video conducted in healthy middle aged participants, the video promoted preferences for comfort care, but it was not a randomised trial, was conducted with a younger healthy cohort, and did not follow the stability of preferences over time.[20] Our current study extends this earlier work by showing the efficacy of the video in a randomised controlled trial among older patients. Moreover, the participants in the video group were more likely to have improved knowledge after the video and stable preferences over time. The stability of preferences is a critical consideration in evaluating preferences at the end of life[33-40] and suggests a more accurate reflection of patients' values and wishes.

Strengths and limitations

Our study has several important limitations and numerous strengths. Firstly, the research staff collecting data at baseline and at the immediate and six week follow-up interviews were not blinded to randomisation, which could have introduced bias into our findings. Previous randomised studies of interventions aimed at improving end of life decision making, however, have seldom been blinded because limiting the number of interviewers eases the burden on participants of addressing difficult and often painful subject matter.[8-10] Furthermore, participants might disclose whether they viewed the video or not. We attempted

to reduce the influence of this potential bias by using structured interviews and outcome measures. Secondly, despite randomisation there were some baseline differences between the two groups. This can be expected in a relatively small sample. Thirdly, videos can be manipulated to favour a particular perspective. Our study used one video of a white woman with dementia. We did not assess responses of participants to videos of people of different sex and race. Fourthly, we asked participants for their preferences in the context of a research study. The next step would be to investigate whether patients and physicians would document their preferences in the medical record or complete an advance directive.[41] Finally, our sample was primarily white and African-American and drawn from primary care clinics in teaching hospitals in metropolitan Boston. Thus, our findings might not be generalisable to other minority groups (such as Latinos and Asian-Americans).

Policy implications and future research

Previous uses of video decision support tools have primarily focused on helping patients to make treatment or screening decisions.[42] Our use of video redirects attention to the underlying health state by clarifying the nature of the condition about which patients are expected to make decisions. Our use of video portrays the illness to add a sense of verisimilitude that might be lacking in verbal descriptions. Moreover, these images might offer a more objective and straightforward approach to describe complex medical conditions, which is particularly pertinent to patients with low health literacy. In the US such patients are more likely to be elderly and African-American[31] and are among the most vulnerable populations in our healthcare system. As the video led to better knowledge of advanced dementia, our study supports the claim made by others that pictorial or visual methods improve decision making processes.[16-22]

Previous studies have suggested that non-white people receive and opt for more aggressive end of life care.[43-49] The reason for this observation is not well elucidated but might be, in part, because of variation in the quality of counselling they receive and their understanding of that counselling. As we have shown elsewhere,[22] our study lends additional support to the notion that health literacy potentially mediates the role of race in end of life decision making, and video decision support tools offer an approach to circumvent this disparity. Future work is needed to explore this finding as health literacy was highly correlated with other variables and our study lacked adequate power to conduct detailed analyses of mediation.

The next step in using videos is to explore other diseases and the goals of care with video portrayals. We suspect that numerous other diseases and interventions, such as advanced cancer and cardiopulmonary resuscitation, would also be more accurately conveyed to patients through a visual medium than solely by verbal descriptions. As we have shown here and elsewhere,[29] criteria regarding the necessary content and editing of each video portrayal must be carefully considered before clinical application of these videos.

Active debate exists surrounding the development of decision support technologies, especially when highly subjective content (patients' narratives and testimonials) and non-traditional media (video) are used.[50,51] While important steps have been taken to develop objective criteria for the development and field testing of decision support tools,[52,53] more research is needed particularly as they apply to the use of video.

Including patients in the decision making process has been an important yet complex advance in modern medical care. To secure the delivery of high quality end of life care, patients must be informed regarding their decision making. Education of patients using video decision support tools can improve their comprehension of disease states such as advanced dementia that are difficult to envision solely with words. Future work could extend the use of video decision support tools to other disease states such as advanced cancer and the goals of care. We have shown that video decision support tools enhance elderly patients' decision making by ensuring that it is both more informed and consistent over time.

We thank the patient and her family, who agreed to be filmed for the purposes of educating and improving decision making at the end of life for all patients. We also thank Claus Hamann, Katherine Hesse, Juergen Bludau, Laurel Yates, Peggy Edson, Ruth Kandel, and Aretha Delight Davis for their contribution to the project.

Contributors: AEV, MKP-O, MJB, MRG, EFC, EDA, and SLM were responsible for study concept and design. AEV and AE-J acquired the data, which was analysed and interpreted by AEV, MKP-O, YC, and SLM. AEV, MKP-O, MJB, MRG, EDA, and SLM drafted the manuscript. AEV, MKP-O, MJB, MRG, KLM, EFC, EDA, AE-J, and SLM supervised the study and critically revised the manuscript for important intellectual content. YC did the statistical analysis. AEV, MJB, MRG, and SLM obtained funding. AEV, MKP-O, MJB, KLM, EFC, EDA, AE-J, and SLM were responsible for administrative, technical, and material support. AEV is guarantor.

Funding: AEV was supported by a George Bennett Fellowship from the Foundation for Informed Medical Decision Making, a New Investigator Research Grant from the Alzheimer's Association, and a Center for Excellence Career Development Award from the Hartford Foundation. Role of the sponsor: None of the foundations participated in the collection, analysis, or interpretation of the data or in preparation, review, or approval of the manuscript.

Competing interests: None declared.

Ethical approval: The protocol was approved by the institutional review board of each participating institution and all participants provided informed consent

1 Lo B, Snyder L. Care at the end of life: guiding practice where there are no easy answers. *Ann Intern Med* 1999;130:772-4.
2 Council on Scientific Affairs. Good care of the dying patient. American Medical Association. *JAMA* 1996;275:474-8.
3 Lynn J. Measuring quality of care at the end of life: a statement of principles. *J Am Geriatr Soc* 1997;45:526-7.
4 Sachs GA, Ahronheim JC, Rhymes JA, Volicer L, Lynn J. Good care of dying patients: the alternative to physician-assisted suicide and euthanasia. *J Am Geriatr Soc* 1995;43:553-62.
5 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.
6 Gillick MR. A broader role for advance medical planning. *Ann Intern Med* 1995;123:621-4.
7 Finucane TE, Shumway JM, Powers RL, D'Alessandri RM. Planning with elderly outpatients for contingencies of severe illness: a survey and clinical trial. *J Gen Intern Med* 1988;3:322-5.

8 SUPPORT. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators. *JAMA* 1995;274:1591-8.
9 Hofmann JC, Wenger NS, Davis RB, Teno J, Connors AF Jr, Desbiens N, et al. Patient preferences for communication with physicians about end-of-life decisions. SUPPORT Investigators. Study to Understand Prognoses and Preference for Outcomes and Risks of Treatment. *Ann Intern Med* 1997;127:1-12.
10 Covinsky KE, Fuller JD, Yaffe K, Johnston CB, Hamel MB, Lynn J, et al. Communication and decision-making in seriously ill patients: findings of the SUPPORT project. The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. *J Am Geriatr Soc* 2000;48(5 suppl):S187-93.
11 Tulsky JA, Fischer GS, Rose MR, Arnold RM. Opening the black box: how do physicians communicate about advance directives? *Ann Intern Med* 1998;129:441-9.
12 Roter DL, Larson S, Fischer GS, Arnold RM, Tulsky JA. Experts practice what they preach: a descriptive study of best and normative practices in end-of-life discussions. *Arch Intern Med* 2000;160:3477-85.
13 Smith AK, Ries AP, Zhang B, Tulsky JA, Prigerson HG, Block SD. Resident approaches to advance care planning on the day of hospital admission. *Arch Intern Med* 2006;166:1597-602.
14 Tulsky JA, Chesney MA, Lo B. See one, do one, teach one? House staff experience discussing do-not-resuscitate orders. *Arch Intern Med* 1996;156:1285-9.
15 Tulsky JA, Chesney MA, Lo B. How do medical residents discuss resuscitation with patients? *J Gen Intern Med* 1995;10:436-42.
16 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.
17 Morgan MW, Deber RB, Llewellyn-Thomas HA, Gladstone P, Cusimano RJ, O'Rourke K, 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.
18 Edwards A, Elwyn G, Mulley A. Explaining risks: turning numerical data into meaningful pictures. *BMJ* 2002;324:827-30.
19 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.
20 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.
21 Volandes AE, Ariza M, Abbo ED, Paasche-Orlow M. Overcoming educational barriers for advance care planning in latinos with video images. *J Palliat Med* 2008;11:700-6.
22 Volandes AE, Paasche-Orlow M, Gillick MR, Cook EF, Shaykevich S, Abbo ED, et al. Health literacy not race predicts end-of-life care preferences. *J Palliat Med* 2008;11:754-62.
23 Wolfson C, Wolfson DB, Asgharian M, M'Lan CE, Ostbye T, Rockwood K, et al. A reevaluation of the duration of survival after the onset of dementia. *N Engl J Med* 2001;344:1111-6.
24 Larson EB, Shadlen MF, Wang L, McCormick WC, Bowen JD, Teri L, et al. Survival after initial diagnosis of Alzheimer disease. *Ann Intern Med* 2004;140:501-9.
25 Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc* 1975;23:433-41.
26 Sclan SG, Reisberg B. Functional assessment staging (FAST) in Alzheimer's disease: reliability, validity, and ordinality. *Int Psychogeriatr* 1992;4(suppl 1):55-69.
27 National Cancer Institute. *Making health communication programs work: a planner's guide*. Bethesda, MD: US National Institutes of Health, 1997. www.cancer.gov/pinkbook.
28 Grant BK, Sloniowski J. *Documenting the documentary: close readings of documentary film and video*. Detroit: Wayne State University Press, 1998.
29 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, NY: Nova Science Publishers, 2007:193-206.
30 Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Fam Med* 1993;25:391-5.
31 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.
32 Davis TC, Wolf MS, Bass PF 3rd, Thompson JA, Tilson HH, Neuberger M, et al. Literacy and misunderstanding prescription drug labels. *Ann Intern Med* 2006;145:887-94.
33 Everhart MA, Pearlman RA. Stability of patient preferences regarding life-sustaining treatments. *Chest* 1990;97:159-64.
34 Danis M, Garrett J, Harris R, Patrick DL. Stability of choices about life-sustaining treatments. *Ann Intern Med* 1994;120:567-73.
35 Emanuel LL, Emanuel EJ, Stoeckle JD, Hummel LR, Barry MJ. Advance directives. Stability of patients' treatment choices. *Arch Intern Med* 1994;154:209-17.
36 Carmel S, Mutran EJ. Stability of elderly persons' expressed preferences regarding the use of life-sustaining treatments. *Soc Sci Med* 1999;49:303-11.
37 Rosenfeld KE, Wenger NS, Phillips RS, Connors AF, Dawson NV, Layde P, et al. Factors associated with change in resuscitation preference of seriously ill patients. The SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. *Arch Intern Med* 1996;156:1558-64.
38 Ditto PH, Smucker WD, Danks JH, Jacobson JA, Houts RM, Fagerlin A, et al. Stability of older adults' preferences for life-sustaining medical treatment. *Health Psychol* 2003;22:605-15.
39 Straton JB, Wang NY, Meoni LA, Ford DE, Klag MJ, Casarett D, et al. Physical functioning, depression, and preferences for treatment at the end of life: the Johns Hopkins Precursors Study. *J Am Geriatr Soc* 2004;52:577-82.
40 Lockhart LK, Ditto PH, Danks JH, Coppola KM, Smucker WD. The stability of older adults' judgments of fates better and worse than death. *Death Stud* 2001;25:299-317.
41 Abbo ED, Sobotka S, Meltzer DO. Patient preferences in instructional advance directives. *J Palliat Med* 2008;11:555-62.
42 O'Connor AM, Stacey D, Rovner D, Holmes-Rovner M, Tetroe J, Llewellyn-Thomas H, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* 2001;(3):CD001431.
43 Garrett JM, Harris RP, Norburn JK, Patrick DL, Danis M. Life-sustaining treatments during terminal illness: who wants what? *J Gen Intern Med* 1993;8:361-8.
44 Caralis PV, Davis B, Wright K, Marcial E. The influence of ethnicity and race on attitudes toward advance directives, life-prolonging treatments, and euthanasia. *J Clin Ethics* 1993;4:155-65.
45 O'Brien LA, Grisso JA, Maislin G, LaPann K, Krotki KP, Greco PJ, et al. Nursing home residents' preferences for life-sustaining treatments. *JAMA* 1995;274:1775-9.
46 Blackhall LJ, Frank G, Murphy ST, Michel V, Palmer JM, Azen SP. Ethnicity and attitudes towards life sustaining technology. *Soc Sci Med* 1999;48:1779-89.
47 Crawley L, Payne R, Bolden J, Payne T, Washington P, Williams S. Palliative and end-of-life care in the African American community. *JAMA* 2000;284:2518-21.
48 Hopp FP, Duffy SA. Racial variations in end-of-life care. *J Am Geriatr Soc* 2000;48:658-63.
49 Klessig J. The effect of values and culture on life-support decisions. *West J Med* 1992;157:316-22.
50 Khangura S, Bennett C, Stacey D, O'Connor AM. Personal stories in publicly available patient decision aids. *Patient Educ Couns* 2008;73:456-64.
51 Winterbottom A, Bekker HL, Conner M, Mooney A. Does narrative information bias individual's decision making? A systematic review. *Soc Sci Med* 2008;67:2079-88.
52 Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ* 2006;333:417.
53 Evans R, Elwyn G, Edwards A, Watson E, Austoker J, Grol R. Toward a model for field-testing patient decision-support technologies: a qualitative field-testing study. *J Med Internet Res* 2007;9:e21.

Accepted: 3 February 2009

Are Opioid Dependence and Methadone Maintenance Treatment (MMT) Documented in the Medical Record? A Patient Safety Issue

Alexander Y. Walley, MD, MSc[1,2,3], Danielle Farrar, MMS[2], Debbie M. Cheng, ScD[1,4], Daniel P. Alford, MD, MPH[1,2], and Jeffrey H. Samet, MD, MPH, MA[1,2,3,5]

[1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center, Boston, MA, USA; [2]Boston University School of Medicine, Boston, MA, USA; [3]Substance Abuse Prevention and Treatment Services, Boston Public Health Commission, Boston, MA, USA; [4]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; [5]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA.

BACKGROUND: Opioid-dependent patients often have co-occurring chronic illnesses requiring medications that interact with methadone. Methadone maintenance treatment (MMT) is typically provided separately from medical care. Hence, coordination of medical care and substance use treatment is important to preserve patient safety.

OBJECTIVE: To identify potential safety risks among MMT patients engaged in medical care by evaluating the frequency that opioid dependence and MMT documentation are missing in medical records and characterizing potential medication-methadone interactions.

METHODS: Among patients from a methadone clinic who received primary care from an affiliated, but separate, medical center, we reviewed electronic medical records for documentation of methadone, opioid dependence, and potential drug-methadone interactions. The proportions of medical records without opioid dependence and methadone documentation were estimated and potential medication-methadone interactions were identified.

RESULTS: Among the study subjects (n=84), opioid dependence documentation was missing from the medical record in 30% (95% CI, 20%–41%) and MMT documentation was missing from either the last primary care note or the last hospital discharge summary in 11% (95% CI, 5%-19%). Sixty-nine percent of the study subjects had at least 1 medication that potentially interacted with methadone; 19% had 3 or more potentially interacting medications.

CONCLUSION: Among patients receiving MMT and medical care at different sites, documentation of opioid dependence and MMT in the medical record occurs for the majority, but is missing in a substantial number of patients. Most of these patients are prescribed medications that potentially interact with methadone. This study highlights opportunities for improved coordination between medical care and MMT.

Received December 12, 2008
Revised May 11, 2009
Accepted June 4, 2009
Published online July 4, 2009

KEY WORDS: methadone; medication interactions; patient safety; care coordination.

J Gen Intern Med 24(9):1007–11
DOI: 10.1007/s11606-009-1043-4

INTRODUCTION

Both the Institute of Medicine and the Joint Commission have focused attention on the role that medication discrepancies and errors play in adverse events and decreased quality of care.[1,2] For hospital, ambulatory, and behavioral healthcare, a Joint Commission goal is to "accurately and completely reconcile medications across the continuum of care." The Institute of Medicine specifically called the improvement of communication between medical and substance use treatment providers fundamental to improving the quality of healthcare for patients with mental and substance-use conditions.[3] Methadone maintenance treatment (MMT) for approximately 260,000 opioid-dependent patients in the United States is restricted to federal- and state-regulated clinics.[4] Typically, MMT clinics are in locations that are separate from general medical care with increased confidentiality protections required for substance use related health information,[5,6] which can be a barrier to the coordination of care.

MMT is a chronic therapy for opioid dependence, a chronic relapsing disease that often requires lifelong treatment.[7] Common co-occurring conditions among opioid-dependent patients receiving MMT, that also require chronic pharmacotherapy include HIV infection,[8] mood disorders,[9–11] chronic pain,[12] osteoporosis,[13] and diabetes.[14–16] Many medications for these conditions interact with methadone in potentially important ways, including prolongation of the QT interval resulting in increased risk of torsade de pointes,[17–21] increased or decreased metabolism of methadone or the interacting medication by the cytochrome P450 enzyme system, and increased sedation by other opioids and benzodiazepines.[22–24] Thus, medication-methadone interactions potentially contribute to clinically significant adverse events, including cardiac arrhythmias, overdoses, decreased cognitive function, opioid withdrawal symptoms, and relapses to illicit opioid use. Ideally, when patients on MMT engage in outpatient or inpatient medical care, their treating physicians are aware of the MMT

and document methadone on the medication list in addition to opioid dependence on the medical problem list.

In this study of MMT patients receiving medical care, our objectives were to 1) quantify the documentation of opioid dependence in the medical record, 2) quantify the documentation of MMT in the medical record, and 3) describe the possible drug interactions that could arise from the use of methadone with other prescribed medications.

METHODS

Study Design and Population

We reviewed electronic medical records (EMRs) of patients from the Boston Public Health Commission (BPHC) MMT program who received their primary medical care at Boston Medical Center (BMC), an affiliated, but separate, medical center. Entry criteria included the following: 1) enrollment in methadone treatment on or before July 2, 2007; 2) a signed release of information permitting 2-way communication between the MMT program and primary care physician; 3) a primary care physician based at BMC; and 4) at least 1 primary care progress note or 1 discharge summary in the BMC EMR between September 1, 2002, and July 2, 2007, and during the period of MMT. This study was approved by the institutional review board of Boston University Medical Campus.

Data Collection and Measures

Methadone dose and "take home" dose information were collected from the MMT EMR for the date immediately preceding the medical center inpatient admission or primary care visit. "Take home" doses refer to unobserved methadone dosing, a privilege given only to patients who have at least 90 days of abstinence from illicit drug use. To assess whether our sample was similar to the MMT program population, we obtained information on age, gender, race/ethnicity, methadone dose, and number of days on methadone as of July 2, 2007, from the MMT EMR for all patients.

From the medical center EMR, we determined insurance status as public (e.g., Medicare or Medicaid), private, or no insurance. To gauge the burden of comorbidity, we searched the active problem list, which is edited and maintained by treating providers, for chronic conditions (e.g., HIV/AIDS), typically treated with medications that potentially interact with methadone, and substance-related conditions (e.g., alcohol dependence).

To quantify both MMT and opioid-dependence diagnosis documentation, the most recent discharge summary and primary care note in the medical center EMR between September 1, 2002, and July 2, 2007, were read and searched by one of the authors (DF). Missing documentation of MMT was defined as no listing of methadone as a medication in either the primary care progress note or the discharge summary. Subjects without any combination of "heroin, opioid, or opiate" and "use, abuse, or dependence" in the discharge summary, primary care note, or the problem list were categorized as not having documentation of opioid dependence. Both the problem list and the medication lists were cumulative, based on previous treatment episodes, and editable by each provider.

To describe possible methadone-drug interactions, 8 medication interaction categories were pre-specified prior to reviewing records and included: 1) may decrease methadone effects (e.g., efavirenz, phenytoin); 2) may increase methadone effects (e.g., fluconazole, fluvoxamine); 3) has altered metabolism with methadone (e.g., zidovudine); 4) benzodiazepines (e.g., clonazepam); 5) other opioids (e.g., oxycodone); and 3 categories of QT interval prolonging medications; 6) risk (generally accepted risk of causing torsade de pointes); 7) possible risk (associated with torsade de pointes and/or QT prolongation but lack substantial evidence); and 8) conditional risk (weakly associated with torsade de pointes and/or QT prolongation).

The first 3 categories were determined based on a published review of methadone-medication interactions.[22] We created categories for benzodiazepines and other opioids because of the risks of oversedation when combined with methadone. We adopted the 3 categories of QT prolonging medications at www.qtdrugs.org to identify medications that increase the risk of torsade de pointes. www.qtdrugs.org is a website funded by the federal Agency for Healthcare Research and Quality.

Analysis

Descriptive statistics were obtained for all subject characteristics to describe the study sample. We determined the proportions with lack of opioid dependence and MMT documentation and the exact binomial 95% confidence intervals for these proportions. All analyses were performed using SAS software version 9.1 (SAS Institute, Cary, NC).

RESULTS

Among the 350 patients enrolled in MMT on July 2, 2007, 157 (45%) had no release of information signed for communication with a primary care physician and 86 (25%) had an existing release of information but not for BMC. Among the 107 patients with an existing release of information with a BMC primary care physician, 84 had at least 1 inpatient or primary care visit before July 2, 2007, and during the period of MMT and comprised the study sample. (Fig. 1)

Characteristics of the study sample (n=84) were mean age of 43.4 years; mean methadone dose of 85 mg; mean MMT duration of 4.4 years; 56% women; and 50% white, 30% African-American, and 20% Hispanic (Table 1). These characteristics were similar to the summary statistics available for

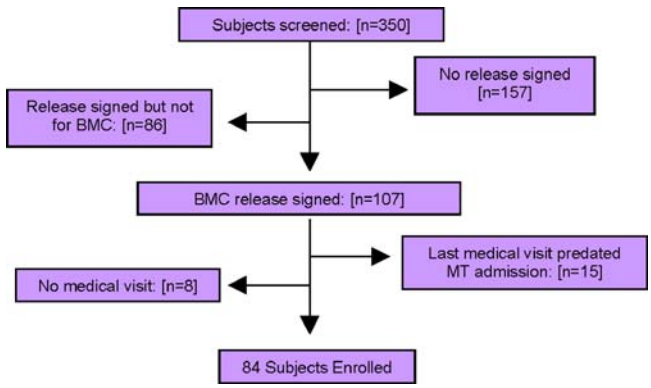


Figure 1. Study sample selection of methadone maintenance treatment patients engaged in medical care. BMC Boston Medical Center, MT methadone treatment.

Table 1. Demographic and Clinical Characteristics of Methadone Maintenance Treatment Subjects Engaged in Medical Care (n=84)

	Mean	(Standard deviation)
Age, years	43.4	(10.8)
Years on MMT at last medical visit	4.4	(4.6)
Milligrams of methadone at last medical visit	85.3	(36.1)
	N	(%)
Female	47	(56)
Race/Ethnicity		
African-American	25	(30)
White	42	(50)
Hispanic	17	(20)
Insurance status		
Public insurance	72	(86)
Private insurance	7	(8)
Free care/ Uninsured	5	(6)
On take homes at time of last medical visit	25	(30)
Medical problems		
Hepatitis C	50	(60)
Pain condition	37	(44)
Depressive disorder	36	(43)
Anxiety disorder	29	(35)
Tobacco use	26	(31)
HIV-infection	19	(23)
Hypertension	18	(21)
Diabetes	11	(13)
Alcohol abuse or dependence	6	(7)
Cocaine use, abuse, or dependence	6	(7)
Renal insufficiency	2	(2)
Seizure disorder	1	(1)

all clinic patients (n=350) at the time of the chart review (i.e., mean age of 39.7 years, mean methadone dose of 81 mg, MMT duration of 3.7 years, 62% women, 57% white, 24% African-American, 17% Hispanic, and 2% other).

Documentation of opioid dependence diagnosis was missing from the medical record in 30% (95% CI, 20%-41%) of subjects. (Table 2) Documentation of MMT was missing from either the last discharge summary or last primary care note in 11% (95% CI, 5%-19%) of subjects. Among the 82 (98%) subjects with a primary care note, documentation of MMT was missing in 7% (95% CI, 3%–15%). Among 41 (49%) subjects with a discharge summary, documentation of MMT was missing in 10% (95% CI, 3%–23%). Among 39 subjects with both a discharge summary and primary care note, documentation of MMT was missing from both notes for 1 subject or 3% (95% CI, 0.1%-13%).

Table 2. Proportions Without Documentation of Opioid Diagnoses and Methadone Maintenance Treatment in Medical Record

	No.	%	95% CI
Overall			
Opioid dependence diagnosis missing from medical record (n=84)	25	30%	20%-41%
Methadone as a medication missing from medical record* (n=84)	9	11%	5%-19%
Subgroups			
Methadone as a medication missing from:			
Last primary care (PC) note (n=82)	6	7%	3%-15%
Last discharge summary (n=41)	4	10%	3%-23%
Both PC note and discharge summary (n=39)	1	3%	0.1%-13%

Methadone not documented in either last primary care note or last discharge summary
CI confidence interval

At least 1 potential methadone interaction was identified in 58 (69%) of subjects; 16 (19%) of subjects had 3 or more such medications with potential interactions. (Figure 2) The frequency of specific interaction types ranged from 9 (11%) for "may decrease methadone effects" to 33 (39%) for potentially QT-interval prolonging effect (Figure 3). For medications within the QT-prolonging subcategories, there were 13 (15%) subjects in the "possible risk" category and 20 (24%) in the "conditional risk" category. No subjects were on any medication, other than methadone, in the "risk" category.

Among the 9 subjects who had MMT documentation missing from either note, 7 (78%) had at least 1 medication that potentially interacted with methadone. Notably, 5 (56%) subjects were on other opioids, the most common interaction category among these subjects.

DISCUSSION

This study highlights two aspects in which care coordination between medical and substance use treatment providers could impact patient safety. First, even when consent for communication between substance-use treatment providers and medical providers exists, medical providers do not always document that a patient has opioid dependence or that a patient is receiving MMT. Second, most patients engaged in medical care and MMT are on 1 or more medications that potentially interact with methadone and may lead to adverse events, such as oversedation, overdose, opioid withdrawal, or cardiac dysrythmias.

This study sample likely represented a "best case" for thorough documentation, because we selected methadone clinic subjects with known primary care physicians at a medical center with comprehensive services, a standardized EMR, and an historical affiliation with the methadone clinic. The coordination problems identified in this study likely loom larger among patients receiving both MMT and primary medical care from unaffiliated providers. Limited communication between separate systems of medical care and MMT was previously documented in a Spanish study in which 89.5% of primary care physicians received no information from the methadone clinic about their patients.[25]

Patients receiving methadone for opioid dependence and engaged in primary medical care have substantial medical comorbidity.[8–16] Thus, it is not surprising that a substantial number of subjects in this study were on medications for

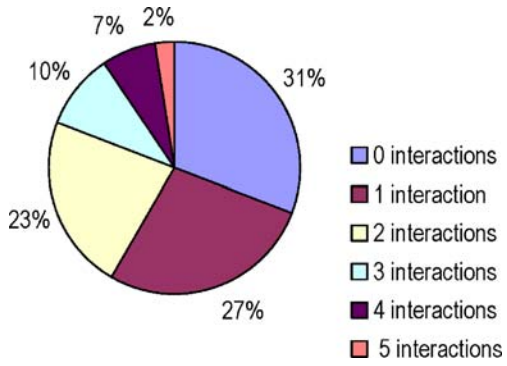


Figure 2. Distribution of the number of potentially harmful medication interactions with methadone among 84 subjects on methadone maintenance.

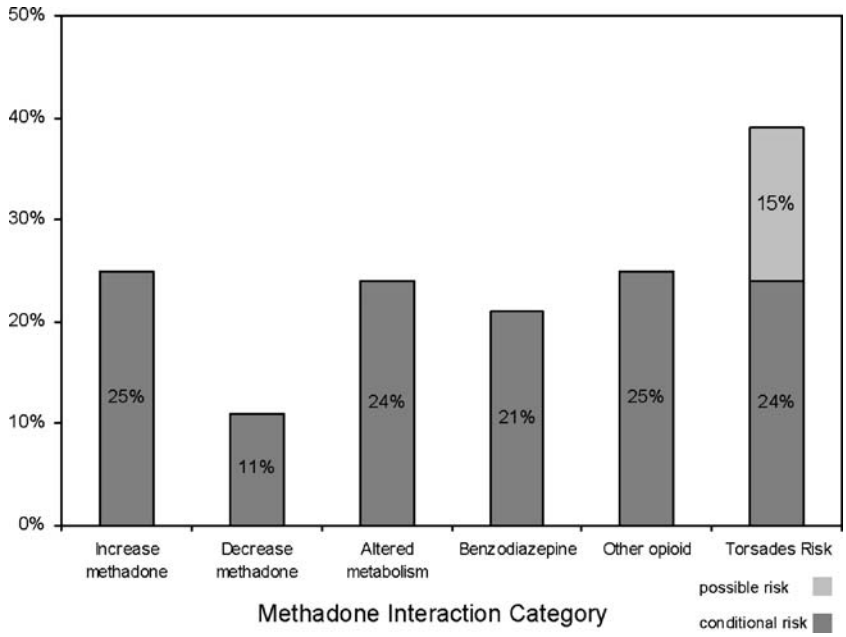


Figure 3. Medications in each category:
Increase methadone: fluoxetine, fluvoxamine, paroxetine, sertraline, omeprazole, metronidazole, ketoconazole, diltiazem
Decrease methadone: efavirenz, nevirapine, lopinavir+ritonavir, butalbital, ascorbic acid
Altered metabolism: didanosine, zidovudine, amytriptyline, doxepin, nortriptyline, clomethiazole., promethazine, nifedpine
Benzodiazepines: clonazepam, diazepam, lorazepam
Other opioids: codeine, fentanyl, morphine, oxycodone
Possible risk: atazanavir, azithromycin, levofloxacin, ofloxacin, quetiapine, risperidone, venlafaxine
Conditional risk: amitriptyline, nortriptyline, paroxetine, sertraline, citalopram, doxepin, fluoxetine, ketoconazole

depression, anxiety, HIV infection, hypertension, or pain. Not only do many of these medications potentially interact with methadone, but multiple interacting medications are common. While none of the identified interacting medications are contraindicated in conjunction with methadone, prescribers should be aware when patients are taking medications that interact, so they can monitor and manage potential interactions.

Among the 9 subjects with methadone missing from a medical note, "other opioids" was the most common category. For patients with pain and opioid dependence, it may be appropriate to be treated with both methadone and other opioids, yet it is critical that both treatments are coordinated to avoid overdose.

While not a focus of our study, the recruitment of eligible subjects points to an additional challenge to care coordination; written consent for communication between medical and substance use treatment providers is commonly not present. Because of federally mandated privacy protections,[5,6] the first step of care coordination between substance abuse treatment and medical providers is obtaining a signed release of information. In our study, almost half of methadone clinic patients did not have a signed release of information permitting communication with their primary care physician. The reasons for this are many: no primary care physician, difficulty for clinic staff to obtain releases because of separation of programs, and stigma among MMT patients with resulting reticence to allow communication with medical providers.

This study's strengths include examination of the EMRs from clinical substance-use treatment and the medical care system to describe the burden of potential methadone-medication interactions and missing documentation between methadone and medical care providers that has not previously received critical attention.

This study was limited by its retrospective design and limited number of subjects. Yet, we used electronic medical records and a data collection methodology that was systematic and the subjects examined were similar to patients in the entire clinic. The generalizability of our findings is limited by focusing on patients from one methadone clinic receiving medical care at an affiliated medical center. We recognize that it is possible that physicians were aware of their patients receiving MMT or being opioid dependent, but did not document it. While methadone is routinely administered during inpatient treatment at the medical center if the treatment team is aware the patient is enrolled in MMT, we did not have access to inpatient pharmacy records to confirm this. We did not evaluate whether medication reconciliation has taken place at the MMT site. However, as a consequence of this study, the methadone clinic did refine its medication reconciliation procedures to regularly document prescribed medications. Lastly, the retrospective design precluded measurement of adverse events from methadone-medication interactions.

This study demonstrates opportunities to improve communication, care coordination, and patient safety among patients receiving medical and substance-use treatment. Patients in MMT frequently require multiple medications and these often interact with methadone and potentially lead to adverse

events. Hence, communication and coordination among substance-use treatment and medical providers has room for improvement so as to mitigate and manage the potential adverse effects of methadone and interacting medications.

Acknowledgements: *The authors would like to thank Eileen Brigandi, Donna Beers, RN, and the staff at Substance Abuse Prevention and Treatment Services, Boston Public Health Commission, for their support and cooperation during this study. We would also like to thank Courtney Pierce for her assistance in manuscript preparation. This project was funded in part by the National Institute on Drug Abuse R25-DA13582. Parts of this work were presented at the Society of General Internal Medicine annual meeting, April 11, 2008; the College on Problems in Drug Dependence annual meeting, June 19, 2008; the Addiction Health Services Research conference, October 21, 2008; and the Association for Medical Education and Research in Substance Abuse annual meeting, November 8, 2008.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *Alexander Y. Walley, MD, MSc; Section of General Internal Medicine, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA(e-mail: awalley@bu.edu).*

REFERENCES

1. **Kohn LT, Corrigan JM, Donaldson MS.** Committee on quality of health care in america, institute of medicine. To err is human: building a safer health system. Washington, DC: National Academy Press; 2000.
2. Joint Commission. 2008 National Patient Safety Goals. Available at: www.jointcommission.org/PatientSafety/NationalPatientSafetyGoals/. Accessed May 31, 2009.
3. Improving the quality of healthcare for mental and substance-use conditions. Institute of Medicine. 504. 2005. Washington, DC: The National Academy Press: 2005.
4. **Kleber HD.** Methadone maintenance 4 decades later: thousands of lives saved but still controversial. JAMA. 2008;300:2303–5.
5. Bureau of National Affairs. **Beckerman JZ, Pritts J, Goplerud E, Leifer J, Borzi PC, Rosenbaum S**. Health information privacy, patient safety, and health care quality: Issues and challenges in the context of treatment for mental health and substance use.16. Available at: ihcrp.georgetown.edu/pdfs/pritts0208.pdf. Accessed May 31, 2009.
6. Confidentiality of Alcohol and Drug Abuse Patient Records. Code of Federal Regulations Title 42, Volume 1, Chapter 1, Part 2. Available at: http://www.access.gpo.gov/nara/cfr/waisidx_02/42cfr2_02.html. Accessed May 31, 2009.
7. **Leshner AI.** Science-based views of drug addiction and its treatment. JAMA. 1999;282:1314–16.
8. **Gourevitch MN, Chatterji P, Deb N, Schoenbaum EE, Turner BJ.** On-site medical care in methadone maintenance: associations with health care use and expenditures. J Subst Abuse Treat. 2007;32:143–51.
9. **Brienza RS, Stein MD, Chen M, et al.** Depression among needle exchange program and methadone maintenance clients. J Subst Abuse Treat. 2000;18:331–7.
10. **Teesson M, Havard A, Fairbairn S, Ross J, Lynskey M, Darke S.** Depression among entrants to treatment for heroin dependence in the Australian Treatment Outcome Study (ATOS): prevalence, correlates and treatment seeking. Drug Alcohol Depend. 2005;78:309–15.
11. **Milby JB, Sims MK, Khuder S, et al.** Psychiatric comorbidity: prevalence in methadone maintenance treatment. Am J Drug Alcohol Abuse. 1996;22:95–107.
12. **Rosenblum** A, Joseph H, Fong C, Kipnis S, Cleland C, Portenoy RK. Prevalence and characteristics of chronic pain among chemically dependent patients in methadone maintenance and residential treatment facilities. JAMA. 2003;289:2370–8.
13. **Kim TW, Alford DP, Malabanan A, Holick MF, Samet JH.** Low bone density in patients receiving methadone maintenance treatment. Drug Alcohol Depend. 2006;85:258–62.
14. **Federman AD, Arnsten JH.** Primary care affiliations of adults in a methadone program with onsite care. J Addict Dis. 2007;26:27–34.
15. **Umbricht-Schneiter A, Ginn DH, Pabst KM, Bigelow GE.** Providing medical care to methadone clinic patients: referral vs on-site care. Am J Public Health. 1994;84:207–10.
16. **Selwyn PA, Budner NS, Wasserman WC, Arno PS.** Utilization of on-site primary care services by HIV-seropositive and seronegative drug users in a methadone maintenance program. Public Health Rep. 1993;108:492–500.
17. **Krantz MJ, Martin J, Stimmel B, Mehta D, Haigney MC.** QTc interval screening in methadone treatment. Ann.Intern.Med. 2009;150:387–95.
18. **Krantz MJ, Lewkowiez L, Hays H, Woodroffe MA, Robertson AD, Mehler PS.** Torsade de pointes associated with very-high-dose methadone. Ann Intern Med. 2002;137:501–4.
19. **Krantz MJ, Kutinsky IB, Robertson AD, Mehler PS.** Dose-related effects of methadone on QT prolongation in a series of patients with torsade de pointes. Pharmacotherapy. 2003;23:802–5.
20. **Krantz MJ, Lowery CM, Martell BA, Gourevitch MN, Arnsten JH.** Effects of methadone on QT-interval dispersion. Pharmacotherapy. 2005;25:1523–9.
21. **Martell BA, Arnsten JH, Krantz MJ, Gourevitch MN.** Impact of methadone treatment on cardiac repolarization and conduction in opioid users. Am J Cardiol. 2005;95:915–8.
22. **Leavitt SB.** Methadone-drug interactions. Addiction Treatment Forum - 3rd Edition 2005; November 2005: 1–31. Available at: www.atforum.com/SiteRoot/pages/addiction_resources/Drug_Interactions.pdf. Accessed May 31, 2009.
23. **Bruce RD, Altice FL, Gourevitch MN, Friedland GH**. Pharmacokinetic drug interactions between opioid agonist therapy and antiretroviral medications: implications and management for clinical practice. J Acquir Immune Defic Syndr. 2006;41:563–72.
24. **Cance-Katz EF.** Treatment of opioid dependence and coinfection with HIV and hepatitis C virus in opioid-dependent patients: the importance of drug interactions between opioids and antiretroviral agents. Clin Infect Dis. 2005;41(Suppl 1):S89–S95.
25. **Velarde MC, Peino AJ, Gomez de Caso Canto JA.** The Methadone Maintenance Program for intravenous heroin addicts. What information do primary care physicians have? Aten Primaria. 1996;17:581–4.

Best Practice Updates for Informed Consent and Patient Education in Weight Loss Surgery

Christina C. Wee[1], Janey S. Pratt[2], Robert Fanelli[3,4], Patricia Q. Samour[5], Linda S. Trainor[6] and Michael K. Paasche-Orlow[7]

To update evidence-based best practice guidelines for obtaining informed consent from weight loss surgery (WLS) patients, with an emphasis on appropriate content and communications approaches that might enhance patient understanding of the information, we performed a systematic search of English-language literature published between April 2004 and May 2007 in MEDLINE and the Cochrane database. Keywords included WLS and informed consent, comprehension, health literacy, and patient education; and WLS and outcomes, risk, patient safety management, and effectiveness. Recommendations are based on the most current literature and the consensus of the expert panel; they were graded according to systems used in established evidence-based models. We identified over 120 titles, 38 of which were reviewed in detail. Evidence suggests that WLS outcomes, including long-term rates of relapse, vary by procedure. For some weight loss surgeries, long-term outcomes may not be known. Risks also vary by patient and provider characteristics. Informed consent should incorporate realistic projections of the short- and long-term risks, benefits, and consequences of surgery, as well as alternatives to WLS. For consent to be informed, the education process should continue until the patient demonstrates comprehension of all relevant material and concepts. Confirmation of comprehension can protect patients engaged in the process of consent for WLS. Future research should focus on the outcomes and consequences of WLS, and different approaches that facilitate patient understanding of, and decision making about, WLS.

Obesity (2009) **17,** 885–888. doi:10.1038/oby.2008.567

INTRODUCTION

Evidence-based best practice guidelines for informed consent in weight loss surgery (WLS) have been previously described (1). Previous recommendations focused on understanding vs. disclosure, appropriate content, teaching and learning, and promoting realistic expectations (1). The 2004 task group found no studies on informed consent and WLS. Recommendations were based on three review articles, standard practice at six WLS centers in Massachusetts, and the consensus of the expert panel (1). No studies on WLS and informed consent have been published since that time. Recommendations are, therefore, based on related articles identified through our search strategy. This report adds recommendations to those in the 2005 guidelines, and describes the supporting evidence.

To make sound medical decisions, patients must not only receive adequate and appropriate information, but also understand it (2). Poor comprehension of the risks, benefits, and consequences of surgery can contribute to unrealistic expectations, suboptimal decision making, and potential litigation (3).

The previous report cited a need for studies that assess the effect of different forms of education on patient understanding (1). This update draws on the evolving literature on patient safety and WLS outcomes to make recommendations on informed consent content. It also reviews the literature on patient comprehension in informed consent, including studies on health literacy and on clinical areas (e.g., cardiac surgery) that might apply to WLS.

METHODS AND PROCEDURES

We searched MEDLINE and the Cochrane database for articles on bariatric or elective surgery and informed consent, comprehension, health literacy, and patient education published between April 2004 and May 2007. We also conducted searches on bariatric surgery and outcomes, risk, patient safety management, and effectiveness. In addition, we reviewed WLS guidelines and other potentially relevant articles recommended by the expert panel or cited in the initial articles we identified.

[1]Division of General Medicine and Primary Care and Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA; [2]Department of Surgery and MGH Weight Center, Massachusetts General Hospital, Boston, Massachusetts, USA; [3]Department of Surgery, Berkshire Medical Center, Pittsfield, Massachusetts, USA; [4]Department of Surgery, University of Massachusetts Medical School, Pittsfield, Massachusetts, USA; [5]Department of Nutrition, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA; [6]Minimally Invasive Surgery and Department of Surgery, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA; [7]General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA. Correspondence: Christina C. Wee (cwee@bidmc.harvard.edu)

Received 21 June 2007; accepted 6 September 2007; published online 19 February 2009. doi:10.1038/oby.2008.567

The system used to grade the quality of the evidence has already been described (4). We identified >120 titles; 38 of the most relevant were reviewed in detail. These included randomized controlled trials, prospective and retrospective cohort studies, meta-analyses, case reports, prior systematic reviews, and expert opinion. The focus of the recommendations and the process used to develop them are reported elsewhere (4).

RESULTS

Content

Risks/complications. WLS centers are ethically obligated to provide patients with adequate information on the risks, benefits, consequences, and alternatives to treatment (2). A recent systematic review of randomized controlled trials and observational studies suggests that risks of mortality and complications associated with WLS vary depending on the type of procedure (5). Several observational studies also indicate that risks vary depending on patient characteristics, e.g., older adults, men, and those with greater comorbidity have higher risks of mortality and complications (6–10).

Risks can vary by as much as 20-fold across patient populations (6). Moreover, risks are higher in unselected and more generalizable populations (5,6,8) than in studies of selected patients seen primarily at tertiary surgical centers. The same applies to participants in studies where follow-up rates were low (5). One small single-site study found that patients who do not follow-up after surgery had poorer outcomes (10). Emerging evidence suggests that health provider characteristics, e.g., surgeon training, and hospital and surgeon volume of specific WLS procedures performed, may correlate with actual risk (11–13).

Recommendations

- Provide realistic risk estimates that take into account patient factors (category C) and relevant institutional and health provider characteristics that might affect risk (e.g., experience and outcomes for specific WLS procedures) (category B).
- Discuss short- and long-term risks and complications, and the potential for unknown or unforeseeable long-term risks (category D).

Benefits/effectiveness. Obesity is associated with premature mortality and other adverse health consequences, some of which are improved or reversed with weight loss (14). Controlled trials and observational studies demonstrate that WLS produces significant and sustained weight loss compared with alternative forms of treatment (5,15). In addition, two recent observational studies found that mortality rates of severely obese patients who had WLS were lower than those of severely obese patients who did not (16,17). Both studies were limited, however, by inadequate control for baseline BMI, illness burden, and potential selection bias.

Studies document substantial weight loss and health benefits from WLS at 1 year; longer-term studies suggest that some, but not all, of the short-term weight loss and medical benefits are sustained over time (18). Magnitude and sustainability of weight loss and benefits vary by type of WLS procedure and patient characteristics (5,18). One recent nonrandomized controlled study found that of 1,703 Swedish subjects enrolled for at least 10 years, 74% of gastric bypass and 35% of gastric banding patients, sustained at least a 20% weight loss; however, 9 and 25%, respectively sustained <5% weight loss (18). A small study of patients planning to undergo WLS found that expectations of weight loss from the surgery far exceeded even the best estimates in the literature (19).

Recommendations

- Provide patients with realistic estimates of short- and long-term weight loss, including the potential for weight regain and modest benefits (category B).
- Inform them if long-term data (>5 years) are unavailable (category D).
- Advise patients on the long-term health benefits of weight loss produced by WLS (category B).
- Make them aware that not all preexisting medical and psychosocial consequences of obesity, including eating disorders, will improve with WLS (category C).
- Give realistic estimates for health outcomes if patients decline surgical treatment (categories B and C), and advise them of known factors and interventions that might optimize benefits (category D).
- Consider patient expectations, the value placed on different outcomes, and the risks each candidate is willing to accept; address unrealistic expectations or other misconceptions patients might have (category C).

Consequences. In addition to risks and benefits, WLS is associated with physiological changes that may have an adverse impact on patient quality of life; these include gastrointestinal side-effects, nutritional deficiencies, and excess skin (20).

Recommendation

- Advise patients on required behavioral and dietary changes and other reasonable and foreseeable consequences of WLS that could affect health or quality of life in a substantive way, e.g., gastrointestinal symptoms, cosmetic effects, nutritional restrictions (category D).

Alternative treatments

WLS is currently the most effective treatment for moderate to severe obesity, patients, but patients vary in the value they place on weight loss and the risks and tradeoffs they are willing to make to lose weight (19,21). Risks, benefits, and tradeoffs vary among different WLS procedures and nonsurgical treatments (5,15,18,22).

Recommendations

- Advise patients about alternative WLS procedures and nonsurgical treatment options (e.g., medical and behavioral) (category C).
- Inform them about alternatives even if they are not available through the consenting health provider or institution (category C).

Comprehension of informed consent

The NIH has defined health literacy as the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions (23). Over 90 million American adults have inadequate health literacy skills (24). The prevalence of limited literacy is higher in older adults, in racial or ethnic minorities, and in those with limited education and chronic diseases (25).

The association between limited literacy and adverse health outcomes has been well documented. Seminal reports about the problem of health literacy have been issued by the Institute of Medicine (24), the Agency for Healthcare Research and Quality (26), the American Medical Association (27), and the Joint Commission on the Accreditation of Hospital Organizations (28). However, no studies have examined the relation between limited literacy and the WLS informed consent process.

Several lessons can still be drawn from related areas of research: limited literacy is a barrier to patient comprehension in the process of informed consent (29); clinicians are not able to discern which of their patients have limited literacy; and patients frequently fail to disclose that information due to concerns about shame (30–32). In addition, patients with limited literacy are less likely to ask questions than those with higher literacy skills (33).

Data show that screening patients for limited literacy skills is neither beneficial nor necessary. Conversely, confirmation of comprehension promotes understanding regardless of health literacy level, and has been promoted as a patient safety measure (34). Simplification of informed consent materials and the use of multiple channels of communication can also be useful (35–37). In particular, care is needed with the presentation of numerical concepts; many people lack basic quantitative literacy skills (38).

A recent study found that 66% of participants (174 of 264) were not able to answer a question relating to the odds of getting "heads" from flipping a coin (39). Moreover, clinicians are frequently unaware of the extent to which they communicate with jargon or use concepts that patients do not comprehend (40,41). In addition to consent forms and supporting brochures, websites, videos, and other materials and decision aids (35), patient participation in developing informed consent information can be helpful (42). A teach-to-goal educational approach, in which patient comprehension is evaluated and education continued until the patient exhibits mastery of the content, can help people with limited literacy (29).

Recommendations

- Evaluate each patient's comprehension of the risks, benefits, consequences, and alternatives to WLS (category C).
- Confirm comprehension to protect patients engaged in the informed consent process (category C).

DISCUSSION

The demographics of WLS patients are changing; older patients and those with greater comorbidity are now undergoing surgery. As a result of these changes, historic estimates of risks and benefits extrapolated from earlier studies of WLS may not apply to the current WLS population. The disproportionate prevalence of obesity in many racial and minority groups requires modifications to how the informed consent process is conducted and communicated; low health literacy is more prevalent in ethnic and racial minorities, and can be a barrier to adequate informed consent (24,25).

Our report is limited because of a lack of high-quality studies, particularly long-term, randomized trials on the risks, benefits, and consequences of various surgical and nonsurgical weight loss treatments. There are also few high-quality long-term observational studies on WLS outcomes in diverse populations. Future research is needed to better identify factors that affect WLS surgery outcomes in the long- and short-term so that patients can be cited appropriate and individualized outcomes information.

WLS is a rapidly evolving field. New surgical techniques are being developed and evaluated. The current report addresses the process of informed consent for routine WLS. We have not presented recommendations for experimental procedures. These require an informed consent process that adheres to federal regulations for the protection of human subjects. We did not address informed consent in the pediatric population either. Data on the risks, benefits, consequences, and alternatives to WLS differ from those on adult populations, and are more limited. Informed consent in the pediatric/adolescent arena requires the assent of the patient and informed consent of the patient's parents (43). This subject is covered in more detail in the Pediatric/Adolescent Task Group report (44).

Our recommendations to ensure patient comprehension of informed consent materials are largely based on relevant research in other clinical areas. Future studies should examine whether these approaches will be effective in the WLS arena, and their potential impact on patient satisfaction and WLS outcomes.

WLS is a high-risk procedure in a demographically diverse and clinically complex population. Rather than serve as a mere legal hurdle, informed consent should provide patients with easily understood and complete information needed to authorize the proposed surgery (2). Future research should focus on important gaps in knowledge on the outcomes and consequences of WLS, and different approaches to facilitate patient understanding of, and decision making about, WLS.

SUPPLEMENTARY MATERIAL

To review task group appendices, go to www.mass.gov/dph and search "Weight Loss Surgery."

ACKNOWLEDGMENTS

We thank Frank Hu for advice in manuscript preparation, Leslie Kirle for administrative support, and Rita Buckley for research and editorial services. This report on WLS was prepared for the Betsy Lehman Center for Patient Safety and Medical Error Reduction (Commonwealth of Massachusetts Boston, MA). Manuscript preparation was supported, in part, by the Boston Obesity Nutrition Research Center Grant P30-DK-46200 and the Center for Healthy Living, Division of Nutrition, Harvard Medical School.

DISCLOSURE

The authors declared no conflict of interest.

REFERENCES

1. Sabin J, Fanelli R, Flaherty H *et al*. Best practice guidelines on informed consent for weight loss surgery patients. *Obes Res* 2005;13:250–253.

2. Faden RR, Beauchamp TL. *A History and Theory of Informed Consent*. Oxford University Press: New York, 1986.
3. Kaufman AS, McNelis J, Slevin M, La Marca C. Bariatric surgery claims—a medico-legal perspective. *Obes Surg* 2006;16:1555–1558.
4. Lehman Center Weight Loss Surgery Expert Panel. Commonwealth of Massachusetts Betsy Lehman Center for Patient Safety and Medical Error Reduction Expert Panel on Weight Loss Surgery: executive report. *Obes Res* 2005;13:205–226.
5. Maggard MA, Shugarman LR, Suttorp M *et al*. Meta-analysis: surgical treatment of obesity. *Ann Intern Med* 2005;142:547–559.
6. DeMaria EJ, Portenier D, Wolfe L. Obesity surgery mortality risk score: proposal for a clinically useful score to predict mortality risk in patients undergoing gastric bypass. *Surg Obes Relat Dis* 2007;3:134–140.
7. Encinosa WE, Bernard DM, Chen CC, Steiner CA. Healthcare utilization and outcomes after bariatric surgery. *Med Care* 2006;44:706–712.
8. Jamal MK, DeMaria EJ, Johnson JM *et al*. Impact of major co-morbidities on mortality and complications after gastric bypass. *Surg Obes Relat Dis* 2005;1:511–516.
9. Hutter MM, Randall S, Khuri SF, Henderson WG, Abbott WM, Warshaw AL. Laparoscopic versus open gastric bypass for morbid obesity: a multicenter, prospective, risk-adjusted analysis from the National Surgical Quality Improvement Program. *Ann Surg* 2006;243:657–662.
10. Harper J, Madan AK, Ternovits CA, Tichansky DS. What happens to patients who do not follow-up after bariatric surgery? *Am Surg* 2007;73:181–184.
11. Ballantyne GH, Ewing D, Capella RF *et al*. The learning curve measured by operating times for laparoscopic and open gastric bypass: roles of surgeon's experience, institutional experience, body mass index and fellowship training. *Obes Surg* 2005;15:172–182.
12. Nguyen NT, Moore C, Stevens CM, Chalifoux S, Mavandadi S, Wilson SE. The practice of bariatric surgery at academic medical centers. *J Gastrointest Surg* 2004;8:856–860.
13. Weller WE, Hannan EL. Relationship between provider volume and postoperative complications for bariatric procedures in New York State. *J Am Coll Surg* 2006;202:753–761.
14. U.S. Department of Health and Human Services. The practical guide—identification, evaluation, and treatment of overweight and obesity in adults. Bethesda, MD: National Institutes of Health, 2000. (NIH publication No. 00-4084.)
15. Li Z, Maglione M, Tu W, Mojica W *et al*. Meta-analysis: pharmacologic treatment of obesity. *Ann Intern Med* 2005;142:532–546.
16. Flum DR, Dellinger EP. Impact of gastric bypass operation on survival: a population-based analysis. *J Am Coll Surg* 2004;199:543–551.
17. Christou NV, Sampalis JS, Liberman M *et al*. Surgery decreases long-term mortality, morbidity, and health care use in morbidly obese patients. *Ann Surg* 2004;240:416–423.
18. Sjostrom L, Lindroos AK, Peltonen M *et al*. Swedish Obese Subjects Study Scientific Group. Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. *N Engl J Med* 2004;351:2683–2693.
19. Wee CC, Jones DB, Davis RB, Bourland AC, Hamel MB. Understanding patients' value of weight loss and expectations for bariatric surgery. *Obes Surg* 2006;16:496–500.
20. Pietras SM, Usdan LS, Apovian CM. Preoperative and postoperative management of the bariatric surgical patient. *J Clin Outcomes Manage* 2007;14:262–274.
21. Wee CC, Hamel MB, Davis RB, Phillips RS. Assessing the value of weight loss among primary care patients. *J Gen Intern Med* 2004;19:1206–1211.
22. O'Brien PE, Dixon JB, Laurie C *et al*. Treatment of mild to moderate obesity with laparoscopic adjustable gastric banding or an intensive medical program: a randomized trial. *Ann Intern Med* 2006;144:625–633.
23. National Institutes of Health Program Announcement: Understanding and Promoting Health Literacy (RO1). <http://grants.nih.gov/grants/guide/pa-files/PAR-07-020.html>. Accessed 25 August 2007.
24. Nielsen-Bohlman L, Panzer AM, Kindig DA. Institute of Medicine. Health Literacy: A Prescription to End Confusion. National Academies Press. <http://books.nap.edu/catalog.php?record.id=10883>. Accessed 25 August 2007.
25. Paasche-Orlow MK, Parker RM, Gazmararian JA *et al*. The prevalence of limited health literacy. *J Gen Intern Med* 2005;20:175–184.
26. Berkman ND, DeWalt DA, Pignone MP *et al*. Literacy and Health Outcomes. Evidence Report/Technology Assessment No. 87 (Prepared by RTI International-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016). AHRQ Publication No. 04-E007-2. January 2004. Rockville, MD: Agency for Healthcare Research and Quality.
27. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs AMA. Health literacy: report of the Council on Scientific Affairs. *JAMA* 1999;281:552–557.
28. Health Literacy and Patient Safety. Joint Commission on Accreditation of Healthcare Organizations. <http://www.jointcommission.org/PublicPolicy/health_literacy.htm>. 2006. Accessed 25 August 2007.
29. Sudore RL, Landefeld CS, Williams BA *et al*. Use of a modified informed consent process among vulnerable patients: a descriptive study. *J Gen Intern Med* 2006;21:867–873.
30. Parikh NS, Parker RM, Nurss JR *et al*. Shame and health literacy: the unspoken connection. *Patient Educ Couns* 1996;27:33–39.
31. Bass PF III, Wilson JF, Griffith CH, Barnett DR. Residents' ability to identify patients with poor literacy skills. *Acad Med* 2002;77:1039–1041.
32. Lindau ST, Tomori C, Lyons T *et al*. The association of health literacy with cervical cancer prevention knowledge and health behaviors in a multiethnic cohort of women. *Am J Obstet Gynecol* 2002;186:938–943.
33. Katz MG, Jacobson TA, Veledar E, Kripalani S. Patient literacy and question-asking behavior during the medical encounter: a mixed-methods analysis. *J Gen Intern Med* 2007;22:782–786.
34. Learning from Early Adopters: Improving Patient Safety Through Informed Consent for Patients with Limited English Proficiency/Low-Literacy Populations. National Quality Forum. <http://216.122.138.39/projects/completed/informed_consent.asp>. Accessed 25 August 2007.
35. O'Connor AM, Stacey D, Entwistle V *et al*. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* 2003;CD001431.
36. Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: a systematic review. *JAMA* 2004;292:1593–1601.
37. Moseley TH, Wiggins MN, O'Sullivan P. Effects of presentation method on the understanding of informed consent. *Br J Ophthalmol* 2006;90:990–993.
38. Fagerlin A, Ubel PA, Smith DM, Zikmund-Fisher BJ. Making numbers matter: present and future research in risk communication. *Am J Health Behav* 2007;in press.
39. Aggarwal A, Speckman JL, Paasche-Orlow MK, Roloff KS, Battaglia TA. The role of numeracy on cancer screening among urban women. *Am J Health Behav* 2007;31(Suppl 1):S47–S56.
40. Castro CM, Wilson C, Wang F, Schillinger D. Babel babble: Physicians' use of unclarified medical jargon with patients. *Am J Health Behav* 2007;31(Suppl 1):S85–S95.
41. Kusec S, Oreskovic S, Skegro M *et al*. Improving comprehension of informed consent. *Patient Educ Couns* 2006;60:294–300.
42. Bottrell MM, Alpert H, Fischbach RL, Emanuel LL. Hospital informed consent for procedure forms: facilitating quality patient-physician interaction. *Arch Surg* 2000;135:26–33.
43. Inge TH, Krebs NF, Garcia VF *et al*. Bariatric surgery for severely overweight adolescents: concerns and recommendations. *Pediatrics* 2004;114:217–223.
44. Pratt J, Lenders CM, Dionne E *et al*. Best practice updates for pediatric/adolescent weight loss surgery. *Obesity*, this issue.

HEALTH POLICY

High Quality Care and Ethical Pay-for-Performance: A Society of General Internal Medicine Policy Analysis

J. Frank Wharam, MB, BCh, BAO, MPH[1], Michael K. Paasche-Orlow, MD, MPH[2], Neil J. Farber, MD[3], Christine Sinsky, MD[4], Lisa Rucker, MD[5], Kimberly J. Rask, MD, PhD[6], M. Kathleen Figaro, MD, MPH[7], Clarence Braddock III, MD, MPH[8], Michael J. Barry, MD[9], and Daniel P. Sulmasy, OFM, MD, PhD[10]

[1]Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA, USA; [2]Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [3]University of California, San Diego, CA, USA; [4]Medical Associates Clinic and Health Plan, Dubuque, IA, USA; [5]Albert Einstein College of Medicine, Bronx, NY, USA; [6]Department of Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, GA, USA; [7]Department of General Internal Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA; [8]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; [9]Medical Practices Evaluation Center, General Medicine Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; [10]St. Vincent's Hospital, Manhattan and New York Medical College, New York, USA.

BACKGROUND: Pay-for-performance is proliferating, yet its impact on key stakeholders remains uncertain.

OBJECTIVE: The Society of General Internal Medicine systematically evaluated ethical issues raised by performance-based physician compensation.

RESULTS: We conclude that current arrangements are based on fundamentally acceptable ethical principles, but are guided by an incomplete understanding of health-care quality. Furthermore, their implementation without evidence of safety and efficacy is ethically precarious because of potential risks to stakeholders, especially vulnerable patients.

CONCLUSION: We propose four major strategies to transition from risky pay-for-performance systems to ethical performance-based physician compensation and high quality care. These include implementing safeguards within current pay-for-performance systems, reaching consensus regarding the obligations of key stakeholders in improving health-care quality, developing valid and comprehensive measures of health-care quality, and utilizing a cautious evaluative approach in creating the next generation of compensation systems that reward genuine quality.

KEY WORDS: ethics; health policy; pay-for-performance; quality improvement; physician reimbursement.

J Gen Intern Med 24(7):854–9

DOI: 10.1007/s11606-009-0947-3

Pay-for-performance systems seek to improve health-care quality by providing bonus dollars to physicians, practice groups, or hospitals whose patients achieve certain health goals.[1] These arrangements are proliferating, yet their impact on key stakeholders remains uncertain.[2–7] The Society of General Internal Medicine (SGIM), through its Ethics Committee, systematically evaluated ethical issues raised by performance-based physician compensation. Investigations included literature review, in-depth interviews with key informants, focus groups among SGIM members, open forums at national SGIM meetings, and discussions among SGIM committees and leadership. A comprehensive report of the Ethics Committee's findings and recommendations is available at http://www.sgim.org/index.cfm?pageId=806.[8]

This position paper begins by examining the fundamental principles of pay-for-performance and setting forth our organization's definition of health-care quality. Based on this exploration, we present our conclusions regarding the manner of implementation of pay-for-performance and its potential effects on key stakeholders. We propose four major strategies for moving toward more ethical and effective performance-based physician compensation, emphasizing the need to implement immediate safeguards to protect vulnerable populations.

FUNDAMENTAL PRINCIPLES OF PAY-FOR-PERFORMANCE; UNDERSTANDING HEALTH-CARE QUALITY

The fundamental principles of pay-for-performance include rewarding quality health care and aligning physicians' financial incentives with the best interests of patients.[9] Although this inherent appeal to physician self-interest might be in tension with professional ideals of altruism and beneficence,[10–13] the principles that inform pay-for-performance are not inherently unethical. It seems just, for example, to financially reward physicians who demonstrate outstanding levels of patient-centered and evidence-based care.

Nevertheless, systems intending to improve medical care must be guided by evidence and a precise definition of health-care quality to ensure that they are effective, valid, and fair. We define health-care quality in a manner that prioritizes patient-

Received July 26, 2008
Revised December 23, 2008
Accepted January 26, 2009
Published online March 18, 2009

centered care[14] while recognizing the importance of population-level health improvement:[15]

> Health-care quality is the degree to which physicians and health-care institutions fulfill their care obligations to individual patients and the degree to which patients, physicians, and health-care institutions enable these obligations to be fulfilled justly across the population.[16]

This understanding of health-care quality informs our criticisms of current pay-for-performance arrangements and provides a roadmap to high quality care and ethical performance-based physician compensation.

Potential Ethical Problems in the Implementation of Pay-for-Performance

In light of these principles, we see the following potential ethical problems in the implementation of pay-for-performance systems.

1. *Lack of proven safety and benefit for patients:* Studies of performance-based physician compensation have generally shown scant evidence of quality improvement.[2–7] Implementation without proof of safety and effectiveness is ethically problematic. It is unclear, for instance, why a new drug to be used by several dozen individuals requires proof of safety and efficacy, while policy changes affecting millions do not. From an ethical perspective, pay-for-performance is a potentially risky experiment in health-care delivery.[17] Further, current pay-for-performance systems generally lack key safeguards against readily anticipated adverse effects [1] (discussed below), and we are concerned that negative outcomes may already be unfolding.
2. *Inadequate definitions of quality:* Although commentators have proposed many definitions of health-care quality,[18–27] none are universally accepted, and they provide little guidance regarding accountability or how quality can be validly measured. Furthermore, current pay-for-performance arrangements are guided by a highly incomplete understanding of quality that does not resemble any published or well-reasoned definition. This understanding typically equates quality with the achievement of non-individualized, pre-determined health goals for broad populations and fails to consider contributions from stakeholders other than physician entities[1] (such as health plans) that also have partial responsibility for ensuring quality. This approach has severely limited our ability to capture the myriad of elements comprising quality care, let alone the most complex but essential feature of a praxis like medicine—the exercise of correct judgment—which is only readily assessed by peers.[28]
3. *Inadequate measures of quality:* Because they are based on inadequate definitions, existing pay-for-performance measures lack validity and comprehensiveness in assessing health-care quality. Measures typically cover only isolated and readily quantifiable aspects of physician clinical performance and fail to assess crucial realms such as judgment, compassion, and communication skills.

Quantifying health-care quality is notoriously difficult, and basing payment incentives upon inadequate measures and definitions of quality will make consequences difficult to control. Unfortunately, this approach is often used to make judgments about individual practitioners when variability in case mix and patient preferences precludes making valid judgments. For example, in a patient with difficult-to-control diabetes, a decline in hemoglobin A_1C from 10.0 to 8.0 might be a remarkable achievement and more validly represent high quality care than a decline from 7.3 to 6.9 in another patient.

4. *Misallocating the locus of accountability for quality improvement:* Many pay-for-performance measures hold physicians accountable for aspects of quality beyond their control, such as health-care delivery problems, lack of incentives for coordinated care, and even social determinants of health. Some may hold physicians accountable for the care of patients with whom they have not had a continuous relationship.
5. *Potential for adverse effects on patients and vulnerable populations:* Performance targets used today may have detrimental effects on quality. For example, it may seem reasonable to require that diabetic patients achieve hemoglobin A_1C levels below 7.0. In patients with previous hypoglycemic episodes, however, this target might be life-threatening.

Or, consider a patient with a hemoglobin A_1C of 7.5 who frequently skips preventive visits but happens to present with back pain. If bonuses are provided for reducing glucose levels, a physician might prefer to discuss diabetes control rather than ruling out life-threatening causes of back pain. Such "treating the measure" might worsen outcomes.

Pre-determined population-centered measures might also induce physicians to avoid patients who are less likely to meet targets. Such patients are often society's most vulnerable members—those with multiple chronic conditions, the poor, the educationally disadvantaged, those with limited English proficiency, and members of racial minority groups.

Because physicians serving disadvantaged patients might receive lower compensation, less well-off practices would be left with fewer resources to improve care. This could create a vicious cycle of worsening quality for the most vulnerable patients.

Poorly designed pay-for-performance systems could therefore limit access to care for vulnerable populations, worsen health-care quality, and erode patient trust.[29–38]

6. *Potential for adverse effects on physicians:* In systems using a limited set of population-level measures, physician professionalism and morale could decline. Some clinicians might "treat the measure" or select the "best" patients to enhance income. Others might provide optimal care despite reduced income, but grow frustrated. Pay-for-performance is also likely to increase the complexity of the reimbursement system, and metrics might be used against physicians for legal, credentialing, or recertification purposes. Such changes would decrease physician job satisfaction, with detrimental effects on patient care and the attractiveness of medicine (especially primary care) as a profession.
7. *Potential for adverse effects on society:* The potential detrimental effects above would have broader implications for society. A decreasing supply of primary care physicians would exacerbate problems in access and quality. Truly valid and comprehensive measurements might require overly burdensome or expensive systems, and could make

the marginal value of performance-based compensation negligible. Deteriorating value could also result if physicians drive up expenses by ordering unnecessary tests or referrals to specialists.[36]

Ultimately, insurers could face a backlash by patients and physicians against an effort that might be viewed cynically as another cost-containment attempt, offered disingenuously as quality improvement.

8. *Lack of structured monitoring for adverse outcomes:* A substantial literature advocates structured oversight of any risky intervention not meant to directly benefit individuals.[39–53] Although pay-for-performance is intended to improve patient care, some would argue that it is primarily a population-centered cost control measure with unclear effectiveness and a substantial risk-benefit ratio for certain populations.[33] We believe the risks from pay-for-performance outlined above are serious enough and have a high enough probability of occurring to engender an ethical obligation for structured monitoring of key outcomes (discussed below).

POLICY RECOMMENDATIONS

SGIM supports evidence-based, ethical, and comprehensive efforts to improve health-care quality and physician compensation. While carefully designed pay-for-performance systems could be a component of such an approach, current iterations fail to reach acceptable ethical standards for the reasons above. We therefore advocate the following four major strategies to achieve high quality health care and ethical performance-based physician compensation (Tables 1, 2, 3).

Current Pay-for-Performance Systems should Rapidly Adopt Safeguards to Protect Vulnerable Populations

Until researchers develop valid and comprehensive quality measures, pay-for-performance systems must prioritize the protection of vulnerable populations and minimize readily anticipated adverse consequences (Table 3). Pay-for-performance leaders should institute the following safeguards to achieve these aims:

1. *Balance current population-level measurements with the best available measures of quality from the patient perspective.* The non-patient-centered nature of current pay-for-performance systems could be partially remedied by appropriate measures. For example, the Consumer Assessment of Healthcare Providers[54] places a strong emphasis on measuring how well health-care providers communicate with patients. A growing body of research[55,56] could inform the development of valid measures in the outpatient setting.
2. *Reduce or stabilize the percentage of physicians' salaries at stake.* Policy makers should limit bonus amounts to reduce temptations to "game" the system, especially in arrangements that do not adjust for case mix. Current levels of approximately 5% of physicians' salaries seem reasonable in systems that adjust for case mix, while lower levels would be appropriate for those that do not.
3. *Provide adequate off-setting compensation to physicians serving vulnerable patients.* For example, the 2006 Massachusetts health-care reform legislation included provisions to base Medicaid hospital rate increases on quality improvement, including the reduction of health-care dis-

Table 1. Potential Ethical Problems in the Implementation of Pay-for-Performance

I. Lack of proven safety and benefit for patients
II. Inadequate definitions of quality
III. Inadequate measures of quality
IV. Misallocating the locus of accountability for quality improvement
V. Potential for adverse effects on patients and vulnerable populations
VI. Potential for adverse effects on physicians
VII. Potential for adverse effects on society
VIII. Lack of structured monitoring for adverse outcomes

Table 2. Major Strategies to Achieve High Quality Health Care and Ethical Performance-Based Physician Compensation

I. Current pay-for-performance systems should rapidly adopt safeguards to protect vulnerable populations (see Table 3)
II. Key stakeholders should develop consensus regarding their responsibilities in improving health-care quality
III. Researchers and policy makers should develop valid and comprehensive quality measures for use in the next generation of compensation systems that reward genuine quality
IV. Researchers and policy makers should use a cautious evaluative approach to long-term development of compensation systems that reward quality

Table 3. Recommended Safeguards to Protect Vulnerable Populations and Prevent Unintended Consequences Within Current Pay-for-Performance Systems

1. Balance current population-level measurements with the best available measures of quality from the patient perspective
2. Reduce or stabilize the percentage of physicians' salaries at stake (except as in point 3 below)
3. Provide adequate off-setting compensation for physicians serving vulnerable patients
4. Population-level measures should:
a. Be evidence-based and clearly linked to valued patient outcomes
b. Assess domains clearly within the influence of the physician or physician group, especially for complex patients
c. Assess quality at the level of large physician practices rather than individual physicians
d. Assess improvement toward goals in addition to achievement of cut-points
5. If systems utilize population-level *outcomes* measures, they should:
a. Explicitly assess patient complexity and vulnerability
b. Carefully adjust for case-mix based on relevant patient factors
c. Carefully adjust for the manner in which responsibility for patient outcomes is shared between physicians, patients, health plans, and other health-care institutions
6. Initiate monitoring before and after implementing the above changes. Monitoring should assess:
a. Patient satisfaction, access, continuity, and coordination of care; effects on vulnerable patients as a particularly important focus
b. Physician satisfaction and professionalism, administrative burden, effects on the patient-physician relationship
c. Effects on disparities between physician practices serving vulnerable and non-vulnerable populations
d. Payer satisfaction and value for health-care expenditures

parities.[57] If such provisions are designed meticulously and fairly,[58] financial incentives could encourage and reward physicians for serving patients with low levels of expendable income, complex medical conditions, non-adherence to recommended treatments, or limited health literacy.

4. *Recommendations regarding population-level measures.* Pre-determined population-level measures of quality must be instituted carefully because they are inherently non-patient-centered. Because such measures are pervasive in modern pay-for-performance systems, we recommend several strategies to maximize the protection of vulnerable patients:

   4.a. *Utilize population-level measures that are evidence-based and clearly linked to valued patient outcomes.* For example, pneumonia and influenza immunizations have been proven to prevent potentially debilitating illnesses while having minimal adverse effects. Other commonly utilized measures may fail to reach these standards; hemoglobin A1C targets are based on evidence from randomized control trials, but the applicability to individual patients on real-life physician panels is often unclear.[35,59]

   4.b. *Population-level measures should assess domains clearly within the influence of the physician or physician group, especially for complex patients.* Basic process measures, such as vaccination rates and the frequency of diabetic eye exams, are imperfect measures of quality, but are more within the influence of physicians and practice groups than outcomes measures. Process measures seem less likely than outcomes measures to cause avoidance of vulnerable patients and physician frustration.

   4.c. *Measures should assess quality at the level of large physician practices rather than individual physicians.* Experts skilled in statistical analysis should determine minimum patient population sizes for each measure to provide optimal data and avoid statistical error. Only practice groups with sufficient numbers of patients should initially be measured.

   4.d. *Measures should assess improvement toward goals in addition to achievement of cut-points.* This could apply to both process and outcomes measures. For example, physician groups could be rewarded both for achieving vaccination rates at a pre-determined level as well as for annual improvements toward the target.

5. *Recommendations regarding population-level outcomes measures.* Population-level *outcomes* measures are methodologically complex, and the validity of current measures is uncertain. This will likely preclude their use in an ethically defensible manner in the short-term unless provisions that maximize validity are closely followed, including:

   5.a. *Explicitly assess patient complexity and vulnerability.* This would require integrating patient survey data and medical record data regarding sociodemographic characteristics and medical comorbidities.

   5.b. *Carefully adjust for case-mix based on relevant patient factors.* For example, it would be inappropriate to reduce systolic blood pressure levels below 140 mmHg in an 85-year-old diabetic patient with multiple comorbidities taking three antihypertensive medications. Proper case-mix adjustment might allow this patient's physician to prioritize other care, while a lack of adjustment could induce either dangerous efforts to lower blood pressure or substantial physician frustration.

   5.c. *Carefully adjust for the manner in which responsibility for patient outcomes is shared between physicians, patients, health plans, and other health-care institutions.* For example, consider two physicians who must eventually prescribe three hypoglycemic medications to similar diabetic patients whose initial hemoglobin A1C levels were 9.5. The first patient has generous health insurance, enabling him to purchase all three medications and lower his hemoglobin A1C to 6.5. The second patient must pay the full cost of medications, and she can only afford two. She only lowers her hemoglobin A1C to 7.5. A proper system would adjust for health insurance status.

6. *Pay-for-performance leaders should initiate monitoring before and after implementing the above changes.* Monitoring should assess important patient outcomes not often included in pay-for performance studies, such as satisfaction, access, continuity, and coordination of care. Effects on vulnerable patients should be a particularly important focus. Studies should also assess physician satisfaction and professionalism, administrative burden, effects on the patient-physician relationship, and the impact on disparities between physician practices serving more vulnerable and less vulnerable populations. Monitoring should examine payer satisfaction and value for health-care expenditures.

Key Stakeholders should Develop Consensus Regarding their Responsibilities in Improving Health Care Quality

A crucial first step in achieving ethically defensible health-care quality improvement will be for key stakeholders to develop consensus regarding their shared and unique obligations to individual patients and patient populations. For example, to improve blood glucose control among diabetic patients, physicians must recommend evidence-based, patient-centered management strategies, practice groups must provide access to testing facilities, health insurers must facilitate receipt of affordable medications and testing, and patients must adhere to therapeutic plans.

Bringing health insurers, patients, employers, and physicians to the table would highlight opportunities to improve coordination and continuity of care; new paradigms for quality improvement that integrate assessment at the individual physician level and institution level could emerge.

Researchers and Policy Makers should Develop Valid and Comprehensive Quality Measures for Use in the Next Generation of Compensation Systems that Reward Genuine Quality

A long-term strategy for quality improvement will be guided by a framework of accountability in which physicians, practice groups, health plans, and public payers are measured based on how well they fulfill well-defined obligations to individual patients and populations.

For example, measures of *physician* quality should assess multiple domains, such as accessibility, adherence to evidence-based but patient-centered care, and communication skills. Appropriate measures would account for individualized patient-physician goals, be based on the best available evidence, and minimize administrative burden and expense.

Measures of *health-care institution* quality (e.g., physician groups, hospitals, and public and private payers) should assess domains such as how well these groups foster teamwork, facilitate achievement of patient goals, strengthen the doctor-patient relationship, and improve access, coordination, and continuity of care for individual patients.

Equally important will be development of valid *population-level* health-care quality measures. In addition to measuring how well physicians and health-care institutions fulfill obligations to individual patients, comprehensive quality measures would assess the degree to which patients, physicians, and health-care institutions maximize health-care resources available to the population, distribute them fairly,[60] and fulfill their obligations justly.

Measures should be developed under strict principles of transparency. For example, all persons involved in creating new measures should, at minimum, be required to state potential conflicts of interest.

Researchers and Policy Makers Should Use a Cautious Evaluative Approach to Long-Term Development of Compensation Systems that Reward Quality

After developing evidence-based measures of physician, health-care institution, and population-level quality, policy makers should implement carefully planned, small-scale pilot programs that reward physician and health-care institution quality. Benefits and adverse effects should be monitored. Those entities implementing innovations in payment and quality improvement should take the lead in funding these studies.

Even with results from well-designed studies, judgments about the ethics of pay-for-performance will remain challenging. One approach might be to give preferential consideration to outcomes among vulnerable patients.

We base our suggestion to begin with pilot programs upon an ethical principle of precaution. However, efforts should be scaled up if benefits prove sufficient, health disparities are reduced and adverse outcomes are minimized.

THE ROLE OF SGIM

In order to aid in the above processes, SGIM is committed to having general internists participate in articulating the quality-related obligations that physicians and health-care institutions have to patients and the population. SGIM encourages its members to take the following actions: (1) help develop measures of physician, health-care institution, and population-level health-care quality, (2) evaluate pay-for-performance measures and programs, and (3) participate in the ongoing monitoring of effects of pay-for-performance on vulnerable populations and physicians. SGIM will continue to develop collaborative alliances with other key national organizations to ensure fair, valid, and comprehensive measures and to promote ethical compensation reform.

CONCLUSIONS

Performance-based physician compensation, if carefully guided by a comprehensive understanding of health-care quality and evidence-based evaluations, might improve patient care, narrow health disparities, and promote fair physician compensation while increasing health-care value. If research and monitoring determine that improved payment systems can benefit patients, physicians, and payers while minimizing risks, they could be ethical arrangements. However, until such data are available, widespread expansion of untested pay-for-performance systems poses substantive ethical issues associated with potential harm to patients, clinicians, and organizations.

Acknowledgements: *We would like to thank members of the SGIM Ethics Committee for their thoughtful comments on several versions of this article. We are also grateful to members of the SGIM Council, the Health Policy Committee, the Disparities Task Force, and the Clinical Practice Committee for reviewing and critiquing the manuscript. We especially thank SGIM Presidents Eugene Rich of Creighton University Medical Center and Lisa Rubenstein of the University of California, Los Angeles, for their suggestions and guidance. We are indebted to the SGIM members who volunteered to facilitate focus groups that informed this position paper: Alexia Torke of the University of Chicago, Tom Staiger and Eliza Sutton of the University of Washington, and Louise Walter of the University of California, San Francisco. We are similarly grateful to Ann-Marie Rosland of the University of Michigan and Matthew Frank of the Harvard Divinity School for assistance in analysis of focus group transcripts and Audrey Mertens Tedeman of the Massachusetts General Hospital for assistance with transcription of focus groups. There were no sources of funding for this work, and no financial support was provided to authors or contributors.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *J. Frank Wharam, MB, BCh, BAO, MPH; Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 133 Brookline Avenue, 6th Floor, Boston, MA 02114, USA (e-mail: jwharam@partners.org).*

REFERENCES

1. **Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM.** Pay for performance in commercial HMOs. N Engl J Med. 2006; 35518: 1895–1902.
2. **Rosenthal MB, Frank RG.** What is the empirical basis for paying for quality in health care? Med Care Res Rev. 2006; 632: 135–157.
3. **Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S.** Does pay-for-performance improve the quality of health care? Ann Intern Med. 2006; 1454: 265–272.

4. **Amundson G, Solberg LI, Reed M, Martini EM, Carlson R.** Paying for quality improvement: compliance with tobacco cessation guidelines. Jt Comm J Qual Saf. 2003; 292: 59–65.
5. **Fairbrother G, Hanson KL, Friedman S, Butts GC.** The impact of physician bonuses, enhanced fees, and feedback on childhood immunization coverage rates. Am J Public Health. 1999; 892: 171–175.
6. **Kouides RW, Bennett NM, Lewis B, Cappuccio JD, Barker WH, LaForce FM.** Performance-based physician reimbursement and influenza immunization rates in the elderly. The Primary-Care Physicians of Monroe County. Am J Prev Med. 1998; 142: 89–95.
7. **Roski J, Jeddeloh R, An L, et al.** The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines. Prev Med. 2003; 363: 291–299.
8. **Wharam JF, Farber NJ, Paasche-Orlow MK, Sinsky C, Rask KJ, Rucker L, Figaro K, Braddock C, Barry M, Sulmasy D.** Can Pay-for-Performance be Ethical? An Analysis by the Society of General Internal Medicine Ethics Committee. Available at: http://www.sgim.org/index.cfm?pageId=806.
9. **Epstein AM, Lee TH, Hamel MB.** Paying physicians for high-quality care. N Engl J Med. 2004; 3504: 406–410.
10. **Blumenthal D.** The vital role of professionalism in health care reform. Health Aff (Millwood). 1994; 131: 252–256.
11. **Swick HM.** Toward a normative definition of medical professionalism. Acad Med. 2000; 756: 612–616.
12. **Wynia MK, Latham SR, Kao AC, Berg JW, Emanuel LL.** Medical professionalism in society. N Engl J Med. 1999; 34121: 1612–1616.
13. **Beauchamp TL, Childress JF.** Principles of biomedical ethics. 5th ed. New York, N.Y.: Oxford University Press; 2001.
14. **Davis K, Schoenbaum SC, Audet AM.** A 2020 vision of patient-centered primary care. J Gen Intern Med. 2005; 2010: 953–957.
15. **Povar GJ, Blumen H, Daniel J, et al.** Ethics in practice: managed care and the changing health care environment: medicine as a profession managed care ethics working group statement. Ann Intern Med. 2004; 1412: 131–136.
16. **Wharam JF, Paasche-Orlow MK, Farber NJ, Sinsky C, Rucker L, Rask KJ, Figaro MK, Braddock C, Barry M, Sulmasy DP.** High Quality Care and Ethical Pay-for-Performance: A Position Statement of the Society of General Internal Medicine. Available at: http://www.sgim.org/index.cfm?section=site&pageId=806. Accessed December 10, 2008.
17. **Wharam JF, Daniels N.** Toward evidence-based policy making and standardized assessment of health policy reform. Jama. 2007; 2986: 676–679.
18. Committee on Quality of Health Care in America IoM. Crossing the Quality Chasm: A New Health System for the 21st Century. Available at: http://www.nap.edu/catalog/10027.html#toc. Accessed November 14, 2006.
19. **Lohr KN, Donaldson MS, Harris-Wehling J.** Medicare: a strategy for quality assurance, V: Quality of care in a changing health care environment. QRB Qual Rev Bull. 1992; 184: 120–126.
20. **Brook RH, McGlynn EA, Shekelle PG.** Defining and measuring quality of care: a perspective from US researchers. Int J Qual Health Care. 2000; 124: 281–295.
21. **Campbell SM, Roland MO, Buetow SA.** Defining quality of care. Soc Sci Med. 2000; 5111: 1611–1625.
22. **Maxwell RJ.** Dimensions of quality revisited: from thought to action. Qual Health Care. 1992; 13: 171–177.
23. Quality of care: 1. What is quality and how can it be measured? Health Services Research Group. Cmaj. Jun 15 1992;146(12):2153-2158.
24. **Donabedian A.** The seven pillars of quality. Arch Pathol Lab Med. 1990; 11411: 1115–1118.
25. **O'Leary DS, O'Leary MR.** From quality assurance to quality improvement. The Joint Commission on Accreditation of Healthcare Organizations and Emergency Care. Emerg Med Clin North Am. 1992; 103: 477–492.
26. **Harteloh PP.** The meaning of quality in health care: a conceptual analysis. Health Care Anal. 2003; 113: 259–267.
27. **Blumenthal D.** Part 1: Quality of care–what is it? N Engl J Med. 1996; 33512: 891–894.
28. **O'Neill O.** Trust with accountability? J Health Serv Res Policy. 2003; 81: 3–4.
29. **Wharam JF, Paasche-Orlow M, Sinsky C, Farber N, Rask K, Rucker L, Sulmasy D, Barry M, Figaro K.** Ethical Issues in Pay-for-Performance: Opinions of SGIM Members. Unpublished. 2008.
30. **Casalino LP.** The unintended consequences of measuring quality on the quality of medical care. N Engl J Med. 1999; 34115: 1147–1150.
31. **Casalino LP, Alexander GC, Jin L, Konetzka RT.** General internists' views on pay-for-performance and public reporting of quality scores: a national survey. Health Aff (Millwood). 2007; 262: 492–499.
32. **Casalino LP, Elster A, Eisenberg A, Lewis E, Montgomery J, Ramos D.** Will pay-for-performance and quality reporting affect health care disparities? Health Aff (Millwood). 2007; 263: w405–414.
33. **Fisher ES.** Paying for performance–risks and recommendations. N Engl J Med. 2006; 35518: 1845–1847.
34. **Hayward RA.** All-or-nothing treatment targets make bad performance measures. Am J Manag Care. 2007; 133: 126–128.
35. **Hayward RA.** Performance measurement in search of a path. N Engl J Med. 2007; 3569: 951–953.
36. **Wodchis WP, Ross JS, Detsky AS.** Is P4P really FFS? Jama. 2007; 29815: 1797–1799.
37. **Wachter RM, Shojania KG.** The unintended consequences of measuring quality on the quality of medical care. N Engl J Med. 2000; 3427: 520.
38. **Volandes AE, Paasche-Orlow MK.** Health literacy, health inequality and a just healthcare system. Am J Bioeth. 2007; 711: 5–10.
39. **Lo B, Groman M.** Oversight of quality improvement: focusing on benefits and risks. Arch Intern Med. 2003; 16312: 1481–1486.
40. **Bellin E, Dubler NN.** The quality improvement-research divide and the need for external oversight. Am J Public Health. 2001; 919: 1512–1517.
41. **Casarett D, Karlawish JH, Sugarman J.** Determining when quality improvement initiatives should be considered research: proposed criteria and potential implications. Jama. 2000; 28317: 2275–2280.
42. **Kofke WA, Rie MA.** Research ethics and law of healthcare system quality improvement: the conflict of cost containment and quality. Crit Care Med. 2003; 313 Suppl: S143–152.
43. **Perneger TV.** Why we need ethical oversight of quality improvement projects. Int J Qual Health Care. 2004; 165: 343–344.
44. **Nerenz DR, Stoltz PK, Jordan J.** Quality improvement and the need for IRB review. Qual Manag Health Care. 2003; 123: 159–170.
45. **Lynn J, Baily MA, Bottrell M, et al.** The ethics of using quality improvement methods in health care. Ann Intern Med. 2007; 1469: 666–673.
46. **Strasberg SM, Ludbrook PA.** Who oversees innovative practice? Is there a structure that meets the monitoring needs of new techniques? J Am Coll Surg. 2003; 1966: 938–948.
47. **Marron JM, Siegler M.** Ethical issues in innovative colorectal surgery. Dis Colon Rectum. 2005; 486: 1109–1113.
48. **Bunker JP, Hinkley D, McDermott WV.** Surgical innovation and its evaluation. Science. 1978; 2004344: 937–941.
49. **Cronin DC 2nd, Millis JM, Siegler M.** Transplantation of liver grafts from living donors into adults–too much, too soon. N Engl J Med. 2001; 34421: 1633–1637.
50. **Reitsma AM, Moreno JD.** Ethical regulations for innovative surgery: the last frontier? J Am Coll Surg. 2002; 1946: 792–801.
51. **Reitsma AM, Moreno JD.** Ethics of innovative surgery: US surgeons' definitions, knowledge, and attitudes. J Am Coll Surg. 2005; 2001: 103–110.
52. **Margo CE.** When is surgery research? Towards an operational definition of human research. J Med Ethics. 2001; 271: 40–43.
53. **Amoroso PJ, Middaugh JP.** Research vs. public health practice: when does a study require IRB review? Prev Med. 2003; 362: 250–253.
54. HCAHPS. CAHPS Hospital Survey. Available at: http://www.hcahpsonline.org/. Accessed February 10, 2008.
55. **Cheraghi-Sohi S, Hole AR, Mead N, et al.** What Patients Want From Primary Care Consultations: A Discrete Choice Experiment to Identify Patients' Priorities. Ann Fam Med. 2008; 6: 107–115.
56. **Rao M, Clarke A, Sanderson C, Hammersley R.** Patients' own assessments of quality of primary care compared with objective records based measures of technical quality of care: cross sectional study. Bmj. 2006; 3337557: 19.
57. Massachusetts Co. An Act Providing Access to Affordable, Quality, Accountable Health Care. Available at: http://www.mass.gov/legis/laws/seslaw06/sl060058.htm.
58. **Weinick RM, Alegria M, Coltin K, Gannon PM, Goel AK.** Disparities in Health Care in the Massachusetts Medicaid Program Recommendations of the Massachusetts Medicaid Disparities Policy Roundtable. Available at: http://www.cms.hhs.gov/MedicaidSCHIPQualPrac/Downloads/disp4p.pdf. Accessed November 24, 2008.
59. **Kent DM, Hayward RA.** Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. Jama. 2007; 29810: 1209–1212.
60. **Daniels N, Sabin JE.** Setting limits fairly : can we learn to share medical resources? Oxford, New York: Oxford University Press; 2002.

The Health and Health Care of US Prisoners: Results of a Nationwide Survey

Andrew P. Wilper, MD, MPH, Steffie Woolhandler, MD, MPH, J. Wesley Boyd, MD, PhD, Karen E. Lasser, MD, MPH, Danny McCormick, MD, MPH, David H. Bor, MD, and David U. Himmelstein, MD

Objectives. We analyzed the prevalence of chronic illnesses, including mental illness, and access to health care among US inmates.

Methods. We used the 2002 Survey of Inmates in Local Jails and the 2004 Survey of Inmates in State and Federal Correctional Facilities to analyze disease prevalence and clinical measures of access to health care for inmates.

Results. Among inmates in federal prisons, state prisons, and local jails, 38.5% (SE = 2.2%), 42.8% (SE = 1.1%), and 38.7% (SE = 0.7%), respectively, suffered a chronic medical condition. Among inmates with a mental condition ever treated with a psychiatric medication, only 25.5% (SE = 7.5%) of federal, 29.6% (SE = 2.8%) of state, and 38.5% (SE = 1.5%) of local jail inmates were taking a psychiatric medication at the time of arrest, whereas 69.1% (SE = 4.8%), 68.6% (SE = 1.9%), and 45.5% (SE = 1.6%) were on a psychiatric medication after admission.

Conclusions. Many inmates with a serious chronic physical illness fail to receive care while incarcerated. Among inmates with mental illness, most were off their treatments at the time of arrest. Improvements are needed both in correctional health care and in community mental health services that might prevent crime and incarceration. (*Am J Public Health.* 2009;99:666–672. doi: 10.2105/AJPH.2008.144279)

The prison population of the United States has quadrupled in the past 25 years, and the country now incarcerates more people per capita than any other nation. Worldwide, imprisonment per 100 000 ranges from 30 in India to 75 in Norway, 119 in China, 148 in the United Kingdom, 628 in Russia, and 750 in the United States.[1]

Currently, nearly 2.3 million US inmates (about 1% of US adults) must rely on their jailers for health care.[2] Although prisoners have a constitutional right to health care through the Eighth Amendment's prohibition of "cruel and unusual" punishment,[3] periodic scandals, as well as previous studies, indicate that prisoners' access to health care and the quality of that care are often deficient.[4,5] Indeed, citing deplorable conditions in California's prison system, a federal judge recently removed prison health care from the state's control.[6] However, there is little nationally representative data on the health and health care of America's prisoners.

Inmates have high rates of chronic medical conditions, especially viral infections. In addition, substance abuse and mental illness are common among inmates.[7,8] We are not aware of any study analyzing the prevalence of common chronic conditions or of access to medical and psychiatric care among the incarcerated population as a whole. Therefore, we sought to determine the prevalence of select chronic diseases, access to health services, and pre- and postincarceration psychiatric treatment among the US inmate population.

METHODS

We analyzed data from the 2004 Survey of Inmates in State and Federal Correctional Facilities (SISFCF) and the 2002 Survey of Inmates in Local Jails (SILJ). The US Census Bureau conducted these surveys for the Bureau of Justice Statistics. Participation in the surveys was voluntary, and prisoners' answers were anonymous and confidential.

Data Sources

The 2004 SISFCF consisted of in-person surveys of state and federal prisoners designed to provide nationally representative data on prison inmates. Between October 2003 and May 2004, inmates provided demographic, criminal justice, and health information to interviewers. The 2002 SILJ employed a virtually identical methodology and questionnaire.

The surveys employed a 2-stage sample design, selecting prisons in the first stage and inmates in the second stage. The Census Bureau preselected the 21 largest state prisons for inclusion in the survey. Remaining state prisons were stratified by census region; those with larger inmate populations were more likely to be included in the survey. Of 1585 state prisons, 301 were selected for participation in the SISFCF, of which 287 participated. Two prisons refused to participate, and 12 were deemed "out of scope": 2 were jails, 1 was under federal jurisdiction, 4 had closed, and 5 no longer housed inmates of the gender for which the facility was originally chosen. Of 16 152 randomly selected inmates, 14 499 completed interviews. The total response rate was 89.1%.

Three federal prisons were preselected. The remaining federal prisons were stratified by security level; those with larger inmate populations were more likely to be included in the survey. Of 148 eligible federal prisons, 40 were selected and 39 participated in the survey (1 prison refused to participate). A computer that was supplied with a list of all inmates selected inmates from within a facility using a random start point and a predetermined skip interval. Of 4253 randomly selected federal inmates, 3686 completed interviews. The total response rate was 84.6%.[9]

The Census Bureau conducted the SILJ from January to April 2002 using a similar 2-stage sample design. Researchers conducting the SILJ preselected 234 jails for inclusion to ensure that facilities with large numbers of men,

women, or juveniles had a higher probability of selection than would jails with smaller numbers of these individual groups. The remaining facilities were stratified by inmate population, and facilities housing larger inmate populations were more likely to be included in the survey. Of 3365 jails, 465, including those that were preselected, were systematically selected, and 417 participated in the survey; 39 refused to participate, and 9 had closed or housed no inmates. From within each institution, interviewers selected inmates using a predetermined random start and sample selection rate. Of 7750 randomly selected jail inmates, 6982 completed interviews. The total response rate was 84.1%.[10] For all 3 surveys, the Census Bureau provided weights that adjusted for nonresponse and sample design in order to yield national estimates.

For both the SISFCF and the SILJ, inmates answered questions about symptoms or medical diagnoses received prior to incarceration, including diabetes mellitus, hypertension, HIV/AIDS, paralysis, prior or current malignancy (breast, cervical, colon, leukemia, lung, ovarian, prostate, testicular, uterine, or other ["other" was not included in the local jail survey]), stroke or brain injury, angina, arrhythmia, arteriosclerosis, prior myocardial infarction, or other heart problem (coronary, congenital, rheumatic). Inmates also reported on persistent problems with kidneys, asthma, cirrhosis, hepatitis, arthritis, or sexually transmitted diseases. Surveyors did not use health records to confirm diagnoses.

Inmates were queried about serious injuries they had sustained since being incarcerated that were caused by an accident or a physical or sexual assault. We defined "serious injuries" as those resulting from knife or gunshot wounds and those causing broken bones, internal injuries, or loss of consciousness.

Inmates also answered questions about their health care since incarceration. Such care included tuberculosis skin test and treatment of a positive test, receipt of prescription medications before and after admission, blood tests (otherwise unspecified), and visits to a doctor, nurse, or other health care worker for a persistent health problem.

The SISFCF and SILJ assessed self-reported mental illnesses, including any prior diagnosis of depressive disorder, bipolar disorder, schizophrenia, posttraumatic stress disorder (PTSD), anxiety or panic disorder, personality disorder, or other mental condition. Inmates answered questions about medications for psychiatric illness at any point in the past, in the year prior to admission, at the time of arrest, and since incarceration. Inmates also reported mental health counseling at any time in the past, in the year prior to admission, at the time of arrest, or following admission.

We determined the self-reported prevalence of common chronic conditions that routinely require ongoing medical treatment, including diabetes mellitus, hypertension, prior myocardial infarction, persistent kidney problems, persistent asthma, cirrhosis, and HIV/AIDS.

In addition, we created another category defining inmates as having "any chronic condition" if they reported any condition likely requiring follow-up medical attention, even if not identified as causing a persistent problem by the inmate. In this category, we included a prior diagnosis of 1 or more of the following: diabetes mellitus, hypertension, HIV/AIDS, paralysis, prior malignancy (excluding skin cancers), prior stroke or brain injury, angina, arrhythmia, arteriosclerosis, myocardial infarction, other heart problems (coronary, congenital, rheumatic), persistent kidney problems, current problems with asthma, and persistent problems with cirrhosis, persistent hepatitis, and arthritis. The SISFCF included a question about "other" types of cancer, a question not included in the SILJ. "Other cancer" adds only 9160 and 704 individuals to state and federal "chronic" indicators, respectively. We did not include pregnancy or sexually transmitted diseases other than HIV/AIDS in our definition of "any chronic condition."

We compared the crude and age-adjusted prevalence of selected chronic conditions among inmates with the prevalence of such conditions among a nationally representative sample of the noninstitutionalized US population from the 2003–2004 National Health and Nutrition Examination Survey (NHANES).[11] The 2003–2004 NHANES included questions regarding a prior diagnosis of diabetes mellitus, hypertension, myocardial infarction, and persistent asthma that were nearly identical to those of the inmate surveys, and staff for the 2003–2004 NHANES tested participants aged 18 to 49 years for HIV. We included comparisons of both crude and age-adjusted prevalences of these chronic conditions among inmates and the nonincarcerated population.

Because most standard access to care measures, such as having a usual source of care or avoiding needed care because of costs, are meaningless in incarceration settings, we developed 5 clinically based access to care measures:

1. *Access to medical examinations.* To assess this measure, we created a marker for inmates with a persistent medical problem routinely requiring medical assessment. For this indicator, we first combined inmates reporting pregnancy at the time of admission with those reporting a persistent problem with diabetes mellitus, the heart or kidneys, hypertension, cancer, stroke or brain injury, paralysis, cirrhosis, arthritis, asthma, hepatitis, or a sexually transmitted disease. (Unfortunately, the surveys did not specifically assess access to care for inmates with HIV.) We then determined whether medical personnel had examined inmates for their persistent conditions at any time since incarceration.
2. *Access to pharmacotherapy.* To assess this measure, we first determined the number of inmates who had a condition routinely treated with pharmacotherapy (hypertension, diabetes mellitus, stroke or brain injury, persistent arthritis, asthma, cirrhosis, or HIV/AIDS) and had been taking a prescription medication at the time of admission. We then determined whether these prisoners continued taking that medication following incarceration. Surveyors did not collect medication names or query inmates about new medications begun during incarceration.
3. *Access to prescription medication.* To further assess access to prescription medication, we determined the number of inmates who had received any prescription drug for any indication prior to incarceration. We then determined the proportion of such inmates who did not receive that medication following incarceration.
4. *Access to laboratory tests.* To assess this measure, we defined prisoners as needing routine laboratory monitoring if they had 1 of the following conditions: diabetes mellitus, persistent hypertension, kidney

problems, cirrhosis, prior myocardial infarction, or HIV/AIDS. We then determined whether these prisoners had undergone at least 1 blood test of any kind since incarceration.

5. *Adequacy of acute care.* To assess this measure, we analyzed data from inmates with a serious injury (knife or gunshot wounds, broken bones, internal injuries, being knocked unconscious, or sexual assault). We then determined whether these prisoners received any medical examination for their injuries.

Finally, we focused on receipt of mental health care. For inmates reporting any prior diagnosis of a mental condition, we determined the proportion ever receiving a medication for that condition. Next, we determined the proportion of this population taking medication at the time of arrest and since incarceration. We also determined the proportion of inmates with any history of a mental condition who had ever received counseling, who had received counseling in the year prior to admission, and who had received counseling since incarceration. Finally, we repeated all mental health analyses using only those inmates with a prior diagnosis of bipolar disorder or schizophrenia.

Statistical Analysis

We used SAS version 9.1 (SAS Institute Inc, Cary, NC) to analyze bivariate relationships. We used SUDAAN version 9.0.3 (Research Triangle Institute, Research Triangle Park, NC) to estimate variance via restricted-access SILJ design variables. For the SISFCF, we calculated variance using the generalized variance estimates available with the survey documentation. We applied sample weights supplied by the Bureau of Justice Statistics to account for nonresponse and survey design and to yield national estimates. We performed direct age standardization via published techniques.[12]

RESULTS

Based on our analysis, US federal prisons held 129196 inmates and state prisons 1225680 in 2004. In 2002, local jails held 631241 inmates. The overwhelming majority of inmates were male, were younger than 35 years, and were disproportionately Black or Hispanic. About 200000 (10%) were military veterans. The majority were parents of minor children at the time of incarceration or at the time of the survey.

Nonresponse to individual items was uncommon. Among federal inmates, 2.1% were missing data on prescription medications at admission and 2.8% on prior diagnosis of PTSD; 6.0% were missing data for HIV testing and 15.8% for duration of incarceration. No data were provided for sexual assault or gunshot wounds in federal prisons. Among state inmates, 1.2% were missing data on prescription medications at admission and 1.7% on prior diagnosis of PTSD; 4.0% were missing data regarding HIV testing and 6.3% for duration of incarceration. Among jail inmates, 0.5% were missing data on the duration of incarceration and 2.2% on prior diagnosis of PTSD; 5.2% were missing data on HIV testing.

Chronic Medical Problems

Chronic conditions were common among inmates; 49702 federal inmates (38.5% [SE=2.2%]), 524116 state inmates (42.8% [SE=1.1%]), and 244336 local jail inmates (38.7% [SE=0.7%]) had at least 1 chronic medical condition (Table 1).

Inmates had rates of diabetes, hypertension, prior myocardial infarction, and persistent asthma comparable to those of the US noninstitutionalized, nonelderly population. However, following age standardization to the 2000 US census, the prevalence of these conditions appeared to be higher for inmates than for the general population, except for prior myocardial infarction among jail inmates (Table 2; see also the appendix to Table 1, available as a supplement to the online version of this article at http://www.ajph.org). More than 20000 inmates reported testing positive for HIV,

TABLE 1—Demographic and Health Characteristics of Inmates in US Federal and State Prisons and in Jails: SISFCF, 2004, and SILJ, 2002

	Federal Inmates		State Inmates		Jail Inmates	
	No.	% (SE)	No.	% (SE)	No.	% (SE)
Total	129196	100	1225680	100	631241	100
Men	120150	93.0 (0.6)	1142989	93.3 (0.4)	558182	88.4 (0.3)
Age, y						
13–35	64692	50.1 (2.0)	654505	53.4 (1.0)	408321	64.7 (0.7)
36–50	50180	38.8 (2.2)	465874	38.0 (1.1)	196420	31.1 (0.7)
>50	14324	11.1 (2.7)	105302	8.6 (1.4)	26500	4.2 (0.3)
Parent of minor child[a]	87618	67.8 (1.6)	706942	57.7 (0.9)	355963	56.4 (0.7)
Race						
Non-Hispanic White	33599	26.0 (2.4)	431449	35.2 (1.2)	226209	35.8 (1.1)
Non-Hispanic Black	55947	43.3 (2.1)	496745	40.5 (1.1)	252116	39.9 (1.2)
Hispanic	32414	25.1 (2.1)	222451	18.2 (1.3)	116316	18.4 (0.9)
Other	7235	5.5 (2.8)	75036	6.1 (1.4)	36600	5.8 (0.4)
Military veteran	12562	9.7 (2.7)	127509	10.4 (1.4)	58761	9.3 (0.5)
Any mental health condition[b]	19117	14.8 (2.6)	312768	25.5 (1.3)	157634	25.0 (0.7)
Any chronic medical condition[c]	49702	38.5 (2.2)	524116	42.8 (1.1)	244336	38.7 (0.7)

Note. SISFCF=Survey of Inmates in State and Federal Correctional Facilities; SILJ=Survey of Inmates in Local Jails. Median duration of incarceration in months (interquartile range) was as follows: for federal inmates, 29 (12–61); for state inmates, 27 (9–67); for jail inmates, 2 (0–4).

[a]Defined as being a parent at time of survey or during incarceration.

[b]Defined as having a prior diagnosis of depressive disorder, bipolar disorder, schizophrenia, posttraumatic stress disorder, anxiety disorder, panic disorder, personality disorder, or other mental health condition.

[c]A chronic condition was defined as affirmative response when asked about the following: HIV/AIDS; prior malignancy (excluding skin cancers) including breast, cervical, colon, leukemia, lung, ovarian, prostate, testicular, uterine, and other ("other" not included in the jail group); hypertension; stroke or brain injury; angina; arrhythmia; arteriosclerosis; myocardial infarction; other heart problem (coronary, congenital, rheumatic); persistent kidney problems; persistent paralysis; current problems with asthma; cirrhosis; persistent hepatitis; persistent arthritis.

TABLE 2—Age-Standardized Prevalence of Select Chronic Conditions Among Adult Federal and State Prisoners, Jail Inmates, and the Noninstitutionalized US Population: SISFCF, 2004, SILJ, 2002, and NHANES, 2003–2004

Condition	Federal Inmates, % (SE)	State Inmates, % (SE)	Jail Inmates, % (SE)	US Population,[a] % (SE)
Diabetes mellitus	11.1 (3.6)	10.1 (2.0)	8.1 (1.7)	6.5 (0.5)
Hypertension	29.5 (2.9)	30.8 (1.5)	27.9 (2.1)	25.6 (1.0)
Prior myocardial infarction	4.5 (4.5)	5.7 (2.8)	2.1 (0.4)	3.0 (0.3)
Persistent kidney problems	6.3 (4.0)	4.5 (1.7)	4.1 (0.8)	. . .
Persistent asthma	7.7 (2.8)	9.8 (1.4)	8.6 (1.0)	7.5 (0.6)
Persistent cirrhosis	2.2 (3.9)	1.8 (1.8)	1.8 (0.7)	. . .
Persistent hepatitis	4.6 (2.9)	5.7 (1.5)	4.6 (1.4)	. . .
HIV[b]	0.9 (3.2)	1.7 (1.8)	1.6 (0.3)	0.5 (0.1)

Note. SISFCF = Survey of Inmates in State and Federal Correctional Facilities; SILJ = Survey of Inmates in Local Jails; NHANES = National Health and Nutrition Examination Survey. Prevalence was standardized to the 2000 US population 18 years and older by direct age standardization. Inmates younger than 18 years represented 0% of federal inmates, less than 1% of state inmates, and 4.8% of jail inmates.

[a]The 2003–2004 NHANES did not include questions regarding persistent kidney problems, cirrhosis, and hepatitis.

[b]For HIV, only populations aged 18–49 years are included to allow comparison with NHANES data, which was derived from laboratory data.

including 1023 federal inmates (1.0% [SE=3.1%]), 15 115 state inmates (1.6% [SE=1.6%]), and 4245 local jail inmates (1.2% [SE=0.2%]); this prevalence was double that of the noninstitutionalized 2003–2004 NHANES population. These percentages did not substantially change when only inmates aged 18–49 years (the age group that underwent HIV testing in the NHANES sample) were included.

Access to Medical Services

Among inmates with a persistent medical problem, 13.9% of federal inmates, 20.1% of state inmates, and 68.4% of local jail inmates had received no medical examination since incarceration. More than 1 in 5 inmates were taking a prescription medication for some reason when they entered prison or jail; of these, 7232 federal inmates (26.3%), 80 971 state inmates (28.9%), and 58 991 local jail inmates (41.8%) stopped the medication following incarceration. Prior to incarceration, slightly more than 1 in 7 inmates were taking a prescription medication for an active medical problem routinely requiring medication (as defined in the Methods section). Of these, 3314 federal (20.9% [SE= 6.7%]), 43 679 state (24.3% [SE=3.3%]), and 28 473 local jail inmates (36.5% [SE=1.7%]) stopped the medication following incarceration.

Only a small portion of prison inmates (3.9% [SE=6.5%] of federal and 6.4% [SE=3.2%] of state inmates) with an active medical problem for which laboratory monitoring is routinely indicated had not undergone at least 1 blood test since incarceration. However, most local jail inmates with such a condition (60.1% [SE=1.8%]) had not undergone a blood test.

Following serious injury, 650 federal inmates (7.7%), 12 997 state inmates (12.0%), and 3183 local jail inmates (24.7%) were not seen by medical personnel (Table 3).

Mental Health

Mental health problems were ubiquitous: 19 117 federal inmates (14.8% [SE=2.6%]), 312 768 state inmates (25.5% [SE=1.3%]), and 157 634 local jail inmates (25.0% [SE=0.7%]) had at least 1 previously diagnosed mental condition (Table 1); most of them had taken medications at some point prior to incarceration. However, a much smaller proportion of inmates with a mental health diagnosis were taking psychiatric medication at the time of their arrest: 25.5% (SE=7.5%) of federal, 29.6% (SE=2.8%) of state, and 38.5% (SE=1.5%) of local jail inmates. Among inmates with a previously diagnosed mental condition who had been treated with a psychiatric medication in the past, 69.1% (SE=4.8%) of federal, 68.6% (SE=1.9%) of state, and 45.5% (SE=1.6%) of local jail inmates had taken a medication for a mental condition since incarceration. A similar pattern was apparent for prearrest and postincarceration counseling (Table 4).

Among prison inmates with schizophrenia or bipolar disorder who had ever been treated with psychiatric medication, the proportion on treatment was approximately 1 in 3 at the time of arrest and nearly 2 in 3 during incarceration (see appendix to Table 2, available as a supplement to the online version of this article at http://www.ajph.org). Among jail inmates with schizophrenia or bipolar disorder, the pattern of low treatment rates at arrest and high treatment rates following incarceration was also present, although less pronounced than in the prison population.

DISCUSSION

Mass incarceration as part of the war on drugs has created a burgeoning inmate population in the United States. Earlier studies of inmates have been based on extrapolations from noninstitutionalized Americans, single institutions, or data from either federal or state prisons alone or jail systems alone. Our study adds to the existing literature by analyzing a large, nationally representative sample of the entire US inmate population. More than 800 000 inmates report having 1 or more chronic medical condition, and their access to medical care appears to be poor, particularly in jails. Our data also demonstrate that prisons are holding and treating many mentally ill people who were off treatment at the time of arrest.

Our age-standardized prevalence estimates for rates of hypertension and diabetes were higher than estimates from earlier population-based projection models (18.3% and 4.8%, respectively).[13] Although the rates of asthma in our study were similar to the rates in the earlier study (8.5%),[13] our figures include only those with active asthma, whereas the earlier estimates included any prior diagnosis. Furthermore, the earlier projections were based on models that used data from NHANES III that included laboratory testing (diabetes) and physical examination (hypertension) as part of diagnostic criteria;

TABLE 3—Access to Medical Care for Inmates of Federal Prisons, State Prisons, and Local Jails: SISFCF, 2004, and SILJ, 2002

Condition	Federal Inmates, No. or % (SE)	State Inmates, No. or % (SE)	Jail Inmates, No. or % (SE)
Persistent medical problem[a]			
Inmates with problem	43 059	465 682	214 812
Inmates with problem not examined by medical personnel	13.9 (4.5)	20.1 (2.1)	68.4 (1.1)
Active medical problem requiring prescription medication[b]			
Inmates on prescription medication at time of incarceration	18 728	181 994	90 283
Inmates not continued on same medication during incarceration	20.9 (6.7)	24.3 (3.3)	36.5 (1.7)
Prescription drug use			
Inmates on prescription drugs at time of incarceration	27 522	280 036	141 133
Inmates not continued on medication during incarceration	26.3 (4.9)	28.9 (2.6)	41.8 (1.4)
Active medical problem routinely requiring blood test[c]			
Inmates with problem	23 467	240 960	106 539
Inmates with problem but with no blood tests since admission[d]	3.9 (6.5)	6.4 (3.2)	60.1 (1.8)
Serious injury[e]			
Inmates with serious injury, no.	8 431	107 989	12 887
Inmates not examined following serious injury, % (SE)	7.7 (10.6)	12.0 (4.6)	24.7 (3.9)

Note. SISFCF = Survey of Inmates in State and Federal Correctional Facilities; SILJ = Survey of Inmates in Local Jails.
[a]Persistent medical problems included pregnancy at time of admission, diabetes mellitus, persistent heart or kidney problems, persistent hypertension, cancer, stroke or brain injury, paralysis, cirrhosis, arthritis, asthma, hepatitis, or a sexually transmitted disease.
[b]Active medical problems included hypertension, stroke, diabetes mellitus, heart problem, kidney, arthritis, asthma, hepatitis, cirrhosis, and HIV/AIDS.
[c]Active medical problems routinely requiring blood tests included diabetes, persistent kidney problems, HIV, persistent hypertension, prior myocardial infarction, and cirrhosis.
[d]Defined as inmates who probably needed blood testing but had not received any since incarceration.
[e]Serious injuries included knife or gunshot wounds, broken bones, sexual assault, internal injuries, and being knocked unconscious. Responses to sexual assault were missing for federal inmates in the SISFCF.

including these measurements as part of the diagnostic criteria among inmates would have increased our prevalence estimates.[13]

Improved management of chronic conditions in prisons and jails may have important implications for community health and in reducing health care disparities, because the vast majority of inmates are eventually released. Approximately 12 million inmates are released annually (William J. Sabol, PhD, chief, Corrections Statistics, Bureau of Justice Statistics, oral communication, April 2008). This high turnover of a population with elevated rates of treatable conditions offers a substantial public health opportunity. Indeed, in response to a congressional request, the National Commission on Correctional Health Care issued an extensive report in 2002 titled *The Health Status of Soon-To-Be-Released Inmates*[8]; although it included recommendations of specific strategies to improve inmates' health, no congressional action has ensued (R. Scott Chavez, PhD, MPA, vice president, National Commission on Correctional Health Care, oral communication, July 2008). Nonetheless, minimizing inmates' physical and mental disability is an important step in reintegrating them into family and employment roles.

The prevalence of HIV in prisons is higher than in the noninstitutionalized population, although it is declining.[14,15] A high incidence of blood-borne illnesses among inmates has also been documented.[16,17] Limited privacy in prison may make prisoners reluctant to comply with treatment of HIV, and sexual coercion and bartering may facilitate transmission. Similarly, untreated bleeding injuries (as documented in our data) pose an obvious transmission risk. Hence, poorly managed HIV may lead prisons to function as "amplifiers" of this and other infectious illnesses and add to the burden of untreated and advanced disease borne by inmates, families, and communities following inmates' release.

We estimate that nearly 500 000 inmates have a previously diagnosed mental condition. Moreover, Bureau of Justice Statistics estimates that include undiagnosed symptoms of mental health problems (such as hallucinations) suggest that the number of inmates with a psychiatric illness may be even higher.[18] The rates of mental illness among inmates are thought to be higher than among the US population as a whole. Although we did not directly compare rates of mental illness among inmates and the general US population, our estimates were derived directly from inmates, as opposed to a representative sampling of unincarcerated Americans.[13]

Sadly, in the United States, many inmates do not receive psychiatric treatment at the time of arrest, even those with schizophrenia or bipolar disorder. However, the low rate of treatment of inmates prior to arrest could be viewed as hopeful news, implying that greater access to outpatient mental health care might reduce the staggering toll of crime and incarceration.[19]

As with indicators for access to medical care, access to psychiatric care appears to be worse in jails than in prisons. The jump in rates of psychiatric treatment during incarceration may reflect limited access to psychiatric treatment among those with mental disorders prior to incarceration, and prisons' new societal role as asylums following the mass closures of inpatient mental health facilities in the 1980s (the largest mental institutions in the United States are urban jails[7]); conversely, psychiatric medications may be overprescribed in prisons. Furthermore, the use of psychiatric medication is measured differently than that of other prescription drugs. The increase in counseling from prearrest to incarceration supports the notion that a genuine improvement in the availability of psychiatric care occurs during incarceration.

Vast improvements in inmate health care are possible. Salutary reforms could include decreasing incarceration rates; making health care systems in prison nonprofit and autonomous from prison authorities; increasing communicable disease education, prevention, and treatment[20–22]; making condoms available[23]; improving care for chronic conditions; providing targeted cancer screening[24]; increasing the

TABLE 4—Prevalence of Diagnosed Mental Conditions Among Inmates of State and Federal Prisons and Local Jails, and Use of Psychiatric Medications and Counseling Before and During Incarceration: SISFCF, 2004, and SILJ, 2002

	Federal Inmates		State Inmates		Jail Inmates	
	No.	% (SE)	No.	% (SE)	No.	% (SE)
Any diagnosed mental condition	19 117	14.8 (2.6)	312 768	25.5 (1.3)	157 634	25.0 (0.7)
Medication[a]						
Ever took medication for emotional or mental problem[b]	13 674	71.6 (3.9)	233 456	74.6 (1.5)	116 011	73.7 (1.1)
Was taking medication at time of arrest[c]	3 481	25.5 (7.5)	69 088	29.6 (2.8)	44 526	38.5 (1.5)
Taking medication since admission[c]	9 455	69.1 (4.8)	160 048	68.6 (1.9)	52 755	45.5 (1.6)
Counseling[a]						
Ever received for mental or emotional problem	12 140	63.6 (4.4)	196 494	62.9 (1.8)	99 906	63.4 (1.3)
Received at any time during 12 mo before arrest[d]	3 754	30.9 (7.6)	66 578	33.9 (3.0)	43 007	43.1 (1.6)
Received since admission[d]	7 090	58.4 (5.9)	126 049	64.2 (2.2)	24 146	24.2 (1.5)

Note. SISFCF = Survey of Inmates in State and Federal Correctional Facilities; SILJ = Survey of Inmates in Local Jails. Mental conditions included prior diagnosis of depressive disorder, bipolar disorder, schizophrenia, posttraumatic stress disorder, anxiety or panic disorder, personality disorder, or other mental condition.
[a]Among those with a mental condition.
[b]Prescribed by a psychiatrist or other doctor.
[c]Among those who were ever prescribed a medication.
[d]Among those who had ever received counseling.

availability of addiction and mental health treatment; providing better supervision to reduce physical and sexual assault; maintaining Medicaid eligibility for inmates[8]; and improving the planning of inmates' discharge and facilitating their reintegration into the community.[25–27]

Limitations

Although access to care in local jails appears to be worse than in federal and state prisons, this result may simply reflect the shorter duration of incarceration among jail inmates. We were unable to validate inmates' responses; however, the anonymous and confidential nature of the survey should have maximized inmates' candor. It is possible that some inmates who reported taking prescription medications that were discontinued at the time of incarceration had actually been switched to a therapeutic equivalent but did not recognize it as such or had a condition that no longer warranted treatment at admission. Furthermore, among those with chronic conditions, no assessment of medications begun following incarceration was possible. Although our measures of access to care among inmates have not been validated, we believe that they have sufficient face validity to support a presumption that health care in prisons and jails is far from adequate. Unfortunately, we have no information on the quality of pharmacological and other medical care. Hence, our data refer only to the most minimal standards of care (i.e., any medical evaluation, any testing, or any treatment).

Conclusions

Providing inmates with health care is politically unpopular. Indeed, former Surgeon General Richard H. Carmona stated that the Bush administration had blocked the release of the Surgeon General's Report, *Call to Action on Corrections in Community Health*, for fear that the report would increase government spending on inmates.[28] However, the constitutional, public health, and human rights imperatives of improving health care in prisons and jails are clear. ■

About the Authors

At the time of this study, Andrew P. Wilper, Steffie Woolhandler, Karen E. Lasser, Danny McCormick, David H. Bor, and David U. Himmelstein were with the Department of Medicine, Cambridge Health Alliance, Cambridge, MA. J. Wesley Boyd is with the Department of Psychiatry, Cambridge Health Alliance, Cambridge.

Requests for reprints should be sent to Andrew P. Wilper, MD, MPH, Boise Veterans Affairs Medical Center, 500 West Fort St, Boise, ID 83702 (e-mail: wilp9522@u.washington.edu).

This article was accepted August 1, 2008.

Contributors

A.P. Wilper designed the study, planned the analysis, performed statistical analysis and data management, and interpreted the analysis. A.P. Wilper, S. Woolhandler, and D.U. Himmelstein drafted the article. J.W. Boyd, K.E. Lasser, D. McCormick, and D.H. Bor performed critical revisions of the article. S. Woolhandler supervised all aspects of the study design, analysis planning, interpretation, and article preparation.

Acknowledgments

Andrew P. Wilper was supported by the Health Resources and Service Administration National Research Service Award (5T32 HP110011).

We thank Robert Griefinger, MD, and William J. Sabol, PhD, for their comments on earlier drafts of the article.

Human Participant Protection

The institutional review board of the Cambridge Health Alliance approved this study.

References

1. King's College London. International Centre for Prison Studies. Entire world—prison population rates per 100,000 of the national population. Available at: http://www.prisonstudies.org. Accessed February 8, 2008.
2. Bureau of Justice Statistics, US Department of Justice. Prison statistics, summary findings. Available at: http://www.ojp.usdoj.gov/bjs/prisons.htm. Accessed February 8, 2008.
3. *Estelle v Gamble*, 429 US 97 (1976).
4. Clemmitt M. Prison health care. *Congr Q.* 2007;17:1–24.
5. Lindquist CH, Lindquist CA. Health behind bars: utilization and evaluation of medical care among jail inmates. *J Community Health.* 1999;24(4):285–303.
6. Moore S. Using muscle to improve health care for prisoners. *New York Times.* August 27, 2007. Available at: http://www.nytimes.com/2007/08/27/us/27prisons.html. Accessed November 18, 2008.
7. Freudenberg N. Jails, prisons, and the health of urban populations: a review of the impact of the correctional system on community health. *J Urban Health.* 2001;78(2):214–235.
8. National Commission on Correctional Health Care. *The Health Status of Soon-to-Be-Released Inmates, Vol. 1.* Available at: http://www.ncchc.org/pubs/pubs_stbr.vol1.html. Accessed July 10, 2008.
9. Bureau of Justice Statistics, US Department of Justice. Survey of inmates in state and federal correctional facilities, 2004. Available at: http://www.icpsr.umich.edu/cocoon/NACJD/STUDY/04572.xml. Accessed August 20, 2007.
10. Bureau of Justice Statistics, US Department of Justice. Survey of inmates in local jails, 2002. Available

at: http://www.icpsr.umich.edu/cocoon/NACJD/STUDY/04359.xml. Accessed August 30, 2007.

11. National Center for Health Statistics, Centers for Disease Control and Prevention. Survey operations manuals, brochures, and consent documents: 1999–current National Health and Nutrition Examination Survey. Available at: http://www.cdc.gov/nchs/about/major/nhanes/currentnhanes.htm. Accessed April 2, 2008.

12. Klein RJ, Schoenborn CA. *Age Adjustment Using the 2000 Projected US Population.* Hyattsville, MD: National Center for Health Statistics; 2001. Healthy People Statistical Notes no. 20.

13. National Commission on Correctional Health Care. *The Health Status of Soon-To-Be-Released Inmates, Vol. 2.* Available at: http://www.ncchc.org/pubs/pubs_stbr.vol2.html. Accessed July 10, 2008.

14. Centers for Disease Control and Prevention. A glance at the HIV/AIDS epidemic. Available at: http://www.cdc.gov/hiv/resources/factsheets/At-A-Glance.htm#1. Accessed February 10, 2008.

15. Maruschak LM. HIV in prisons, 2004. Available at: http://www.ojp.usdoj.gov/bjs/pub/pdf/hivp04.pdf. Accessed February 10, 2008.

16. Macalino GE, Vlahov D, Sanford-Colby S, et al. Prevalence and incidence of HIV, hepatitis B virus, and hepatitis C virus infections among males in Rhode Island prisons. *Am J Public Health.* 2004;94(7):1218–1223.

17. Khan AJ, Simard EP, Bower WA, et al. Ongoing transmission of hepatitis B virus infection among inmates at a state correctional facility. *Am J Public Health.* 2005;95(10):1793–1799.

18. James DJG, Lauren E. Mental health problems of prison and jail inmates. Available at: http://www.ojp.usdoj.gov/bjs/pub/pdf/mhppji.pdf. Accessed February 10, 2008.

19. Lamberti JS, Weisman R, Faden DI. Forensic assertive community treatment: preventing incarceration of adults with severe mental illness. *Psychiatr Serv.* 2004;55(11):1285–1293.

20. Restum ZG. Public health implications of substandard correctional health care. *Am J Public Health.* 2005;95(10):1689–1691.

21. Glaser JB, Greifinger RB. Correctional health care: a public health opportunity. *Ann Intern Med.* 1993;118(2):139–145.

22. MacNeil JR, McRill C, Steinhauser G, Weisbuch JB, Williams E, Wilson ML. Jails, a neglected opportunity for tuberculosis prevention. *Am J Prev Med.* 2005;28(2):225–228.

23. Berkman A. Prison health: the breaking point. *Am J Public Health.* 1995;85(12):1616–1618.

24. Binswanger IA, White MC, Perez-Stable EJ, Goldenson J, Tulsky JP. Cancer screening among jail inmates: frequency, knowledge, and willingness. *Am J Public Health.* 2005;95(10):1781–1787.

25. Binswanger IA, Stern MF, Deyo RA, et al. Release from prison—a high risk of death for former inmates. *N Engl J Med.* 2007;356(2):157–165.

26. Jacobi JV. Prison health, public health: obligations and opportunities. *Am J Law Med.* 2005;31(4):447–478.

27. Lee J, Vlahov D, Freudenberg N. Primary care and health insurance among women released from New York City jails. *J Health Care Poor Underserved.* 2006;17(1):200–217.

28. Donnelly J. Inmate health program: a model. *Boston Globe.* September 3, 2007. Available at: http://www.boston.com/news/local/massachusetts/articles/2007/09/03/inmate_health_program_a_model/?page=2. Accessed on November 18, 2008.

Health Insurance and Mortality in US Adults

Andrew P. Wilper, MD, MPH, Steffie Woolhandler, MD, MPH, Karen E. Lasser, MD, MPH, Danny McCormick, MD, MPH, David H. Bor, MD, and David U. Himmelstein, MD

Objectives. A 1993 study found a 25% higher risk of death among uninsured compared with privately insured adults. We analyzed the relationship between uninsurance and death with more recent data.

Methods. We conducted a survival analysis with data from the Third National Health and Nutrition Examination Survey. We analyzed participants aged 17 to 64 years to determine whether uninsurance at the time of interview predicted death.

Results. Among all participants, 3.1% (95% confidence interval [CI]=2.5%, 3.7%) died. The hazard ratio for mortality among the uninsured compared with the insured, with adjustment for age and gender only, was 1.80 (95% CI=1.44, 2.26). After additional adjustment for race/ethnicity, income, education, self- and physician-rated health status, body mass index, leisure exercise, smoking, and regular alcohol use, the uninsured were more likely to die (hazard ratio=1.40; 95% CI=1.06, 1.84) than those with insurance.

Conclusions. Uninsurance is associated with mortality. The strength of that association appears similar to that from a study that evaluated data from the mid-1980s, despite changes in medical therapeutics and the demography of the uninsured since that time. (*Am J Public Health.* 2009;99:2289–2295. doi:10.2105/AJPH.2008.157685)

The United States stands alone among industrialized nations in not providing health coverage to all of its citizens. Currently, 46 million Americans lack health coverage.[1] Despite repeated attempts to expand health insurance, uninsurance remains commonplace among US adults.

Health insurance facilitates access to health care services and helps protect against the high costs of catastrophic illness. Relative to the uninsured, insured Americans are more likely to obtain recommended screening and care for chronic conditions[2] and are less likely to suffer undiagnosed chronic conditions[3] or to receive substandard medical care.[4]

Numerous investigators have found an association between uninsurance and death.[5–14] The Institute of Medicine (IOM) estimated that 18 314 Americans aged between 25 and 64 years die annually because of lack of health insurance, comparable to deaths because of diabetes, stroke, or homicide in 2001 among persons aged 25 to 64 years.[4] The IOM estimate was largely based on a single study by Franks et al.[5] However, these data are now more than 20 years old; both medical therapeutics and the demography of the uninsured have changed in the interim.

We analyzed data from the Third National Health and Nutrition Examination Survey (NHANES III). NHANES III collected data on a representative sample of Americans, with vital status follow-up through 2000. Our objective was to evaluate the relationship between uninsurance and death.

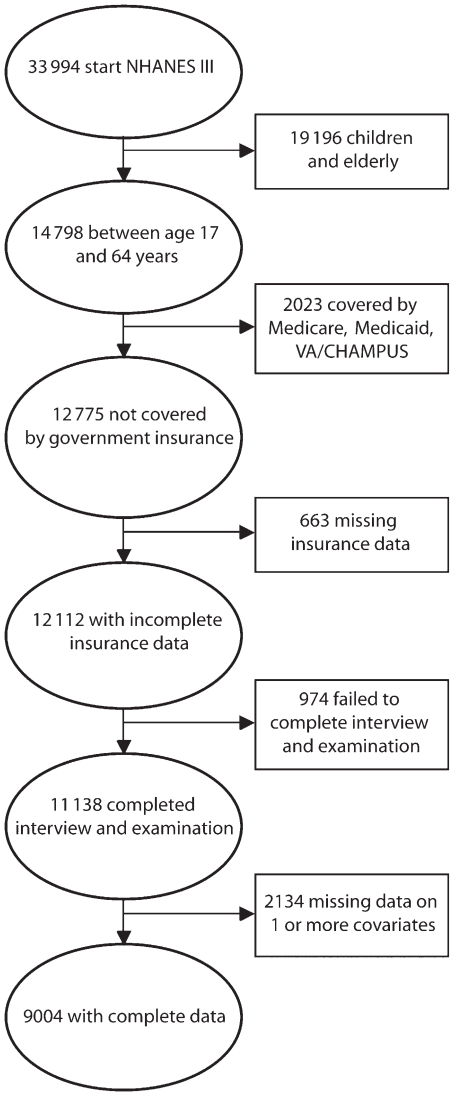
METHODS

The National Center for Health Statistics (NCHS) conducted NHANES III between 1988 and 1994. The survey combined an interview, physical examination, and laboratory testing. NHANES III employed a complex sampling design to establish national estimates of disease prevalence among the noninstitutionalized civilian population in the United States.[15] Staff performed interviews in English and Spanish.

The NHANES III Linked Mortality File matched NHANES III records to the National Death Index (NDI). The NCHS's linkage, which uses a probabilistic matching strategy through December 31, 2000, is described elsewhere.[16] The NCHS perturbed the file to prevent reidentification of survey participants. Vital status was not altered in this process. The publicly released data yield survival analysis results virtually identical to the restricted-use NHANES III Linked Mortality File.[17]

In designing our analysis, we hewed closely to Franks'[5] methodology to facilitate interpretation of time trends. We analyzed data for individuals who reported no public source of health insurance at the time of the NHANES III interview. First, we excluded those aged older than 64 years, as virtually all are eligible for Medicare. Of the 33 994 individuals participating, 14 798 were aged between 17 and 64 years at the time of the interview. In keeping with earlier analyses,[5–7,13] we also excluded nonelderly Medicare recipients and persons covered by Medicaid and the Department of Veterans Affairs/Civilian Health and Medical Program of the Uniformed Services military insurance (n=2023), as a substantial proportion of those individuals had poor health status as a prerequisite for coverage. Of the 12 775 participants not covered by government insurance, we excluded 663 (5.2%) who lacked information on health insurance. We excluded 974 of the remaining 12 112 who were covered by private insurance or uninsured at the time of the interview because of failure to complete the interview and physical examination. Of the remaining 11 138, we included only the 9005 with complete baseline data from both the interview and physical examination in our final analysis (Figure 1). Among those with complete insurance data, those with complete interview and examination data were both less likely to be uninsured (16.4% vs 21.6%; $P<.001$) and less likely to die (3.0% vs 4.5%; $P<.001$).

NHANES III staff interviewed respondents in their homes regarding demographics (including health insurance). Participants responded to questions about race, ethnicity, income, and household size. The sample design permits estimation for 3 racial/ethnic groups: non-Hispanic White, non-Hispanic Black, and



Note. NHANES III = National Health and Nutrition Examination Survey; VA/CHAMPUS = Veterans Affairs/Civilian Health and Medical Program of the Uniformed Services.

FIGURE 1—Study population and exclusions.

Mexican American. The NCHS created a variable that combined family income and the poverty threshold during the year of interview (the poverty income ratio), allowing income to be standardized for family size and compared across the 6 years of data collection.[18]

NHANES III interviewers also collected data on education, employment, tobacco use, alcohol use, and leisure exercise. We analyzed education dichotomously, comparing those with 12 years or more education to those with less than 12 years. We considered respondents to be unemployed if they were looking for work, laid off, or unemployed. All others, including the employed, students, homemakers, and retirees were considered "not unemployed." We considered smokers in 3 categories: current smokers, former smokers (those who had smoked more than 200 cigarettes in their lifetime), and nonsmokers. We labeled those drinking more than 6 alcoholic beverages per week as regular drinkers. We analyzed exercise in 2 groups: those achieving greater than or equal to 100 metabolic equivalents (METs) per month, versus those achieving less than 100 METs per month.[19,20]

NHANES III measured participants' self-perceived health in 5 categories: excellent, very good, good, fair, and poor. We combined the last 2 groups because of small numbers. NHANES physicians performed physical examinations on all participants and provided an impression of overall health status rated as excellent, very good, good, fair, and poor.[21] We combined the final 2 groups because of small numbers. We analyzed body mass index (BMI; weight in kilograms divided by height in meters squared) in 4 categories: less than 18.5; 18.5 to 25; more than 25 to less than 30; and 30 and higher.

NHANES III oversampled several groups, including Black persons, Mexican Americans, the very young (aged 2 months to 5 years), and those aged older than 65 years. To account for this and other design variables we used the SUDAAN (version 9.1.3, Research Triangle Institute, Research Triangle Park, NC) SURVIVAL procedure and SAS (version 9.1, SAS Institute Inc, Cary, NC) PROC SURVEYFREQ to perform all analyses. We (as did Franks et al.[5]) employed unweighted survival analyses and controlled for the variables used in determining the sampling weights (age, gender, and race/ethnicity) because of the inefficiency of weighted regression analyses.[22]

We analyzed the relation between insurance, demographics, baseline health status variables, and mortality by using χ^2 tests. We then used a Cox proportional hazards survival analysis controlling only for age and gender to determine if lack of health insurance predicted mortality. We repeated the analysis of the relationship of insurance to mortality after forcing all covariates in the model. In this Cox proportional hazards analysis, we controlled for gender, age, race/ethnicity (4 categories), income (poverty income ratio), education, current unemployment, smoking status (3 categories), regular alcohol use, self-rated health (4 categories), physician-rated health (4 categories), and BMI (4 categories). We tested for significant interactions between these variables and health insurance status (i.e., $P<.05$). We handled tied failure times by using the Efron method.

We performed multiple sensitivity analyses to analyze the robustness of our results.

We developed a propensity score model and controlled for the variables in our previous models (with the exception of health insurance status), as well as marital status; household size; census region; number of overnight visits in hospital in past 12 months; number of visits to a physician in past 12 months; limitations in work or activities; job or housework changes or job cessation because of a disability or health problem; and number of self-reported chronic diseases, including emphysema, prior nonskin malignancy, stroke, congestive heart failure, hypertension, diabetes, or hypercholesterolemia. Next, we included the propensity score in the multivariable model with the indicator for insurance status. In addition, we tested for the effect of including those covered by Medicaid by using our original Cox model and the propensity score adjusted analysis. In a subsidiary analysis, we excluded employment and self- and physician-rated health, as these covariates may be a result of limited access to health care because of uninsurance.

To facilitate interpretation of our hazard ratio, we first replicated the calculation in the IOM report to estimate the number of US adults who die annually because of lack of health insurance. This approach applies the overall hazard ratio to 9-year age strata and sums these figures to arrive at an annual number of deaths attributable to lack of health insurance. We then recalculated this figure by using the slightly different approach utilized by the Urban Institute, which does not age stratify when calculating total mortality. We believe this approach to be more accurate than that used to produce the IOM estimate, as it calculates mortality from the entire age range that the hazard ratio was calculated from, as opposed to calculating mortality over 10-year age strata.[23]

RESULTS

We display baseline characteristics of the sample in Table 1; 9004 individuals contributed 80 657 person-years of follow-up time between 1988 and 2000. Of these, 16.2% (95% confidence interval [CI]=14.1%, 18.2%) were uninsured at the time of interview.

TABLE 1—Insurance and Mortality Among Nonelderly US Adults Aged 17 to 64 Years: NHANES III (1986–1994) With Follow-Up Through 2000

Characteristic	No. (weighted %)	% Uninsured (SE)	% Died (SE)
Vital status as of December 31, 2000			
Alive	8653 (96.9)	16.2 (1.0)	0
Deceased	351 (3.1)	17.2 (2.8)	100
Insurance status[a]			
Privately insured	6655 (83.8)	0	3.0 (0.3)
Uninsured	2350 (16.2)	100	3.3 (0.6)
Gender			
Female	4695 (50.2)	15.1 (1.1)	2.6 (0.3)
Male	4311 (49.8)	17.3 (1.3)	3.5 (0.4)
Age, y			
17–24	1750 (17.1)	28.5 (2.5)	0.7 (0.2)
25–34	2338 (27.1)	19.7 (1.5)	1.4 (0.4)
35–44	2177 (26.2)	11.6 (1.2)	1.7 (0.3)
45–54	1529 (16.8)	10.8 (1.4)	5.1 (0.9)
55–64	1344 (12.7)	8.9 (1.4)	10.7 (1.1)
Race/ethnicity			
Non-Hispanic White	3484 (78.1)	12.3 (0.8)	3.1 (0.4)
Non-Hispanic Black	2567 (9.9)	22.6 (2.1)	4.1 (0.5)
Mexican American	2598 (5.1)	45.5 (1.9)	3.1 (0.4)
Other	355 (6.9)	29.5 (7.3)	0.9 (0.4)
Education, y			
<12	2917 (19.6)	37.4 (3.0)	4.1 (0.5)
≥12	6087 (80.4)	11.0 (0.7)	2.8 (0.3)
Employment			
Unemployed[b]	511 (4.0)	49.8 (3.9)	5.3 (1.3)
All others	8493 (96.0)	14.8 (0.9)	3.0 (0.3)
Poverty income ratio[c]			
0–1	1678 (9.2)	56.2 (2.7)	4.3 (0.9)
>1–3	4171 (39.7)	22.1 (1.7)	3.0 (0.3)
>3	3155 (51.2)	4.4 (0.5)	3.0 (0.4)
Smoking status			
Current smoker	2465 (29.1)	22.8 (1.8)	4.6 (0.5)
Former smoker[d]	1794 (22.3)	10.4 (1.1)	4.2 (0.7)
Nonsmoker	4745 (48.6)	14.9 (1.1)	1.7 (0.3)
Drinking status, alcoholic drinks/wk			
<6	7193 (78.3)	15.3 (1.1)	4.3 (0.7)
≥6	1811 (21.7)	19.6 (1.5)	2.8 (0.4)
Exercise, METs/mo			
≥100	3475 (42.0)	13.7 (1.1)	2.9 (0.4)
<100	5529 (58.0)	18.0 (1.1)	3.2 (0.4)
Self-rated health			
Excellent	1675 (23.4)	9.3 (1.3)	2.0 (0.4)
Very good	2499 (34.9)	12.0 (0.9)	1.4 (0.4)
Good	3288 (31.7)	20.5 (1.9)	3.3 (0.4)
Fair or poor	1542 (9.9)	33.6 (2.5)	10.8 (1.2)

Continued

TABLE 1—*Continued*

Physician-rated health on examination			
Excellent	4627 (54.2)	16.8 (1.2)	1.8 (0.3)
Very good	2179 (24.4)	13.3 (1.2)	2.6 (0.5)
Good	1858 (18.4)	17.2 (1.4)	4.9 (0.7)
Fair or poor	340 (3.0)	21.7 (4.8)	19.0 (2.6)
Measured BMI			
<18.5	205 (2.7)	19.8 (4.0)	4.0 (1.4)
18.5-25	3764 (46.8)	16.4 (1.2)	2.4 (0.3)
>25-<30	2853 (30.4)	14.9 (1.2)	3.3 (0.7)
≥30	2182 (20.0)	17.2 (1.8)	4.3 (0.8)

Notes. BMI = body mass index (weight in kg divided by height in meters squared); METs = metabolic equivalents; NHANES = National Health and Nutrition Examination Survey.
[a]For those with complete data for all characteristics; excludes those covered by any government insurance.
[b]Looking for work, laid off, or unemployed.
[c]Combines family income, poverty threshold, and year of survey to allow analysis of income data across the 6 years of NHANES III; less than 1 indicates less than the poverty threshold.
[d]Smoked more than 200 cigarettes in lifetime.

Uninsurance was associated with younger age, minority race/ethnicity, unemployment, smoking, exercise (less than 100 METs per month), self-rated health, and lower levels of education and income ($P<.001$ for all comparisons). Regular alcohol use and physician-rated health were also associated with higher rates of uninsurance ($P<.05$ for both comparisons).

By the end of follow-up in 2000, 351 individuals, or 3.1% (95% CI=2.5%, 3.7%) of the sample, had died (Table 1). Significant bivariate predictors of mortality included male gender ($P=.04$), age ($P<.001$), minority race/ethnicity ($P<.001$), less than 12 years of education ($P=.008$), unemployment ($P=.02$), smoking ($P<.001$), regular alcohol use ($P=.04$), worse self-rated health status ($P<.001$), and worse physician-rated health status ($P<.001$).

In the model adjusted only for age and gender, lack of health insurance was significantly associated with mortality (hazard ratio [HR]=1.80; 95% CI=1.44, 2.26). In subsequent models adjusted for gender, age, race/ethnicity, poverty income ratio, education, unemployment, smoking, regular alcohol use, self-rated health, physician-rated health, and BMI, lack of health insurance significantly increased the risk of mortality (HR=1.40; 95% CI=1.06, 1.84; Table 2). We detected no significant interactions between lack of health insurance and any other variables. Our sensitivity analyses yielded substantially similar estimates.

Replicating the methods of the IOM panel with updated census data[24,25] and this hazard ratio, we calculated 27 424 deaths among Americans aged 25 to 64 years in 2000 associated with lack of health insurance. Applying this hazard ratio to census data from 2005[26] and including all persons aged 18 to 64 years yields an estimated 35 327 deaths annually among the nonelderly associated with lack of health insurance. When we repeated this approach without age stratification, (thought by investigators at the Urban Institute to be an overly conservative approach)[23] we calculated approximately 44 789 deaths among Americans aged 18 to 64 years in 2005 associated with lack of health insurance.

DISCUSSION

The uninsured are more likely to die than are the privately insured. We used a nationally representative data set to update the oft-cited study by Franks et al. and demonstrate the persistence of increased mortality attributable to uninsurance. Our findings are in accord with earlier research showing that lack of health insurance increases the likelihood of death in select illnesses and populations.[5–7,13] Our estimate for annual deaths attributable to uninsurance among working-age Americans is more than 140% larger than the IOM's earlier figure.[23]

By using methodologies similar to those used in the 1993 study, we found that being uninsured is associated with a similar hazard for mortality (1.40 for our study vs 1.25 for the 1993 study). Although the NHANES I study methodology and population were similar to those used in NHANES III, differences exist. The population analyzed in the original study was older on average than were participants in our sample (22.8% vs 55.6% aged 34 years or younger). The maximum length of follow-up was less (16 years vs 12 years), and the earlier analysis was limited to White and Black persons, whereas the present study also includes Mexican Americans.

The relative youthfulness and shorter follow-up in our study population would be expected to reduce our power to detect an elevated risk of death. In addition, if gaining Medicare reduces the effect of uninsurance on mortality, then the younger age and shorter length of follow-up in our study might strengthen the association between uninsurance and mortality compared with the earlier study. It is less clear how the differences in the racial and ethnic make-up of our study population would affect our ability to detect difference in risk of death. In fact, the increased likelihood of uninsurance among Mexican Americans who were nonetheless no more likely to die than non-Hispanic Whites might also be expected to reduce our power compared with the earlier study.

The original analysis confirmed vital status by review of decedents' death certificates. The NCHS had developed a probabilistic matching strategy to establish vital status. A subsample underwent death certificate review and verification; 98.7% were found to be correctly classified following this review.[16] Again, it is not clear how any misclassification would bias our results. Moreover, Congress extended Medicare coverage in 1972 to 2 nonelderly groups: the long-term disabled and those with end-stage renal disease.[27] So, although both studies excluded Medicare enrollees, only ours entirely excluded disabled nonelderly adults who are at particularly high risk of death.

TABLE 2—Adjusted Hazards for Mortality Among US Adults Aged 17 to 64 Years: NHANES III, 1988–2000

Characteristic	Hazards Ratio (95% CI)
Insurance status	
Privately insured[a] (Ref)	1.00
Uninsured	1.40 (1.06, 1.84)
Age[b]	1.06 (1.05, 1.07)
Gender	
Female (Ref)	1.00
Male	1.37 (1.13, 1.68)
Race/ethnicity	
Non-Hispanic White (Ref)	1.00
Non-Hispanic Black	1.32 (0.98, 1.79)
Mexican American	0.88 (0.64, 1.19)
Other	0.46 (0.24, 0.90)
Exercise, METs/mo	
≥100 (Ref)	1.00
<100	1.05 (0.80, 1.38)
Smoking status	
Nonsmoker (Ref)	1.00
Current smoker	2.02 (1.43, 2.85)
Former smoker[c]	1.42 (1.09, 1.85)
Drinking status, alcoholic drinks/wk	
<6 (Ref)	1.00
≥6	1.38 (0.99, 1.92)
Education, y	
≥12 (Ref)	1.00
<12	0.98 (0.75, 1.27)
Employment	
Not unemployed[d] (Ref)	1.00
Unemployed	1.40 (0.92, 2.14)
Self-rated health	
Excellent (Ref)	1.00
Very good	0.67 (0.42, 1.09)
Good	1.27 (0.84, 1.90)
Fair or poor	2.26 (1.40, 3.64)
Physician-rated health	
Excellent (Ref)	1.00
Very good	0.99 (0.77, 1.27)
Good	1.17 (0.90, 1.52)
Fair or poor	3.22 (2.26, 4.58)
Measured BMI	
<18.5	1.26 (0.69, 2.29)
18.5–25 (Ref)	1.00
>25–<30	0.87 (0.66, 1.15)
≥30	0.89 (0.69, 1.15)
Poverty income ratio[e]	1.03 (0.95, 1.12)

Notes. BMI = body mass index (weight in kg divided by height in meters squared); CI = confidence interval; METs = metabolic equivalents.
[a]For those with complete data for all characteristics; excludes those covered by any government insurance.
[b]Hazard ratio reflects risk for every 1-year increase in age.
[c]Smoked more than 200 cigarettes in lifetime.
[d]Looking for work, laid off, or unemployed.
[e]Combines family income, poverty threshold, and year of survey to allow analysis of income data across the 6 years of NHANES III; less than 1 indicates less than the poverty threshold. Entered into regression model as a continuous variable. Hazard ratio represents change for every 1 unit increase in the poverty income ratio.

The mechanisms by which health insurance affects mortality have been extensively studied. Indeed, the IOM issued an extensive report summarizing this evidence.[29] The IOM identified 3 mechanisms by which insurance improves health: getting care when needed, having a regular source of care, and continuity of coverage.

The uninsured are more likely to go without needed care than the insured. For instance, Lurie et al. demonstrated that among a medically indigent population in California, loss of government-sponsored insurance was associated with decreased use of physician services and worsening control of hypertension.[28,29] The uninsured are also more likely to visit the emergency department[30] and be admitted to the hospital[31] for "ambulatory care sensitive conditions," suggesting that preventable illnesses are a consequence of uninsurance.

The chronically ill uninsured are also less likely to have a usual source of medical care,[32] decreasing their likelihood of receiving preventative and primary care. Discontinuity of insurance is also harmful; those intermittently uninsured are more likely to die than the insured.[13]

All of these factors likely play a role in the decline in health among middle-aged uninsured persons detected by Baker et al.[33,34] This trend appears to reverse at age 65, when the majority gains access to Medicare coverage.[35] Other studies suggest that extending health insurance not only improves health, but also may be cost effective.[36]

Limitations

Our study has several limitations. NHANES III assessed health insurance at a single point in time and did not validate self-reported insurance status. We were unable to measure the effect of gaining or losing coverage after the interview. Point-in-time uninsurance is associated with subsequent uninsurance.[6] Intermittent insurance coverage is common and accelerates the decline in health among middle-aged persons.[33] Among the near-elderly, point-in-time uninsurance was associated with significant decline in overall health relative to those with private insurance.[13] Earlier population-based surveys that did validate insurance status found that between 7% and 11% of those initially recorded as being uninsured were misclassified.[13] If present, such misclassification might dilute the true effect of uninsurance in our sample. We excluded 29.5% of the sample because of missing data. These individuals were more likely to be uninsured and to die, which might also bias our estimate toward the null.

We have no information about duration of insurance coverage from this survey. Further, we have no data regarding cost sharing (out-of-pocket expenses) among the insured; cost sharing worsened blood pressure control among the poor in the RAND Health Insurance Experiment, and was associated with decreased use of essential medications, and increased rates of emergency department use and adverse events in a random sample of elderly and poor Canadians.[37,38]

Unmeasured characteristics (i.e., that individuals who place less value on health eschew both health insurance and healthy behaviors) might offer an alternative explanation for our findings. However, our analysis controlled for tobacco and alcohol use, along with obesity and exercise habits. In addition, research has found that more than 90% of nonelderly adults without insurance cite cost or lack of employer-sponsored coverage as reasons for being uninsured, whereas only 1% percent report "not needing" insurance.[39] In fact, the variables included in our main survival analysis may inappropriately diminish the relationship between

insurance and death. For example, poor physician-rated health, poor self-rated health, and unemployment may result from medically preventable conditions. Indeed, earlier analyses suggest that the true effect of uninsurance is likely larger than that measured in multivariate models.[13,40] In addition, Hadley found that accounting for endogeneity bias by using an instrumental variable increases the protective effect of health insurance on mortality.[40]

Conclusions

Lack of health insurance is associated with as many as 44 789 deaths per year in the United States, more than those caused by kidney disease (n=42 868).[41] The increased risk of death attributable to uninsurance suggests that alternative measures of access to medical care for the uninsured, such as community health centers, do not provide the protection of private health insurance. Despite widespread acknowledgment that enacting universal coverage would be life saving, doing so remains politically thorny. Now that health reform is again on the political agenda, health professionals have the opportunity to advocate universal coverage. ■

About the Authors

At the time of this research, all authors were with the Department of Medicine at Cambridge Health Alliance, affiliated with Harvard Medical School, Cambridge, MA.

Correspondence should be sent to Andrew P. Wilper, MD, MPH, Veterans Affairs Medical Center, 500 W Fort St, Boise, ID 83702 (e-mail: wilp9522@u.washington.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints/Eprints" link.

This article was accepted on March 16, 2009.

Contributors

A. P. Wilper designed the study, planned the analysis, performed statistical analysis and data management, and interpreted the analysis. A. P. Wilper, S. Woolhandler, and D. U. Himmelstein drafted the article. K. E. Lasser, D. McCormick, and D. H. Bor performed critical revisions of the article. S. Woolhandler supervised all aspects of the study design, analysis planning, interpretation, and article preparation.

Acknowledgments

A. P. Wilper was supported by a Health Resources and Service Administration National Research Service Award (5T32 HP110011).

Human Participant Protection

The institutional review board of Cambridge Health Alliance deemed this study exempt from formal review.

References

1. US Census Bureau. Income, poverty, and health insurance coverage in the United States: 2007. Available at: http://www.census.gov/prod/2008pubs/p60-235.pdf. Accessed August 29, 2008.

2. Ayanian JZ, Weissman JS, Schneider EC, Ginsburg JA, Zaslavsky AM. Unmet health needs of uninsured adults in the United States. *JAMA.* 2000;284:2061–2069.

3. Ayanian JZ, Zaslavsky AM, Weissman JS, Schneider EC, Ginsburg JA. 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.

4. Institute of Medicine. *Care Without Coverage, Too Little, Too Late.* Washington, DC: The National Academies Press; 2002.

5. Franks P, Clancy CM, Gold MR. Health insurance and mortality. Evidence from a national cohort. *JAMA.* 1993;270:737–741.

6. McWilliams JM, Zaslavsky AM, Meara E, Ayanian JZ. Health insurance coverage and mortality among the near-elderly. *Health Aff (Millwood).* 2004;23:223–233.

7. Sorlie PD, Johnson NJ, Backlund E, Bradham DD. Mortality in the uninsured compared with that in persons with public and private health insurance. *Arch Intern Med.* 1994;154:2409–2416.

8. Bennett CL, Horner RD, Weinstein RA, et al. Racial differences in care among hospitalized patients with Pneumocystis carinii pneumonia in Chicago, New York, Los Angeles, Miami, and Raleigh-Durham. *Arch Intern Med.* 1995;155:1586–1592.

9. Ayanian JZ, Kohler BA, Abe T, Epstein AM. The relation between health insurance coverage and clinical outcomes among women with breast cancer. *N Engl J Med.* 1993;329:326–331.

10. Roetzheim RG, Gonzalez EC, Ferrante JM, Pal N, Van Durme DJ, Krischer JP. Effects of health insurance and race on breast carcinoma treatments and outcomes. *Cancer.* 2000;89:2202–2213.

11. Roetzheim RG, Pal N, Gonzalez EC, Ferrante JM, Van Durme DJ, Krischer JP. Effects of health insurance and race on colorectal cancer treatments and outcomes. *Am J Public Health.* 2000;90:1746–1754.

12. Canto JG, Rogers WJ, French WJ, Gore JM, Chandra NC, Barron HV. Payer status and the utilization of hospital resources in acute myocardial infarction: a report from the National Registry of Myocardial Infarction 2. *Arch Intern Med.* 2000;160:817–823.

13. Baker DW, Sudano JJ, Durazo-Arvizu R, Feinglass J, Witt WP, Thompson J. Health insurance coverage and the risk of decline in overall health and death among the near elderly, 1992-2002. *Med Care.* 2006;44:277–282.

14. Bhattacharya J, Goldman D, Sood N. The link between public and private insurance and HIV-related mortality. *J Health Econ.* 2003;22:1105–1122.

15. Centers for Disease Control and Prevention. National Health and Nutrition Examination III data files, documentation, and SAS code. Available at: http://www.cdc.gov/nchs/about/major/nhanes/nh3data.htm. Accessed May 9, 2008.

16. Centers for Disease Control and Prevention. National Health and Nutrition Examination III Linked Mortality File: Matching Methodology. Available at: http://www.cdc.gov/nchs/data/datalinkage/matching_methodology_nhanes3_final.pdf. Accessed April 8, 2008.

17. Wheatcroft G, Cox CS, Lochner KA. Comparative analysis of the NHANES III public-use and restricted-use linked mortality files. Available at: http://www.cdc.gov/nchs/data/datalinkage/nh3_mort_compare_2007_final.pdf. Accessed April 14, 2008.

18. Centers for Disease Control and Prevention. National Health and Nutrition Examination III adult data file documentation. Available at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHANES/NHANESIII/1A/ADULT-acc.pdf. Accessed April 14, 2008.

19. Centers for Disease Control and Prevention. Physical activity for everyone Web site. Available at: http://www.cdc.gov/nccdphp/dnpa/physical/everyone/recommendations/index.htm. Accessed April 16, 2008.

20. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation.* 2007;116:1081–1093.

21. Centers for Disease Control and Prevention. National Health and Nutrition Examination III, Cycle 2 Physician Examiner's Training Manual. Available at: http://www.cdc.gov/nchs/data/nhanes/nhanes3/cdrom/nchs/manuals/phys.pdf. Accessed April 8, 2008.

22. Korn EL, Graubard BI. *Analysis of Health Surveys.* Wiley Series in Probability and Statistics. New York, NY: Wiley; 1999.

23. Dorn S. Uninsured and Dying Because of It: Updating the Institute of Medicine Analysis on the Impact of Uninsurance on Mortality. Available at: http://www.urban.org/UploadedPDF/411588_uninsured_dying.pdf. Accessed April 23, 2008.

24. US Census Bureau. Health insurance coverage status and type of coverage all persons by age and sex: 1999 to 2007. Available at: http://www.census.gov/hhes/www/hlthins/historic/hihistt2.xls. Accessed August 25, 2009.

25. Centers for Disease Control and Prevention, National Center for Health Statistics. Death rates by 10-year age groups: United States and each state, 2000. Available at: http://www.cdc.gov/nchs/data/statab/VS00100_TABLE23A.pdf. Accessed August 25, 2009.

26. Centers for Disease Control and Prevention, National Center for Health Statistics. Death rates by 10-year age groups: United States and each state, 2005. Available at: http://www.cdc.gov/nchs/data/statab/MortFinal2005_Worktable23R.pdf. Accessed August 25, 2009.

27. Centers for Medicare and Medicaid Services. Key milestones in CMS programs. Available at: http://www.cms.hhs.gov/History/Downloads/CMSProgramKeyMilestones.pdf. Accessed May 20, 2008.

28. Lurie N, Ward NB, Shapiro MF, Brook RH. Termination from Medi-Cal–does it affect health? *N Engl J Med.* 1984;311:480–484.

29. Lurie N, Ward NB, Shapiro MF, Gallego C, Vaghaiwalla R, Brook RH. Termination of Medi-Cal benefits. A follow-up study one year later. *N Engl J Med.* 1986;314:1266–1268.

30. Oster A, Bindman AB. Emergency department visits for ambulatory care sensitive conditions: insights into preventable hospitalizations. *Med Care.* 2003;41: 198–207.

31. Weissman JS, Gatsonis C, Epstein AM. Rates of avoidable hospitalization by insurance status in Massachusetts and Maryland. *JAMA.* 1992;268: 2388–2394.

32. Wilper AP, Woolhandler S, Lasser KL, McCormick D, Bor DH, Himmelstein DU. Chronically ill and uninsured: a national study of disease prevalence and access to care among U.S. adults. *Ann Intern Med.* 2008;149: 170–176.

33. Baker DW, Sudano JJ, Albert JM, Borawski EA, Dor A. Lack of health insurance and decline in overall health in late middle age. *N Engl J Med.* 2001;345: 1106–1112.

34. Baker DW, Sudano JJ, Albert JM, Borawski EA, Dor A. Loss of health insurance and the risk for a decline in self-reported health and physical functioning. *Med Care.* 2002;40:1126–1131.

35. McWilliams JM, Meara E, Zaslavsky AM, Ayanian JZ. Health of previously uninsured adults after acquiring Medicare coverage. *JAMA.* 2007;298: 2886–2894.

36. McWilliams JM, Meara E, Zaslavsky AM, Ayanian JZ. Use of health services by previously uninsured Medicare beneficiaries. *N Engl J Med.* 2007;357: 143–153.

37. Newhouse J. *Free for All? Lessons From the RAND Health Insurance Experiment.* Cambridge, MA: Harvard University Press; 1993.

38. Tamblyn R, Laprise R, Hanley JA, et al. Adverse events associated with prescription drug cost-sharing among poor and elderly persons. *JAMA.* 2001;285: 421–429.

39. Graves JA, Long SK. Why do people lack health insurance? Health Policy Online: Timely Analysis of Current Trends and Policy Options. Available at: http://www.urban.org/UploadedPDF/411317_lack_health_ins.pdf. Accessed May 29, 2008.

40. Hadley J, Waidmann T. Health insurance and health at age 65: implications for medical care spending on new Medicare beneficiaries. *Health Serv Res.* 2006;41: 429–451.

41. Kung H, Hyoert D, Xu J, et al. Deaths: final data for 2005. *Natl Vital Stat Rep.* 2008;56(10). Available at: http://www.cdc.gov/nchs/data/nvsr/nvsr56/nvsr56_10.pdf. Accessed February 17, 2009.

Outcomes and predictors of very stable INR control during chronic anticoagulation therapy

Daniel M. Witt,[1-3] Thomas Delate,[2,3] Nathan P. Clark,[1,2] Chad Martell,[2] Thu Tran,[2] Mark A. Crowther,[4] David A. Garcia,[5] Walter Ageno,[6] and Elaine M. Hylek,[7] on behalf of the Warfarin Associated Research Projects and other EnDeavors (WARPED) Consortium

[1]Kaiser Permanente Colorado Clinical Pharmacy Anticoagulation Service, Lafayette; [2]University of Colorado School of Pharmacy, Denver; [3]Kaiser Permanente Colorado Clinical Pharmacy Research Team, Aurora; [4]McMaster University, Hamilton, ON; [5]University of New Mexico School of Medicine, Albuquerque; [6]University of Insubria, Varese, Italy; and [7]Boston University School of Medicine, MA

For patients on warfarin therapy, an international normalized ratio (INR) recall interval not exceeding 4 weeks has traditionally been recommended. Less frequent INR monitoring may be feasible in stable patients. We sought to identify patients with stable INRs (defined as having INR values exclusively within the INR range) and comparator patients (defined as at least one INR outside the INR range) in a retrospective, longitudinal cohort study. Occurrences of thromboembolism, bleeding, and death were compared between groups. Multivariate logistic regression models were used to identify independent predictors of stable INR control. There were 2504 stable and 3569 comparator patients. The combined rates of bleeding and thromboembolism were significantly lower in stable patients. Independent predictors of stable INR control were age older than 70 years and the absence of comorbid heart failure and diabetes. Stable patients were significantly less likely to have target INR of 3.0 or higher or chronic diseases. We hypothesize that many patients demonstrating stable INR control could be safely treated with INR recall intervals greater than the traditional 4 weeks. (Blood. 2009; 114:952-956)

Introduction

Warfarin is effective for the primary and secondary prevention of both arterial and venous thromboembolic disorders. Its variable dose response and narrow therapeutic index mandate periodic monitoring of the international normalized ratio (INR).[1] Target INR ranges of 2.0 to 3.0 or 2.5 to 3.5 have been recommended for most indications because INR values in these ranges are associated with the best combination of thrombosis reduction and bleeding avoidance.[1] Although multiple studies have addressed the optimum target intensity of anticoagulation, few studies have addressed the optimal testing frequency. Current guidelines suggest a time interval not exceeding 4 weeks between INR determinations.[1,2] However, this recommendation is not evidence based, having evolved instead from regional differences in routine clinical practice and expert opinion.[3]

More frequent INR testing has been suggested as a means to increase time in the therapeutic range, especially among patients who self-monitor warfarin using point-of-care technology.[1,4,5] Although more frequent testing may increase the proportion of time within the therapeutic INR range in some patients, it is not likely to benefit those patients who demonstrate long-term INR stability as demonstrated by minimal INR deviation and longitudinal warfarin dose stability. Hypothetically, less frequent INR monitoring may be possible for such patients. Supporting evidence comes from the United Kingdom where anticoagulation providers routinely allow INR recall intervals in stable patients up to 90 days.[6] Recent evidence suggests that longer INR recall intervals may also be associated with improved INR control,[7,8] which has in turn been associated with reduced risk for anticoagulation therapy-related adverse events.[9,10]

Our objectives were to identify a subgroup of patients with very stable (ie, all INR values in the therapeutic range) INR control, to compare the risk of anticoagulation therapy-related adverse events in such patients to the corresponding risk in patients without exclusively therapeutic INR control, and to describe patient characteristics associated with long-term INR stability.

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 more than 480 000 members in the Denver-Boulder metropolitan area. Anticoagulation services at KPCO are provided by a centralized Clinical Pharmacy Anticoagulation Service (CPAS).[9] Working collaboratively with the referring physician and using standardized dosing algorithms,[11] CPAS clinical pharmacists initiate, adjust, and refill anticoagulant medications and order relevant laboratory tests. Dosing algorithms used during the study specified a maximum INR recall interval of 6 weeks. Integrated, electronic medical, pharmacy, and laboratory records system and CPAS database (Dawn-AC; 4S Systems Ltd) were used to identify patients, treatments, and outcomes for this study. Approval to conduct this study was obtained from the KPCO Institutional Review Board.

Submitted February 26, 2009; accepted May 5, 2009. Prepublished online as *Blood* First Edition paper, May 13, 2009; DOI 10.1182/blood-2009-02-207928.

An Inside *Blood* analysis of this article appears at the front of this issue.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

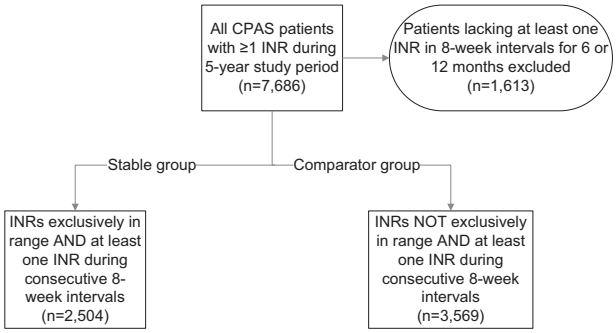


Figure 1. Process for defining the study groups.

Patients

Patients with a duration of warfarin therapy in excess of 90 days, at least one INR determination during the study time frame (January 2000 through December 2005), an age of greater than 18 years, and warfarin therapy continuing throughout a 6-month observation period were included in the study.

Stable patients were defined as having all INR values within the strictly defined therapeutic reference interval for the first identifiable continuous 6-month period (ie, 100% INR control). Comparator patients were those who did not have any continuous 6-month period where all INR values were within the therapeutic range. 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 respective observation periods. The process for defining the study cohorts is depicted in Figure 1.

Data collection

Variables collected for analysis included the primary warfarin indication (atrial fibrillation, venous thromboembolism, heart valve disorder, other), age at start of the observation period, sex, INR target, duration of warfarin therapy, and INR values. Patient-specific factors that could influence the risk for anticoagulant-related complications were also recorded: diabetes mellitus, hypertension, heart failure, prior venous thrombosis, hemorrhage or stroke, cancer, and estrogen therapy. Risk factors were considered present when a coded assessment for a given factor 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 using ambulatory prescription drug data from the observation period.[12] 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 use.[12,13]

The first occurrence of anticoagulant-related complications (thromboembolism, bleeding, and death) was determined as previously described.[14] Briefly, specific complications requiring admission to the emergency department or hospital 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 2 investigators. Events were scored using a modified Naranjo scale to quantify the relationship of the adverse event with warfarin therapy.[15] 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.

Table 1. Baseline characteristics

Characteristic	Stable group, n = 2504	Comparator group, n = 3569	*P*
Mean age,* y (SD)	72.3 (10.9)	68.8 (13.1)	< .001
Age older than 70 y, %	63.0	51.5	< .001
Male, %	52.0	51.5	.688
INR target, %			
2.0	3.9	3.3	.167
2.5	87.0	79.7	< .001
3.0 or more	9.1	17.0	< .001
Primary indication for anticoagulation therapy, %			
Atrial fibrillation	49.9	43.4	< .001
Venous thromboembolism	25.6	25.8	.856
Heart valve disorder	8.0	12.7	< .001
Other	16.5	18.1	.107
Risk factors, %			
Diabetes mellitus†	1.6	3.5	< .001
Hypertension†	18.2	20.2	.046
Heart failure†	5.9	8.7	< .001
Prior venous thrombosis†	2.5	3.7	.012
Prior hemorrhage†	1.2	2.0	.021
Prior stroke†	0.0	0.1	.273
Cancer†	0.2	0.6	.060
Estrogen therapy‡	7.8	10.7	< .001
Mean chronic disease score (SD)	6.5 (2.6)	6.7 (2.7)	< .001
Median duration of warfarin therapy, d§ (IQR)	1166 (554, 2051)	755 (725, 1753)	.743

INR indicates international normalized ratio; IQR, interquartile range; and SD, standard deviation.

*As of date of index INR measurement.

†During the 180 days before the index INR.

‡During the 90 days before the index INR.

§From initiation of warfarin therapy.

Statistical analysis

Data analyses were performed using SAS 9.1.3 statistical software. Patient characteristics were reported as means and standard deviations for interval-level variables (eg, age, warfarin dose, length of warfarin therapy) and percentages for categoric variables (eg, sex, target INR, occurrence of anticoagulation therapy-related complications). Associations between categoric variables were assessed using the chi-square test and continuous variables were compared using the independent samples *t* test or Wilcoxon rank sum test (depending on the distribution of the data). Patient characteristics and risk factors were entered into multivariate logistic regression models to identify variables that independently predict INR stability. The alpha was set at .05.

Results

Records from 7686 patients were screened; of these, 6073 patients had a period where an INR was measured every 8 weeks for at least 6 months. The stable group was composed of 2504 patients with INR values within the desired reference interval on all determinations and the comparator group of 3569 patients with at least one INR outside the desired reference interval (Figure 1).

Baseline characteristics of stable patients and comparators are presented in Table 1. Stable group patients were older than comparator group patients and more likely to have had a target INR of 2.5 and to have been receiving warfarin for atrial fibrillation, but less likely to have had a target INR of 3.0 or higher, to have been receiving warfarin for heart valve replacement, to have comorbid diabetes, heart failure, or prior venous thrombosis, or to be

Table 2. Unadjusted outcomes during 180-day follow-up period

Characteristic	Stable group, n = 2504	Comparator group, n = 3569	*P*
Received heparin,* %	0.3	3.2	< .001
Deceased, n, %	10, 0.4	58, 1.6	< .001
AC-related death, n, %	1, 0.04	5, 0.1	.411†
AC-related thrombosis, n, %	10, 0.4	26, 0.7	.100
AC-related bleeding, n, %	19, 0.8	101, 2.8	< .001
AC-related bleeding or thrombosis, n, %	28, 1.1	127, 3.6	< .001

AC indicates anticoagulation.
*Heparin or low-molecular-weight heparin.
†Fisher exact test.

receiving concurrent estrogen therapy. The mean chronic disease score was also lower in stable group patients. Differences in duration of warfarin therapy between groups prior to inclusion in the study were not statistically significant. The mean proportion of INR values in the therapeutic range for the comparator group was 46.9% (standard deviation [SD] = 22.0). The stable group had a lower mean number of INRs measured per patient, 6.7 (SD = 1.3) during the observation period compared with 10.7 (SD = 4.5) per patient for comparators ($P < .001$).

Rates of anticoagulation therapy–related adverse events (thromboembolism, bleeding, and death) are summarized in Table 2. Compared with stable group patients, the rate of overall mortality was higher in the comparator group ($P < .01$); however, the difference in anticoagulation therapy–related mortality rate was not statistically significant. The rate of anticoagulation-related bleeding complications was higher in the comparator group compared with their stable counterparts ($P < .05$). Compared with stable group patients, the combined complication rates of bleeding or thromboembolism occurred at a higher rate in the comparator group ($P < .001$). Patients in the comparator group were more likely than the stable group to require coadministration of heparin or low-molecular-weight heparin ($P < .001$).

Table 3 summarizes patient characteristics predictive of stable status. Significant predictors of stable group status were age older than 70 years (odds ratio [OR] = 1.54; 95% confidence interval [CI], 1.38-1.72) and the absence of comorbid diabetes (OR = 1.87; 95% CI, 1.3-2.67), heart failure (OR = 1.43; 95% CI, 1.16-1.76), or concurrent estrogen therapy (OR = 1.32; 95% CI, 1.09-1.60). Stable patients were significantly less likely to have a target INR of 3.0 or higher (OR = 0.48; 95% CI, 0.38-0.61) and increasing chronic disease scores (OR = 0.96; 95% CI, 0.94-0.98).

Table 3. Predictors of stable INR control status (c-statistic = 0.61)

Predictor	Odds ratio	95% CI
Age		
Older than 70 y	1.54	1.38-1.72
70 y or younger		
Sex		
Female		
Male	0.98	0.88-1.10
INR target		
2.0	1.12	0.85-1.48
2.5		
3.0 or more	0.48	0.38-0.61
Primary indication for anticoagulation therapy		
Atrial fibrillation		
Venous thromboembolism	0.93	0.81-1.06
Heart valve disorder	1.18	0.89-1.56
Other	0.90	0.78-1.05
Thromboembolic risk factors		
Diabetes mellitus		
Yes		
No	1.87	1.30-2.67
Hypertension		
Yes		
No	1.09	0.95-1.25
Heart failure		
Yes		
No	1.43	1.16-1.76
Prior venous thrombosis		
Yes		
No	1.33	0.97-1.81
Prior hemorrhage		
Yes		
No	1.53	0.99-2.38
Estrogen therapy		
Yes		
No	1.32	1.09-1.60
Chronic disease score	0.96	0.94-0.98

Discussion

In this large retrospective cohort study, we identified 2504 patients with very stable long-term INR control. We identified that age older than 70 years and the absence of comorbid diabetes or heart failure independently predicted this INR stability. Patients with a target INR of 3.0 or higher and those with a greater burden of chronic diseases were less likely to have such long-term INR stability. On average, the proportion of comparator patients' INRs in the therapeutic range was 46.9%, whereas stable patients' INRs were 100% therapeutic. The seemingly suboptimal INR control reflects the absence of these very stable patients from the comparator group. The time in therapeutic range for all patients managed by CPAS is typically about 64%.[9] Although other investigations have examined predictors of very poor INR control,[16-19] to our knowledge this study is the first to assemble a large cohort of anticoagulated patients and carefully evaluate them for predictors of INR stability.

Our findings are important as patients with long-term stable INR control may be adequately treated with less frequent INR monitoring, perhaps as infrequently as every 8 weeks. Extending the INR recall interval in such patients is likely to reduce costs and increase convenience (and therefore perhaps adherence) without impacting the risk for bleeding or thrombosis.

The most surprising observation in our analysis was that age older than 70 years predicted long-term INR stability. This observation is somewhat counterintuitive and should be confirmed in additional studies. This finding argues against innate INR variability associated with advancing age. The possibility that younger patients were more likely to have been receiving warfarin for heart valve indications was explored posthoc by comparing the proportion of patients 70 years or older in both groups on warfarin for this indication. In the stable group, 5.0% had a heart valve indication compared with 7.5% in the comparator group ($P = .09$). An interaction term for age and warfarin indication was tested in the predictive model but was not significant.

Our results are likely to be valid. The data set used to complete this study is robust and has been used previously in health records and data extraction research.[9,14] 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. Real-world patients with a variety of indications for warfarin and therapeutic INR targets were included. Clinical events were comprehensively collected and described, and INR determinations were performed by a single laboratory and systematically captured in an integrated electronic medical record. All clinical events were independently assessed for causality by 2 expert reviewers. The long-term stable cohort was carefully established using a definition for stability (ie, 100% of INR values within the strict INR range) more rigorous than that used by most anticoagulation providers in routine practice. For example, had INR results within 0.2 of the upper and lower limits of the specified INR range qualified as "in-range" (as is common clinical practice in North America), the number of patients with long-term INR stability would have been substantially larger.[20] Most patients observed in our study had been on warfarin therapy for several years. Although differences were not statistically significant between groups, individuals with long-term stability tended to have been on warfarin longer than comparator patients. 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. It is retrospective and relies upon extraction of data from administrative databases and medical records. Not all variables likely to enter into clinical decision making were collected. The observational study design also precludes definitive establishment of cause and effect relationships between study variables and outcomes. Retrospective database analysis is particularly prone to missing clinical events if care is delivered outside participating institutions. 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 nonaffiliated hospitals, it is likely that the vast majority of clinically important events were captured. As patients are provided with comprehensive care by our anticoagulation service, we are confident that all pertinent laboratory values were captured. Our study was conducted within an integrated health care delivery system with a specialized anticoagulation service using standardized warfarin dosing protocols and, thus, the observed results may not directly translate to other health care settings.

We would like to have estimated the actual proportion of anticoagulated patients within KPCO with exclusively therapeutic INR control. As not all patients managed by CPAS met initial eligibility criteria, this was not possible. However, of 7686 anticoagulated patients with at least one measured INR during the 5-year study period, we were able to identify 2504 patients (33%) who had at least 6 months of INR values within their desired therapeutic range. Other researchers have reported that approximately 37% of patients with atrial fibrillation managed in community settings are within the therapeutic INR range 75% or more of the time.[7] Irrespective of the actual proportion, our data suggest that a substantial number of patients would be adequately treated with INR recall intervals in excess of 4 weeks. More frequent INR monitoring would of course be necessary in the presence of new comorbidities or new medications affecting the INR.

In conclusion, our work supports the hypothesis that a subgroup of anticoagulated patients with therapeutically stable INR values over 6 months can be identified. In general, these patients will tend to be older, with a target INR less than 3.0, and without significant chronic disease burden. Patients with such stable INR control experience significantly fewer anticoagulation therapy-related complications. We agree with others who have suggested that INR recall intervals should be individually tailored based on recent INR control rather than being fixed at minimum frequency such as 4 weeks.[7] We acknowledge that our findings need to be validated in future prospective evaluations. Specifically, we suggest a prospective randomized study that will enroll chronically anticoagulated patients and, after a period of INR stability, gradually increase the interval between INR determinations up to 8 weeks and possibly as long as 12 weeks in patients with stable INR values and baseline characteristics predictive of long-term INR stability.

Authorship

Contribution: D.M.W., T.D., and N.P.C. designed the research and extracted information from medical records; T.D. performed the statistical analysis; D.M.W., T.D., N.P.C., M.A.C., D.A.G., W.A., and E.M.H. interpreted the analysis and revised the paper; M.A.C. wrote the initial draft of the paper; and T.T. and C.M. extracted information from medical records and reviewed the paper.

Conflict-of-interest disclosure: E.M.H. reports serving as an advisor to Boehringer Ingelheim, Bristol-Myers Squibb, Sanofi-Aventis, and The Medicines Company, and participating in clinical symposia sponsored by Bayer and Bristol-Myers Squibb. D.A.G. reports serving as a consultant to Roche Diagnostics. The remaining authors declare no competing financial interests.

A complete list of WARPED Consortium members appears in the supplemental Appendix, available on the *Blood* website; see the Supplemental Materials link at the top of the online article.

Correspondence: Daniel M. Witt, PharmD, FCCP, BCPS, CACP; 16601 E Centretech Pkwy, Aurora, CO 80011; e-mail: dan.m.witt@kp.org.

References

1. 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;133:160-198.
2. Fuster V, Ryden LE, Cannom DS, et al. ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients With Atrial Fibrillation): developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. *Circulation.* 2006;114:e257-e354.
3. Fitzmaurice DA. Oral anticoagulation therapy should be managed in the community with treatment aimed at standard targets and increased recall intervals. *J Thromb Haemost.* 2008;6:1645-1646.
4. Heneghan C, Alonso-Coello P, Garcia-Alamino M, et al. Self-monitoring of oral anticoagulation: a systematic review and meta-analysis. *Lancet.* 2006;367:404-411.
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. Guidelines on oral anticoagulation: third edition. *Br J Haematol.* 1998;101:374-387.
7. Rose AJ, Ozonoff A, Henault LE, Hylek EM. Warfarin for atrial fibrillation in community-based practise. *J Thromb Haemost.* 2008;6:1647-1654.
8. Snyder CM, Helms BE, Hall DL. Evaluation of INR monitoring frequency and time in therapeutic range. *J Pharm Technol.* 2008;24:255-260.
9. 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;127:1515-1522.
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;9:283-292.

11. Poller L, Shiach CR, MacCallum PK, et al. Multicentre randomised study of computerised anticoagulant dosage: European Concerted Action on Anticoagulation. *Lancet.* 1998;352:1505-1509.
12. Clark DO, Von KM, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. *Med Care.* 1995;33:783-795.
13. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. *J Clin Epidemiol.* 1992;45:197-203.
14. 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.
15. Naranjo CA, Busto U, Sellers EM, et al. A method for estimating the probability of adverse drug reactions. *Clin Pharm Ther.* 1981;30:239-245.
16. Cavallari LH, Aston JL, Momary KM, et al. Predictors of unstable anticoagulation in African Americans. *J Thromb Thrombolysis.* 2009;27:430-437.
17. Kimmel SE, Chen Z, Price M, et al. The influence of patient adherence on anticoagulation control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Arch Intern Med.* 2007;167:229-235.
18. Limdi NA, Arnett DK, Goldstein JA, et al. Influence of CYP2C9 and VKORC1 on warfarin dose, anticoagulation attainment and maintenance among European-Americans and African-Americans. *Pharmacogenomics.* 2008;9:511-526.
19. Sconce E, Khan T, Mason J, et al. Patients with unstable control have a poorer dietary intake of vitamin K compared to patients with stable control of anticoagulation. *Thromb Haemost.* 2005;93:872-875.
20. Rose AJ, Ozonoff A, Berlowitz DR, Henault LE, Hylek EM. Warfarin dose management affects INR control. *J Thromb Haemost.* 2009;7:94-101.

Journal of Thrombosis and Haemostasis, 7: 1982–1989 DOI: 10.1111/j.1538-7836.2009.03598.x

IN FOCUS

Incidence and predictors of bleeding or thrombosis after polypectomy in patients receiving and not receiving anticoagulation therapy

D. M. WITT,*† T. DELATE,†‡ K. H. MCCOOL,*‡ M. B. DOWD,*‡ N. P. CLARK,*‡ M. A. CROWTHER,§ D. A. GARCIA,¶ W. AGENO,** F. DENTALI,** E. M. HYLEK†† and W. G. RECTOR‡‡ ON BEHALF OF THE WARPED[1] CONSORTIUM

**Kaiser Permanente Colorado Clinical Pharmacy Anticoagulation Service, Lafayette, CO; †Kaiser Permanente Colorado Clinical Pharmacy Research Team, Aurora, CO; ‡University of Colorado School of Pharmacy, Denver, CO, USA; §McMaster University, Hamilton, ON, Canada; ¶University of New Mexico School of Medicine, Albuquerque, NM, USA; **University of Insubria, Varese, Italy; ††Boston University School of Medicine, Boston, MA, USA; and ‡‡Colorado Permanente Medical Group, Denver, CO, USA*

To cite this article: Witt DM, Delate T, McCool KH, Dowd MB, Clark NP, Crowther MA, Garcia DA, Ageno W, Dentali F, Hylek EM, Rector WG, on behalf of the WARPED Consortium. Incidence and predictors of bleeding or thrombosis after polypectomy in patients receiving and not receiving anticoagulation therapy. *J Thromb Haemost* 2009; 7: 1982–9.

See also Douketis JD. Perioperative management of antithrombotic therapy: lifting the fog. This issue, pp 1979–81.

Summary. *Background and aims:* To assess the effect of warfarin anticoagulation therapy (AC) on the incidence of colon bleeding after elective colonoscopy with polypectomy and to identify independent predictors of post-polypectomy colon bleeding. *Methods:* This was a retrospective cohort analysis. Patients interrupting warfarin AC therapy for polypectomy (AC group) were matched on age (± 3 years) with up to two patients who underwent polypectomy but were not receiving AC (non-AC group). Data were extracted from electronic medical, pharmacy and laboratory claims and records and manual medical chart review. Incidence rates of colon bleeding requiring hospitalization, other gastrointestinal bleeding, thrombosis and death in the 30 days post-polypectomy were compared between groups. Multivariate regression techniques were used to identify independent predictors of post-polypectomy colon bleeding. *Results:* A total of 425 AC group patients were matched to 800 non-AC group patients. Post-polypectomy colon bleeding occurred more often in AC group patients (2.6% vs. 0.2%, $P = 0.005$). There were no differences in the rates of other outcomes ($P > 0.05$). Independent predictors of colon bleeding included AC group status [adjusted odds ratio (AOR) = 11.6; 95% confidence interval (CI) = 2.3–57.3], number of polyps removed (AOR = 1.2; 95% CI = 1.1–1.4) and male gender (AOR = 9.2, 95% CI = 1.1–74.9). *Conclusions:* The incidence of post-polypectomy colon bleeding was higher in patients receiving AC even although warfarin was interrupted for the procedure. Independent predictors of colon bleeding were identified as: receiving AC, removal of multiple polyps and male gender. Our findings suggest that additional methods to reduce the likelihood of post-polypectomy colon bleeding in AC patients should be investigated.

Keywords: anticoagulation, bleeding, colonoscopy, polypectomy, warfarin.

Introduction

Colonoscopy is the most commonly performed gastrointestinal (GI) endoscopic procedure in the United States [1]. Complications of colonoscopy are uncommon but can be life threatening [2]. For example, polypectomy increases the likelihood of complications nearly 9-fold compared with colonoscopy without polypectomy [3], of which colon bleeding is the most common serious complication.

Patients receiving anticoagulant therapy (AC) with warfarin routinely undergo colonoscopy with polypectomy. Current guidelines suggest warfarin therapy should be interrupted for several days prior to polypectomy in an attempt to minimize the risk of post-polypectomy bleeding [4]. For patients at high risk of thrombosis in the absence of warfarin, cross-coverage with low-molecular-weight heparin (so-called bridging therapy) is frequently prescribed [4]. Anticoagulation therapy has been reported to increase the likelihood of post-polypectomy colon bleeding by up to

Correspondence: Daniel M. Witt, Kaiser Permanente Colorado, 16601 E. Centretech Pkwy, 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 19 June 2009, accepted 19 August 2009

thirteenfold [5,6]. These estimates, however, may be unreliable as they are derived from investigations with only a few AC patients (e.g. 23 out of 5152 patients) [6]; limiting their generalizability. In contrast, another study reported an equivocal relationship between AC and bleeding risk post-polypectomy [7]. Additionally, the potential influence of important factors such as indication for AC, number of warfarin doses held in preparation for, and anticoagulation intensity immediately prior to the procedure on the rate of bleeding in AC patients undergoing polypectomy has not been reported.

The purpose of this investigation was to assess the effect of warfarin AC on the incidence of serious colon bleeding after elective colonoscopy with polypectomy and to identify independent predictors of post-polypectomy bleeding in two large cohorts of patients who underwent polypectomy: one requiring interruption of warfarin AC for the procedure and the other not prescribed AC.

Methods

Setting

This analysis was conducted at Kaiser Permanente Colorado (KPCO), an integrated healthcare delivery system providing care to approximately 450 000 members in the Denver–Boulder metropolitan area, more than 7000 of which are patients treated with warfarin. All anticoagulated patients receive comprehensive services, including plans for peri-procedural warfarin interruption, from a centralized Clinical Pharmacy Anticoagulation Service (CPAS) [8]. Details of these plans are recorded in the patient's electronic medical record (EMR) as well as an anticoagulant database (Dawn-AC; 4S Systems, Cumbria, UK).

Study design

All study activities were approved by the KPCO Institutional Review Board. This was a matched retrospective analysis. All KPCO members 18 years or older who had an elective, outpatient colonoscopy with polypectomy performed between January 2000 and December 2006 were eligible for study inclusion. Patients interrupting chronic warfarin therapy for polypectomy (AC group) were randomly matched using the GREEDY Matching Algorithm (Mayo Clinic College of Medicine 2003) on age (± 3 years) with up to two patients who underwent polypectomy but were not receiving warfarin at that time (Non-AC group). If a patient had more than one polypectomy during the study period, only data from the first episode were analyzed.

Study population

All included patients were continuously enrolled in the Kaiser Foundation Health Plan during the 180 days prior to and 30 days after polypectomy (unless death occurred during the 30 days following). Patients who received colonoscopy to evaluate emergent GI bleeding, had a history of bleeding from intestinal arteriovenous malformations (ICD-9 code 569.85), had insufficient EMR information to assess study-related variables, had their warfarin monitored outside of CPAS, or were receiving prophylaxis with warfarin to prevent venous thromboembolism after orthopedic surgery (because the duration of warfarin therapy in this setting is typically ≤ 6 weeks) were excluded.

Study outcomes

The primary outcome was the incidence rate of colon bleeding requiring hospitalization [including emergency department (ED) visits] within 30 days of polypectomy. The 30-day follow-up period was chosen as the majority of post-polypectomy bleeding complications have been reported to occur within this time frame [5,9]. Descriptions of each case were summarized. Secondary outcomes included the incidence rates of hospitalization for non-colon GI bleeding, hospitalization for thromboembolism and death attributable to bleeding or thromboembolism during the 30-day post-polypectomy follow-up. In addition, independent predictors of post-polypectomy colon bleeding were identifed.

Data collection

Patients who had undergone elective colonoscopy with polypectomy were identified by querying integrated KPCO electronic administrative databases using International Classification of Diseases, Ninth Revision (ICD-9) codes 45.42 and 48.36 and Current Procedural Terminology (CPT) codes 44392, 44394, 45384 and 45385. Information on concurrent AC use and primary indication for AC at the time of the colonoscopy were identified via the Dawn AC management system. Electronic membership databases were queried to identify patient age (on date of polypectomy), gender and KPCO membership eligility (including date of death when applicable). Integrated electronic medical office databases were queried to identify co-morbid diagnoses including intestinal arteriovenous malformations, previous history of hypertension, diabetes, stroke/transient ischemic attack, recent heart failure, active malignancy, bleeding and arterial or venous thrombosis.

Manual medical chart reviews were perfomed on all included patients to document information on indication for colonoscopy (anemia, colon cancer screening, constipation, diarrhea, irritable bowel, follow up from previous polypectomy and non-emergent rectal bleeding); number, location (ascending, transverse, descending, sigmoid and rectum) and size of polyps removed; description of polyp morphology (diminutive, flat, hyperplastic, laterally spreading, mixed, pedunculated and sessile); method of polyp removal (forceps, cold snare, snare cautery and combination of methods); and the use of aspirin and/or other antiplatelet therapy at the time of colonoscopy. A sum of all polyps removed was calculated for each patient. A global variable was created to indicate if the largest polyp

removed was > 10 mm (as per the endoscopist's dictated report).

Additional information collected for AC group patients only included indication for AC therapy, number of warfarin doses omitted prior to polypectomy, international normalized ratio (INR) value prior to polypectomy, use of unfractionated heparin or the low-molecular-weight heparin enoxaprin during the 7 days prior to and 30 days after polypectomy (bridging therapy), elapsed time (days) between polypectomy and resumption of warfarin (and enoxaparin if applicable) and INR value recorded at the time of anticoagulation therapy-related complications (when available and applicable).

During the 30 days after polypectomy, hospitalizations (including ED visits) were identified through queries of integrated electronic KPCO inpatient and ambulatory claims databases using predefined colon bleeding and thromboembolic event ICD-9 codes (available upon request). Thromboembolic complications included an embolic or thrombotic cerebrovascular accident, pulmonary embolism, deep vein thrombosis, or other systemic thromboembolic events. All events were verified and assessed for the association with AC therapy (where applicable) via manual medical chart review using a standardized abstraction form and a modified Naranjo scale. The Naranjo scale is a scoring system used to quantify the relationship between an adverse event and a given drug therapy [10]. All outcomes were reviewed by two adjudicators blinded to AC therapy status, disagreements regarding event categorization were resolved by a third adjudicator. Cause of death was verified through medical record and/or death certificate review.

Data analysis

Patient characteristics were reported as means, medians and standard deviations for interval-level variables and proportions for categorical variables. Colon bleeding rate was calculated by dividing the number of patients experiencing at least one bleeding episode by the total number of patients undergoing polypectomy. Rates of secondary outcomes were similarly calculated.

Conditional logistic regression with adjustment for the matching variable (age) and clustering of matched observations was used to assess differences in baseline characteristics and outcomes between groups. Wilcoxon rank-sum tests and Fisher's exact tests were used to assess differences between patients with and without a primary bleeding outcome for interval-level and categorical variables, respectively.

Patient AC status, age, gender, colonoscopy indications, co-morbidities and polyp variables (size, total count, retrieval method, and morphology) were entered into an unconditional, stepwise multivariate logistic regression model to identify independent predictors of colon bleeding after polypectomy among all patients. To preserve power, only variables with a $P < 0.30$ were entered into the model and kept in the final model with a $P < 0.10$. Two-way interactions were not included in the model because the primary outcome occurred infrequently. Unconditional regression was used because the incidence of colon bleeding among the clustered, matched patients was low. Similar predictive modeling was performed among AC group patients only; but with the addition of a variable indicating the use of enoxaparin.

Results

A total of 614 potential AC group patients were identified. Of these, 83 with insufficient medical record information, 61 with colonoscopy to evaluate emergent GI bleeding, 39 with warfarin monitored outside of CPAS, 4 with non-continuous health plan enrollment, 1 with intestinal arteriovenous malformation and 1 for warfarin prophylaxis after orthopedic surgery were excluded, resulting in 425 eligible-for-matching AC patients. A potential pool of 9219 non-AC group patients was identified; however, 125 and 91 were excluded for non-continuous health plan enrollment and colonoscopy to evaluate emergent GI bleeding, respectively. Of the remaining 9003 eligible-for-matching non-AC patients, 800 were randomly matched on age (± 3 years) to 425 AC patients.

Baseline characteristics for the AC and Non-AC groups are summarized in Table 1. The groups were well-matched with regards to age. Patients in the AC group were more likely to be male, to have had their colonoscopy for non-emergent rectal bleeding follow up, to have co-morbid heart failure and prior

Table 1 Baseline characteristics by anticoagulant use status at time of colonoscopy

Characteristic	AC group (n = 425)	Non-AC group (n = 800)	P-value*
Mean age† (SD)	69.5 (9.6)	69.6 (9.4)	0.484
Male (%)	61.9	53.5	0.003
Primary indication for colonoscopy (%)			
Anemia	7.8	5.4	0.064
Cancer screening	40.1	49.7	0.001
Constipation	0.0	0.6	0.998
Diarrhea	2.6	2.4	0.898
Irritable bowel	0.2	0.4	0.607
Previous polypectomy follow-up	27.1	26.0	0.624
Non-emergent rectal bleeding follow-up	18.2	13.6	0.043
Other	4.2	2.0	0.025
Antiplatelet use (%)			
Yes	19.1	6.1	< 0.001
Unknown	49.9	92.1	< 0.001
Co-morbidity (%)			
Diabetes mellitus‡	3.1	3.5	0.775
Hypertension‡	37.4	32.1	0.053
Heart failure‡	11.8	1.4	< 0.001
Prior venous thrombosis‡	10.4	3.0	< 0.001
Any prior hemorrhage‡	12.9	9.6	0.060
Prior stroke‡	0.0	0.0	1.000
Cancer‡	0.9	0.3	0.180

AC, anticoagulant therapy; SD, standard deviation. *Adjusted for clustering of matched observations and matching age variable. †As of date of colonoscopy. ‡During the 180 days prior to the colonoscopy.

Table 2 Outcomes during the 30-day follow-up period by anticoagulant use status at time of colonoscopy

Outcome	AC group (n = 425)	Non-AC group (n = 800)	P-value*
Colon bleeding (n, %)	11, 2.6	2, 0.2	0.005
Death (n, %)	1, 0.2	1, 0.1	0.624
Thrombosis (n, %)	1, 0.2	1, 0.1	0.624
Non-colon GI bleed (n, %)	1, 0.2	0, 0.0	0.347

AC, anticoagulant therapy; GI, gastrointestinal. *Adjusted for clustering of matched observations and age matching variable.

venous thrombosis (all $P < 0.05$). Patients in the Non-AC group were more likely to have had colonoscopy for cancer screening ($P = 0.001$).

Overall, 13 out of 1225 patients (1.06%) experienced post-polypectomy colon bleeding. Colon bleeding was more common in the AC group (2.6% vs. 0.2%, $P = 0.005$) (Table 2). Case summaries of all primary outcome events are presented in Table 3. An average of 9 days (median = 7, range = 3–18) elapsed between polypectomy and the primary outcome. No bleeding complications occurred during the colonoscopy. Secondary outcomes (thrombosis, death and non-colon GI bleeding) were uncommon and differences between groups were not statistically significant ($P > 0.05$ for all) (Table 2). Details pertaining to the quantity and morphology of polyps removed in the two groups as well as the polyp retrieval method are summarized in Table 4. Patients in the AC group were more likely to have had polyps retrieved using snare cautery whereas patients in the non-AC group were more likely to have had cold snare retrieval (both $P < 0.05$).

Patients hospitalized with post-polypectomy colon bleeding were more likely to have been receiving AC ($P < 0.001$), to be male ($P = 0.005$) and to have had a higher number of polyps removed ($P < 0.001$) (Table 5). Predictive modeling indicated that independent predictors of post-polypectomy colon bleeding included being in the AC group [adjusted odds ratio (AOR) = 11.6; 95% confidence interval (CI) = 2.3–57.3], number of polyps removed [AOR = 1.2 (i.e. 20% increased likelihood of bleeding for each additional polyp removed); 95% CI = 1.1–1.4] and male gender (AOR = 9.2, 95% CI = 1.1–74.9) (model fit c-statistic = 0.89). Age was included in the final model iteration but was not a significant predictor ($P = 0.13$). Among patients who received colonoscopy for colon cancer screening, the rates of colon bleeding were 1.76% and 0.50% in the AC and non-AC groups, respectively.

For patients in the AC group only, those experiencing post-polypectomy colon bleeding were more likely to have had a primary indication for AC of coronary artery disease compared with AC patients without bleeding (16.7% vs. 3.4%, $P = 0.034$). No other differences in AC group baseline characteristics were detected (all $P > 0.05$; data not shown). Patients in the AC group who experienced colon bleeding also had a higher number of polyps removed ($P < 0.001$) (Table 6).

For most AC group patients, warfarin was stopped 4 days prior to and resumed on the evening of colonoscopy. Differences in the use of bridging therapy between AC group patients who did and did not experience colon bleeding were not statistically significant. Among AC group patients, 5.6% ($n = 4$) of those who received bridging therapy after polypectomy experienced colon bleeding (all received therapeutic enoxaparin before and after the procedure) compared with 2.0% ($n = 7$) of those who did not receive bridging (Table 3). The most common bridging strategy was therapeutic enoxaparin before and after the procedure (56.6% of all bridging therapies). Most patients did not receive bridging therapy (Table 6). The mean INR on the day prior to colonoscopy was 1.4 ± 1.4 and 1.3 ± 1.2 in AC group patients with and without colon bleeding, respectively ($P > 0.05$).

For patients in the AC group, the number of polyps removed was the sole independent predictor of colon bleeding (AOR = 1.3; 95% CI = 1.1–1.6; model fit c-statistic = 0.86). The cold snare retrieval method ($P = 0.070$) and male gender ($P = 0.088$) were included in the final multivariate analysis but were not independently associated with bleeding. The mean INR at the time of post-polypectomy bleeding for AC patients was 2.3 (median = 2.4, range = 1.4–3.1).

Discussion

In our cohort of 1225 patients undergoing polypectomy during routine colonoscopy (425 of which interrupted AC therapy for the procedure), we detected an overall incidence of subsequent colon bleeding episodes related to the procedure of 11 per 1000 polypectomies. However, the majority of these complications (85%) were observed in patients receiving warfarin (26 per 1000 AC patients undergoing polypectomy). Our observations suggest that patients interrupting warfarin are at a substantially increased risk (elevenfold higher) of colon bleeding after polypectomy compared with patients not prescribed AC. This confirms observations from the aforementioned studies involving limited AC populations that also reported high rates of post-polypectomy colon bleeding in AC patients [5,6].

In addition to AC use, we identified other factors that independently predict colon bleeding in patients undergoing polypectomy including male gender and the number of polyps retrieved. Greater numbers of polyps may be more common in men and, therefore, compared with women, more polyps are likely removed during colonoscopy [2]. We found the count of polyps was a predictor independent of patient gender. We were unable to address if there is an interaction between gender and count of polyps in our model as a result of the low incidence of bleeding in our cohorts. Nevertheless, we did find that men had a higher mean count of polyps removed than women (2.6 vs. 2.2, $P < 0.001$). Although data regarding polyp location were collected, they were not included in regression modeling because many patients had polyps removed from multiple locations and we were unable to link the removal site to the bleeding site. Among AC group patients, only the number of

Table 3 Summary of all colon bleeding events

Pre-polypectomy INR	Event INR	Enoxaparin bridge therapy	ASA dose	AC indication	LOS (days)	Event description	Transfusion (# of units)
Non-AC group patients							
na	na	None	UNK	None	6	Developed venous thrombosis following polypectomy, had just started enoxaparin and warfarin (INR = 1.0) when presented 11 days after polypectomy with hemodynamically stable BRBPR, and ↓ Hgb ≥ 2 g dL^{-1}, intact polypecto my sites without bleeding noted on exam	3
na	na	None	UNK	None	1	Presented 7 days after polypectomy with bloody diarrhea and ↓ Hgb ≥ 2 g dL^{-1}, observed in ED overnight, two ulcers at polypectomy sites noted on exam and cauterized	0
AC group patients							
1.7	3.0	Therapeutic enoxaparin pre- and post-polypectomy	UNK	Atrial fibrillation	2	Daily rectal bleeding reported following polypectomy, presented 6 days after polypectomy with ↓ Hgb ≥ 2 g dL^{-1}, bleeding resolved without intervention during hospital admission	1
1.5	UNK	None	UNK	Atrial fibrillation	10	Presented 3 days after polypectomy with hypotensive BRBPR and ↓ Hgb ≥ g dL^{-1}, required hemicolec tomy complicated by two episodes of recurrent bleeding	18
1.3	2.4	None	81 mg	Dialysis access thrombosis	2	Presented 18 days after polypectomy with maroon stools, lightheadedness and ↓ Hgb ≥ 2 g dL^{-1}, small ulcer right colon without active bleeding no ted on exam	2
1.1	3.1	Therapeutic enoxaparin pre- and post-polypectomy	UNK	Atrial fibrillation	2	Presented 7 days after polypectomy with hypoten sion, exam revealed bleeding from a previously cauterized lesion, bleeding resolved following repeat cauterization	2
1.3	2.2	None	UNK	Atrial flutter	0	Presented 6 days after polypectomy with BRBPR and ↓ Hgb ≥ 2 g dL^{-1}; bleeding resolved without intervention during ED observation	0
1.4	2.54	Therapeutic enoxparin pre- and post-polypectomy	None	Mechanical mitral valve	2	Presented 10 days after polypectomy with bloody diarrhea, a bleeding ulceration at polypectomy site was noted on exam and clipped	0
1.9	2.47	None	325 mg	Coronary artery disease	9	Presented 15 days after polypectomy with acute worsening of chronic blood in stools, lightheaded ness and ↓ Hgb ≥ 2 g dL^{-1}, bleeding polypectomy ulcers and an antral ulcer were noted on exam and cauterized	3
1.4	1.9	None	81 mg	Atrial fibrillation	6	Presented 10 days after polypectomy with rectal bleeding and ↓ Hgb ≥ 2 g dL^{-1}, a cecal ulcer was noted on exam and injected and clipped, subse quently required repeat colonoscopy (× 3) due to persistent bleeding	> 10

Table 3 Summary of all colon bleeding events

Pre-polypectomy INR	Event INR	Enoxaparin bridge therapy	ASA dose	AC indication	LOS (days)	Event description	Transfusion (# of units)
UNK	1.4	None	325 mg	Coronary artery disease	1	Presented 4 days after polypectomy with rectal bleeding and ↓ Hgb ≥ 2 g dL^{-1}, bleeding at polypectomy site noted on exam and injected, resection of cecum required 2 weeks later due to large persistent pneumoperitoneum	5
UNK	1.73	None	UNK	Pulmonary embolism	1	Presented 13 days after polypectomy with abdominal pain and BRBPR and ↓ Hgb ≥ 2 g dL^{-1}, bleeding resolved following discontinuation of warfarin	0
1.3	2.51	Therapeutic enoxaparin pre- and post-polypectomy	UNK	Transient ischemic attack	4	Presented 7 days after polypectomy with hypotension, dizziness and syncope and ↓ Hgb ≥ 2, became non-responsive with respiratory arrest requiring resuscitation, no active bleeding from polypectomy sites noted on exam	9

ASA, aspirin; AC, anticoagulation; LOS, length of stay; na, not applicable; UNK, unknown; Hgb, hemoglobin; BRBPR, bright red blood per rectum; ED, emergency department.

Table 4 Polyp characteristics by anticoagulant use status at time of colonoscopy

Characteristic	AC group (n = 425)	Non-AC group (n = 800)	P-value*
Largest polyp > 10 mm (% yes)	21.7	17.9	0.100
Mean total count of polyps removed (median, SD)	2.6 (2.0, 2.3)	2.4 (2.0, 2.4)	0.103
Gross polyp morphology (%)			
Diminutive	5.5	6.6	0.297
Flat	0.2	0.6	0.403
Hyperplastic	1.0	1.4	0.472
Laterally spreading	0.0	0.1	0.999
Mixed polyp types	12.2	13.3	0.543
Pedunculated	9.8	5.6	0.008
Sessile	48.5	48.0	0.973
Not stated	22.9	24.3	0.535
Polyp-retrieval method (%)			
Cold biopsy forceps	8.1	6.9	0.483
Cold snare	2.4	21.2	< 0.001
Combination of methods	17.1	13.0	0.045
Snare cautery	57.8	43.4	< 0.001
Other	2.8	3.3	0.609
Not stated	11.9	12.3	0.854

AC, anticoagulant therapy. *Adjusted for clustering of matched observations and age matching variable.

polyps removed independently predicted subsequent colon-bleeding, suggesting that the likelihood of post-polypectomy colon bleeding increases by 30% for each polyp removed.

To date, recommendations on the management of patients receiving AC during the period surrounding colonoscopy are limited [4]. Recommendations regarding the need for bridging therapy during warfarin interruption vary for polypectomy (a high-bleeding risk procedure) based on the underlying risk of thromboembolism [11]. The standardized approach employed by KPCO CPAS pharmacists reserves bridging therapy for only those patients at highest thromboembolic risk, such as those with mechanical heart valves or atrial fibrillation with previous thromboembolic events [4,11]. For AC patients with post-polypectomy colon bleeding in our study, a median of 4 (SD = 1.4) doses of warfarin were withheld prior to colonoscopy resulting in a median pre-procdedure INR of 1.4 (SD = 0.2). Warfarin was resumed on the evening of colonoscopy for the vast majority of AC patients; loading doses of warfarin were used commonly when therapy was resumed. However, AC patients who did and did not experience post-polypectomy colon bleeding were managed similarly (Table 6). For patients who received bridging therapy, the last dose of enoxaparin was administered approximately 24 h prior to colonoscopy and resumed, unless otherwise directed by the endoscopist, the morning after the procedure. Both once and twice daily doses of enoxaparin were used. The small number of bleeding events precludes any conclusion about what (if any) increase in the risk of colon bleeding might be attributed to enoxaparin use for bridging therapy.

Although bleeding was the primary focus of this study, preventing thromboembolic complications during AC therapy

Table 5 Potential colon bleeding predictors by colon bleeding status

Characteristic	Bleeding (n = 13)	No bleeding (n = 1212)	P-value
AC group status (%)	84.6	34.2	< 0.001
Mean age* (SD)	72.8 (6.4)	69.5 (9.5)	0.208
Male (%)	92.3	56.0	0.009
Primary indication for colonoscopy (%)			
Non-emergent rectal bleeding	15.4	15.2	1.000
Colon cancer screening	38.5	46.4	0.781
Mean total count of polyps removed (median, SD)	5.2 (5.0, 2.4)	2.4 (2.0, 2.4)	< 0.001
Risk factors (%)			
Hypertension†	38.5	33.9	0.772
Heart failure†	15.4	4.9	0.134
Prior venous thrombosis†	7.7	5.5	0.552
Prior hemorrhage†	7.7	10.8	0.719
Pedunculated polyp (%)	0.0	7.1	0.319
Largest polyp > 10 mm (% yes)	38.5	18.9	0.085
Polyp-retrieval method (%)			
Cold snare	15.4	14.7	0.943
Combination	30.8	14.2	0.103
Snare cautery	38.5	48.3	0.582

AC, anticoagulant therapy. *As of date of colonoscopy. †During the 180 days prior to the colonoscopy.

interruption was also assessed. Thrombotic events were uncommon and not significantly different between AC and non-AC groups. Althought this observation suggests the strategies used to interrupt AC therapy in this study minimized the likelihood of thromboembolic complications, the very small number of events precludes definitive conclusions regarding this outcome.

Our conclusions are potentially confounded by the lack of information regarding aspirin use, particularly in the non-AC group. We performed two *post-hoc* regression analyses with aspirin use added to the model under the assumption that all patients with unknown aspirin use status were either (i) not taking aspirin or (ii) were all taking aspirin. Neither of these assumptions altered the results of our origninal analysis to an appreciable extent. At KPCO, discontinuation of aspirin or non-steroidal anti-inflammatory drugs (NSAIDs) prior to colonoscopy is managed on a case-by-case basis. It is possible that antiplatelet agents were used more commonly by patients in the AC group and that some of the excess bleeding in that cohort could be attributable to antiplatelet therapy (rather than warfarin). That possibility notwithstanding, several studies have reported that the concurrent use of aspirin or NSAIDs does not increase the risk of post-polypectomy bleeding [5,6,12,13]. Current American Society for Gastrointestinal Endoscopy guidelines recommend that, in the absence of a pre-existing bleeding disorder, endoscopic procedures may be performed on patients taking aspirin and NSAIDs in standard doses [4]. However, no information on the concommitant use of aspirin, NSAIDs and/or AC has been reported; nevertheless, it is plausible that interactions between these agents may increase bleeding.

The small number of primary outcome events limits our ability to discern predictors of post-polypectomy colon bleeding. Nevertheless, to our knowledge, this is by far the largest evaluation of post-polypectomy bleeding in AC patients. In addition, we included a comprehensive set of predictive factors to evaluate. Finally, as KPCO employs a centralized anticoagulation management program, processes may have been in place that decreased the risk of bleeding. Other healthcare systems without anticoagulation management service may experience different rates of post-polypectomy colon bleeding among their AC patients.

Our results are likely to be valid. We examined real-world patients with a variety of indications for warfarin and colonoscopy. The dataset used to complete this study is

Table 6 Anticoagulant patient characteristics overall and by primary bleeding status

Characteristic	Overall (n = 425)	Bleeding (n = 11)	No bleeding (n = 414)
Mean warfarin doses held prior to colonoscopy (n, median, SD)	3.6 (404, 4.0, 1.0)	3.6 (9, 4.0, 1.4)	3.6 (395, 4.0, 1.0)
Bridging therapy (%)			
Prior to but not after colonoscopy*	0.9	0.0	1.0
After but not prior to colonoscopy†	6.4	0.0	6.5
Prior to and after colonoscopy‡	15.8	36.4	15.2
None	76.9	63.6	77.3
Mean INR prior to colonoscopy (n, median, SD)	1.3 (388, 1.2, 0.2)	1.4 (9, 1.4, 0.2)	1.3 (379, 1.2, 0.2)
Resumed warfarin with a loading dose after colonoscopy (%)	72.3	54.6	72.8
Mean days to re-starting warfarin after colonoscopy (n, median, SD)	0.5 (412, 0, 1.5)	0.5 (11, 0, 1.0)	0.5 (400, 0, 1.5)
Mean total count of polyps removed (median, SD)	2.6 (2.0, 2.3)	5.4 (5.0, 2.4)§	2.5 (2.0, 2.3)
Largest polyp > 10 mm (% yes)	21.7	27.3	21.5
Cold snare polyp-retrieval (%)	2.4	9.1	2.2
Mean days to enoxaparin after colonoscopy (n, median, SD)	1.1 (89, 1.0, 0.8)	1.0 (4, 1.0, 0.0)	1.1 (85, 1.0, 0.8)

SD, standard deviation; INR, international normalized ratio. *Therapeutic dose enoxaparin (n = 1), other enoxaparin dose (n = 3). †Prophylactic dose enoxaparin (n = 14), therapeutic dose enoxaparin (n = 10), other enoxaparin dose (n = 4). ‡Therapeutic dose enoxaparin before and after (n = 56), prophylactic dose enoxaparin before and after (n = 6), therapeutic dose enoxaparin before prophylactic dose enoxaparin after (n = 2); intravenous unfractionated heparin before and after, other enoxaparin dose before and after, and prophylactic dose enoxaparin before therapeutic dose enoxaparin after (n = 1 each). §P < 0.001.

robust and has been used previously in health records and data extraction research [8,14]. 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 all clinical events were independently assessed for causality by two expert reviewers.

Conclusions

Exposure to AC therapy around the time of colonoscopy with polypectomy increased the likelihood of post-polypectomy colon bleeding by approximately elevenfold. As no bleeding complications occurred during colonoscopy, it appears that a strategy of witholding warfarin to achieve an INR below 1.5 and giving any pre-procedure bridging doses of enoxaparin at least 24 h prior to colonoscopy is sufficient to allow safe polyp retrieval. However, prompt resumption of warfarin and/or enoxaparin therapy after colonoscopy confers an increased likelihood of colon bleeding in the 30 days post procedure. As each polyp retrieved increased bleeding likelihood by about 20%, resumption of AC therapy could be delayed selectively when removal of a large number of polyps is required, especially in men. We are unable to determine if delaying the resumption of AC therapy in this manner would attenuate bleeding risk or would unacceptably increase the risk of thromboembolism. Local strategies to control bleeding such as endoscopic clip application immediately after polypectomy might also reduce bleeding risk in patients requiring resumption of AC therapy. These strategies should be explored in future research.

Addendum

Contribution: D.M. Witt, T. Delate, K.H. McCool, M.B. Dowd, and N.P. Clark designed the research and extracted information from medical records; T. Delate performed the statistical analysis; D.M. Witt and F. Dentali wrote the initial draft of the manuscript; D.M. Witt, T. Delate, K.H. McCool, M.B. Dowd, N.P. Clark, M.A. Crowther, D.A. Garcia, W. Ageno, F. Dentali, E.M. Hylek and W.G. Rector interpreted the analysis and revised the manuscript.

Acknowledgements

The authors would like to acknowledge M. L. Smith, JD, PharmD, BCPS, M. Kershner and D. Winn for their contributions to the data collection. D.M. Witt had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. This study was funded in its entirety by Kaiser Permanente Colorado.

Disclosure of Conflict of Interests

The authors state that they have no conflict of interest.

References

1 Pignone M, Rich M, Teutsch SM, Berg AO, Lohr KN. Screening for colorectal cancer in adults at average risk: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002; **137**: 132–41.
2 Rabeneck L, Paszat LF, Hilsden RJ, *et al.* Bleeding and perforation after outpatient colonoscopy and their risk factors in usual clinical practice. *Gastroenterology* 2008; **135**: 1899–906.
3 Levin TR, Zhao W, Conell C, Rabeneck L, Paszat LF, Hilsden RJ, Saskin R, Leddin D, Grunfeld E, Wai E, Goldwasser M, Sutradhar R, Stukel TA. Complications of colonoscopy in an integrated health care delivery system. *Ann Intern Med* 2006; **145**: 880–6.
4 Eisen GM, Baron TH, Dominitz JA, *et al.* Guideline on the management of anticoagulation and antiplatelet therapy for endoscopic procedures. *Gastrointest Endosc* 2002; **55**: 775–9.
5 Hui AJ, Wong RM, Ching JY, *et al.* Risk of colonoscopic polypectomy bleeding with anticoagulants and antiplatelet agents: analysis of 1657 cases. *Gastrointest Endosc* 2004; **59**: 44–8.
6 Kim HS, Kim TI, Kim WH, *et al.* Risk factors for immediate postpolypectomy bleeding of the colon: a multicenter study. *Am J Gastroenterol* 2006; **101**: 1333–41.
7 Timothy SK, Hicks TC, Opelka FG, Timmcke AE, Beck DE. Colonoscopy in the patient requiring anticoagulation. *Dis Colon Rectum* 2001; **44**: 1845–8.
8 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.
9 Rathgaber SW, Wick TM. Colonoscopy completion and complication rates in a community gastroenterology practice. *Gastrointest Endosc* 2006; **64**: 556–62.
10 Naranjo CA, Busto U, Sellers EM, *et al.* A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; **30**: 239–45.
11 Douketis JD, Berger PB, Dunn AS, *et al.* The perioperative management of antithrombotic therapy: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 2008; **133**: 299S–339S.
12 Gibbs DH, Opelka FG, Beck DE, *et al.* Postpolypectomy colonic hemorrhage. *Dis Colon Rectum* 1996; **39**: 806–10.
13 Yousfi M, Gostout CJ, Baron TH, *et al.* Postpolypectomy lower gastrointestinal bleeding: potential role of aspirin. *Am J Gastroenterol* 2004; **99**: 1785–9.
14 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**: 960–7.

Use of Patient Self-Report Oral Health Outcome Measures in Assessment of Dental Treatment Outcomes

Wanda G. Wright, DDS, MS, MSD; Judith A. Jones, DDS, MPH, DScD; Avron Spiro III, PhD; Sharron E. Rich, MPH; Nancy R. Kressin, PhD

Abstract

Objective: *To assess the sensitivity of a newly developed brief measure of oral health-related quality of life (OQOL).* ***Methods:*** *Self-assessed oral health and OQOL were measured in three groups of patients who had presented for either prophylaxis (n = 32), endodontic care (n = 15), or for a denture (n = 16) in a dental school setting before and after treatment. Main outcome measures included the single-item self-report of oral health (OH-1) and the 6- and 12-item versions of a new OQOL instrument. General linear modeling was used to compute means of self-reported oral health by treatment group.* ***Results:*** *Of the 63 patients who completed the baseline questionnaire, 44 (70 percent) returned questionnaires after treatment. The sample averaged 43 ± 15 years, 48 percent male and 55 percent with some college education. Ethnic representation included 35 percent White, 33 percent Black, and 32 percent other – mostly Latino. The mean self-reported number of teeth was 20.6. In terms of sensitivity, significant differences were observed between the treatment groups on the items assessing being upset (P < 0.05), feeling depressed (P < 0.05), and uncomfortable about the appearance of teeth or dentures (P < 0.05). However, magnitude of change, as measured by an effect size, was characterized as minimal to small in the recall and endodontic groups and borderline moderate in the denture group.* ***Conclusion:*** *The measure was sensitive to differences within groups, with a small to borderline magnitude of change.*

Key Words: oral health, quality of life, sensitivity, dental treatment, outcome measures

Introduction

Oral health-related quality of life (OQOL) represents the self-perceived impact of oral conditions on daily functioning and well-being. These functions include physical, psychological, and social functioning; performance of self-care; perceived health and symptomology; and the presence of pain or distress (1). Combined with clinical and other indices, these measures of oral health status help provide a more comprehensive assessment of an individual's overall health.

Over the past several decades several OQOL instruments have been developed and have undergone considerable testing, establishing their validity and reliability (2). These instruments have been used to describe the impact of disease on patient's daily functioning (3); and as outcome measures to evaluate the effectiveness of interventions (4). Less attention has been paid to the use of these instruments to measure the magnitude and extent of longitudinal change. Only a few population-based studies have examinedchanges in OQOL as a result of dental intervention, and much of this research has focused on the replacement of teeth using conventional or implantable prosthesis (5-8), temporomandibular joint dysfunction (9), or to evaluate dental care programs (10). Failure to consider a broader range of interventions is an important gap in our knowledge base given the increased recognition of the importance of these instruments for quantifying the treatment benefit in clinical trials or investigating the impact of illness over time (11). If OQOL instruments are to be of value in assessing outcomes from clinical interventions, then their ability to describe the magnitude and extent of change must be determined (12). Furthermore, establishing the sensitivity of OQOL instruments would assist investigators in selecting the most appropriate measure and assist health professionals to interpret the meaning of changes in scores derived from the instrument.

The present study of dental school patients examines the impact of treatment of dental conditions on patients' quality of life. These data provide us the opportunity to assess the sensitivity to change of a newly developed brief measure of quality of life. This new measure, consisting of 6- and 12-item scales, is comprised of items from three existing OQOL measures: the Oral Health

Send correspondence and reprint requests to Wanda G. Wright, Department of Health Policy & Health Services Research, Boston University Goldman School of Dental Medicine, 715 Albany Street, 560, no. 342, Boston, MA 02118. Tel.: 617-638-4914; Fax: 617-638-6381; e-mail: wgwright@bu.edu. Wanda G. Wright and Sharron E. Rich are with the Department of Health Policy & Health Services Research, Boston University Goldman School of Dental Medicine. Judith A. Jones is with the Department of General Dentistry, Boston University School of Dental Medicine. Avron Spiro III is with the Massachusetts' Veterans Epidemiologic Research and Information Center, Veterans Affairs Boston Healthcare System; the Health Policy & Health Services Research, Boston University School of Dental Medicine; and the Department of Epidemiology, Boston University School of Public Health. Nancy R. Kressin is with the Center for Health Quality, Outcomes, & Economic Research, Edith Nourse Rogers Memorial Veterans Hospital; and the Department of General Internal Medicine, Boston University School of Medicine. Sources of Support: NICDR Grant U54 DE 14264/K24 DE 00419, Delta Dental Plans of America. Presented as a poster at the 2004 NOHC Meeting. Manuscript received: 10/15/07; accepted for publication: 8/27/08.

DOI: 10.1111/j.1752-7325.2008.00106.x

Quality of Life instrument (OHQOL) (13), the General Oral Health Assessment Index (GOHAI) (14), and the Oral Health Impact Profile (OHIP) (15). Both the 6- and 12-item scales of this new measure have demonstrated sound psychometric properties including excellent validity, reliability, internal consistency, and limited floor and ceiling effects (16). Both the 6- and 12-item scales can be used in the clinical setting as an outcome measure. However, the observed differences in internal consistency reliability suggest that the 6-item scale is appropriate for use in comparing groups of patients while the 12-item scale is appropriate for use in assessing outcomes among individual patients (16).

The present work using this new 6- and 12-item measure was developed based on a conceptual model of health and quality of life proposed by Patrick and Erickson (17) and applied to oral health by Gift and Atchison (18,19). This model contains five health-related quality of life concepts including a) opportunity; b) perceptions; c) three functional states: physical, social, and psychological functioning; d) impairments; and e) survival. Opportunity reflects the impact that oral health has on one's ability to function in social and work roles. Perceptions include self-rating of oral health as well as satisfaction with this self-rating and perceived need for treatment. Physical functioning of the teeth and oral cavity includes activity restrictions such as a patient's ability to eat, chew, speak, or sleep without discomfort. Social functioning includes the impact of oral health on social roles such as speaking, smiling, eating in public, and being able to meet one's obligations such as work and family responsibilities. Psychological functioning includes a patient's satisfaction with the esthetics of their dentition, comfort with interpersonal relations as well as worry, concern, embarrassment about, or lack of confidence because of problems with teeth or gums. Impairments include self-reported symptoms or other indication of discomfort or pain. Finally, survival can be measured by tooth loss or mortality, e.g., from oral cancer.

Our aim was to assess the sensitivity to change of this new measure as a consequence of dental interventions on three groups of dental patients: denture replacement, endodontic, and patients on recall with no apparent disease. These groups were selected based on the rationale that patients with more severe conditions (endodontic and denture replacement) would have a greater level of change in their oral health status, and that OQOL indicators should be more sensitive to the impact of these conditions. We expected little change in the self-reported oral health status of the recall group. Baseline and follow-up oral quality of life scores were evaluated using both the 6- and 12-item OQOL scales (summary scores and individual items in each), as well as a global assessment of OQOL.

Materials and Methods

Design. We used a repeated measure design to examine the effects of dental treatment on patient-assessed outcomes of dental care in three groups of patients (recall, endodontic, or denture) in a dental school setting. This design is characterized by having more than one measurement of at least one variable for each subject. We compared baseline (immediately before treatment) and follow-up (3 months after treatment) quality of life scores. The study protocol received approval from the Institutional Review Board at Boston University. Consent to participate was obtained according to Institutional Review Board requirements.

Setting and Participants. Participants were a convenience sample of adults aged ≥18 years seeking treatment in the dental clinics at Boston University Goldman School of Dental Medicine, Boston, Massachusetts, between June 2002 and May 2004. Three categories of patients were recruited: those coming to the general dentistry clinics for "recall" visits for checkups and prophylaxis (RECALL group), those presenting to the endodontic clinic in pain (ENDO group), and patients presenting to the prosthodontic clinic for removable dentures (DENTURE group). Patients were excluded if they were less than 18 years of age or if they had a diagnosis other than those specified earlier. Research staff recruited subjects in the waiting areas of the three clinics from Monday through Friday between 9 AM and 5 PM. All treatments were provided by undergraduate dental students under supervision by staff from the General Dentistry Department at Boston University Goldman School of Dental Medicine.

Data Collection. At prearranged times, research staff visited each clinic (general dentistry, endodontic, prosthodontic) and attempted to recruit subjects in the clinic waiting areas. Of the 122 subjects approached (RECALL = 51, ENDO = 39, DENTURE = 32), 43 subjects declined study participation (RECALL = 14, ENDO = 17, DENTURE = 12). The reasons provided for declining were inconvenience ($n = 20$), no time ($n = 14$), or no reason given ($n = 9$). Five recall patients were under the age of 18 years and were excluded. Four DENTURE patients presented to the clinic for reasons other than the insertion of a removable prosthesis, and seven of the endodontic patients were not undergoing emergency endodontic treatment and were excluded. Ultimately, 63 patients (RECALL = 32, ENDO = 15, DENTURE = 16) participated, 44 completed follow-up.

Subjects in the recall and endodontic groups were asked to complete the baseline questionnaire at their initial dental visit. Patients in the denture group were asked to complete the baseline questionnaire immediately prior to the insertion of the removable prosthesis. All patients were asked to complete the instrument again 3 months later and return it by mail in the stamped addressed envelope provided. We attempted to reach all patients by phone as a reminder to send in the second questionnaire. We sent a second copy of

the questionnaire to subjects not returning their questionnaire within the designated time period.

OQOL Measures. OQOL (16) was measured using the 6- and 12-item scales developed from items in the OHQOL, the GOHAI, and the OHIP. For details of development, please see the original article (16). Briefly, the 64 items from the combined three quality of life surveys were administered along with a clinical oral examination to two veterans study samples ($n = 827$). The hypothesized framework included four primary dimensions: physical function, psychosocial function (with three subdimensions of role function, distress, and worry), impairment, and perceptions.

Each of the 64 items from the three OQOL instruments was independently categorized into one of the theoretical domains by the authors. An iterative series of multitrait scaling analysis was conducted to examine the fit of the items to the hypothesized domains (20). These analyses examine item-level characteristics including internal consistency, equality of item-scale correlations, and discriminant validity. The results provide information about scale distribution characteristics, reliability of scale scores, and correlations among hypothesized scales.

The conceptual model was altered to include five dimensions: physical function, impairment/disease, and three dimensions of psychosocial function: role function, distress, and worry. Five scales to correspond with the said dimensions, a separate denture subscale (3 items), and a summary scale comprised of all items were created.

Forward stepwise regression was conducted to develop a short-form version of the measure. For each scale, the total scale score was used as the dependent variable, with data from the two veteran samples. Items that explained either 80 percent of the variance or the first five items, whichever was greater, were selected resulting in five scales, each with five items. All of the scales had excellent internal consistency reliability, ranging from 0.78 to 0.92.

The five scales and three denture items were then administered to a sample of dental patients ($n = 113$). Using multitrait analysis, the number of items was reduced further by eliminating items contributing least to each scale's internal consistency reliability and retaining items which conceptually best represented the subscale. One 12-item measure (Cronbach's alpha = 0.90) and a second 6-item measure (Cronbach's alpha = 0.80) were developed.

The association of the 6- and 12-item measures with clinical indices was examined using the clinical data from the two sets of veterans. Both scales were significantly correlated overall with number of teeth ($r = 0.35$ and -0.23, for the 6- and 12-item scales, respectively), coronal decay ($r = 0.09$ and 0.14), periodontal status ($r = 0.19$ and 0.20), and root caries ($r = 0.14$ and 0.12) (14). The associations detected between the 6- and 12-item scales and clinical indices are similar to those of other published findings (21,22).

The 12-item measure contains 3-item subscales for three scales: distress, worry and social function (role), and single items assessing dimensions titled physical function, denture, and pain. The 6-item measure includes single items assessing distress, worry, social function, physical function, denture, and pain as listed (see Table 1 for the scales).

We also included a separate single item that is not part of the 6- and 12-item scales. This 5-point global self-report of oral health (OH-1) has been used in prior studies and asks, "How would you describe the health of your teeth and gums? Would you say it is excellent, very good, good, fair or poor?" Responses are scored from 1 (excellent) to 5 (poor).

Scoring. Some items (GOHAI 10 and OHQOL B31) were reversed so that a higher score consistently indicated worse oral quality of life. GOHAI 10 score of 3 becomes a 1, score of 2 remains a 2, score of 1 becomes a 3; and OHQOL B31 score of 5 becomes a 1, score of 4 becomes a 2, score of 3 remains a 3, score of 2 becomes a 4, and score of 1 becomes a 5.

All item scores were then converted to a scale of 0-100. Items scored on a 0-4 scale (all OHIP items) were converted to a scale of 0-100 by having 0 = 0, 1 = 25, 2 = 50, 3 = 75, and 4 = 100. The item initially scored on a 1-3 scale (GOHAI 10) was converted as follows after the item was reversed: 1 = 0, 2 = 50, and 3 = 100. Items scored on a 1-5 scale (OHQOL B31 and OH-1) were converted to a scale of 0-100 by having 1 = 0, 2 = 25, 3 = 50, 4 = 75, and 5 = 100.

Final scores for each scale (6- and 12-item) and subscales (distress, worry, and social function) were created by computing the mean of the responses to items represented by each scale.

Analysis. A combination of bivariable and multivariable statistical methods was used for this analysis. We measured differences in mean age by group using analysis of variance (ANOVA). Categorical demographic variables were tested using chi-squared and Fisher's exact tests to examine differences between groups. Demographic variables that differed between groups were adjusted for in the multivariate analysis. Baseline OQOL scores were computed and compared using ANOVA. Duncan's multiple range test was used to control the Type I error rate. We used general linear modeling to examine between-group effects in OQOL scores, mean change scores were reported as least squares means after adjusting for baseline OQOL score and covariates such as age and gender. Change scores were derived for individual items and total scores by subtracting posttreatment scores from baseline scores. Positive scores indicated an improvement and negative scores indicated deterioration following treatment. The magnitude of change was assessed as an effect size, calculated by dividing the mean of change scores by the standard deviation (SD) of the related baseline score (23).

Table 1
6- and 12-Item Short-Form Oral Health-Related Quality of Life Measures

During the past 3 months how often have you experienced the following difficulties because of problems with your teeth, mouth, or dentures? (Circle one answer)	Never	Hardly ever	Occasionally	Fairly often	Very often
1. Have you had to avoid eating some foods? (Physical function; OHIP 28)*	0	1	2	3	4
2. Have you found it difficult to relax? (Distress; OHIP 35)*	0	1	2	3	4
3. Have you felt depressed? (Distress; OHIP 36)	0	1	2	3	4
4. Have you been upset? (Distress; OHIP 34)	0	1	2	3	4
5. Have you felt uncomfortable about the appearance of your teeth, mouth, or dentures? (Worry; OHIP22)	0	1	2	3	4
6. Have you been worried by dental problems? (Worry; OHIP 19)	0	1	2	3	4
7. Have you had trouble getting along with other people? (Social function; OHIP 41)	0	1	2	3	4
8. Have you avoided going out? (Social function; OHIP 39)*	0	1	2	3	4
9. Have you been totally unable to function? (Social function; OHIP 48)	0	1	2	3	4

	Never	Sometimes	Always
10. In the past 3 months, how often did you feel nervous or self-conscious because of problems with your teeth, gums, or dentures? (Worry; GOHAI 10)*	1	2	3

	None at all	A little bit	Some	Quite a bit	A great deal
11. During the past 3 months, how much pain or distress has your teeth or gums caused you? (Pain; OHQOL B31)*	1	2	3	4	5

If you have removable denture appliances, please answer the following question:

During the past 3 months, how often have you had the following problem with your dentures?	Never	Hardly ever	Occasionally	Fairly often	Very often
12. Have you had uncomfortable dentures? (Denture; OHIP 18)*	0	1	2	3	4

* Indicate items in 6-item measure.
OHIP, Oral Health Impact Profile; GOHAI, General Oral Health Assessment Index; OHQOL, Oral Health Quality of Life.

All analyses were conducted in SAS version 9.1 (SAS Corporation, Cary, NC, USA). We used $P < 0.05$ as a cutoff for statistical significance.

Results

Characteristics of Participants. At baseline, 63 subjects participated, 32 in the recall group, 15 in the endodontic group, and 16 in the denture group. The sociodemographic characteristics (see Table 2) of the baseline sample demonstrate the expected association of age with treatment group; denture patients tended to be older whereas the recall group tended to be younger ($P = 0.0002$) and more highly educated ($P = 0.003$). The mean number of teeth by self-report was 26 in the recall group, 23 in the endodontic group, and 7.0 in the denture group. Approximately 31 percent of the denture group were completely edentulous.

Of the participants who completed the baseline questionnaire, 44 (70 percent) returned questionnaires after treatment. This sample averaged 45 ± 15.8 (SD) years, 49 percent male and 48 percent with some college education. Ethnic representation included 30 percent White, 38 percent Black, and 32 percent other. There were no differences on any of these dimensions between those who remained in the study and those who did not. Overall, the age, gender, race, and educational status of the sample remained constant over the period of the study (Table 2). No information was collected on those declining participation in the study.

Baseline Quality of Life Scores. There were no significant differences between the three groups in terms of their summary OQOL or subscale scores at baseline. Details of the summary and individual 6- and 12-item baseline scores for each group are shown in Table 3. Baseline scores for some individual items varied by group, in particular, items assessing pain and distress ($P = 0.0001$) and worry ($P = 0.05$). The

Table 2
Demographic Characteristics of Participants

	Baseline			
	Total ($n = 63$)	RECALL* ($n = 32$)	ENDO† ($n = 15$)	DENTURE‡ ($n = 16$)
Age (mean and SD)	43 (14.9)	38.5 (13.5)[A]	40.9 (16.0)[B]	55.8 (8.3)[B]
Gender (%)				
Female	52	50	47	63
Male	48	50	53	38
Race (%)				
White	35	44	27	25
Black	33	25	27	56
Other	32	31	46	19
Education (%)				
High school graduate or less	46	25[A]	67[B]	69[B]
Some college	54	75[A]	33[B]	31[B]

	Follow-up	
	Participants returning second survey ($n = 44$)	Participants not returning second survey ($n = 19$)
Age (mean and SD)	44.9 (15.8)	40.2 (12.3)
Gender (%)		
Female	50	42
Male	50	58
Race (%)		
White	30	47
Black	38	21
Other	32	32
Education (%)		
High school graduate or less	77	89
Some college	23	11

Values with same superscripts are not significantly different ($P > 0.05$ using Duncan's test).
* Regular users of diagnostic and preventive care and presented for a prophylaxis.
† Presented to the dental school setting in pain; to undergo emergency endodontic treatment.
‡ Will receive a removable prosthesis.
SD, standard deviation.

subjects in the endodontic group expressed more pain than subjects in the recall or denture groups. The denture group expressed less worry than the endodontic or recall groups.

Follow-up Quality of Life Scores. There were no significant differences between the three groups in terms of their follow-up, OH-1 scale, or subscale scores (Table 4). However, there were differences on individual items assessing being upset between the endodontic group and the denture group ($P = 0.05$) and for the item assessing being depressed between the denture group and the recall group ($P = 0.05$). Significant differences were also noted between the recall and denture groups ($P = 0.03$) and the endodontic group and denture group ($P = 0.03$) for the item assessing being uncomfortable with the appearance of the teeth or dentures.

Magnitude of Change in Quality of Life Scores Following Dental Intervention. The effect sizes describe the magnitude of change, and these effect sizes varied by group. Cohen (23) defined effect sizes as small = 0.2, moderate = 0.5, and large = 0.8. Using Cohen's criteria, effect sizes were characterized as minimal to small in the recall and endodontic groups, and borderline moderate in the denture group. There were significant differences between the denture and recall groups ($P = 0.03$) for the subscale regarding social functioning. There were also significant differences between the denture and recall groups ($P = 0.03$) for the items assessing feeling depressed and feeling uncomfortable about the appearance of the teeth, mouth, or dentures (Table 5).

Discussion

We examined whether a newly developed brief measure of OQOL is sensitive to changes in oral health status as a consequence of dental interventions. We hypothesized that patients receiving a removable prosthesis or endodontic care would show greater improvement in OQOL over a 3-month period than patients receiving only a prophylaxis. Our overall findings were that patients for removable prosthesis showed the greatest improvement in OQOL following dental treatment. The greatest

Table 3
Baseline Scale Scores and Items by Group

In the past 3 months, how often have you experienced the following difficulties because of problems with your teeth, mouth, or dentures?	RECALL§ ($n = 32$)	ENDO• ($n = 15$)	DENTURE∞ ($n = 16$)
Summary scales		Mean (SD)	
OH-1	55 (27.4)	58 (32.3)	53 (32.7)
6-item scale	42 (9.6)	41 (12.6)	45 (13.5)
12-item scale	32 (14.3)	39 (17.1)	34 (19.4)
Subscales			
Distress	26 (24.8)	45 (35.9)	26 (33.8)
Worry	48 (15.5)	57 (18.4)	47 (21.9)
Social functioning	10 (17.5)	16 (23.9)	12 (22.8)
Individual items			
Have you been upset?†	32 (29.9)	55 (45.5)	35 (45.1)
Have you found it difficult to relax?*†	27 (28.6)	45 (36.8)	25 (36.5)
Have you felt depressed?†	17 (24.9)	35 (38.7)	14 (30.2)
Do you feel nervous or self-conscious?*‡	69 (32.9)	63 (29.6)	80 (31.6)
Have you been worried about dental problems?‡	44 (30.2)[AB]	60 (38.7)[A]	28 (38.8)[B]
Have you felt uncomfortable about the appearance of your teeth?	32 (36.0)	47 (39.9)	34 (39.6)
Have you avoided going out?*¶	15 (25.2)	23 (33.3)	14 (27.3)
Have you been totally unable to function?¶	4 (11.1)	10 (18.4)	6 (25)
Have you had trouble getting along with others?¶	12 (22.8)	13 (24.7)	12 (18.5)
How much pain or distress do you have?*	24 (26.5)[A]	63 (35.1)[B]	22 (27.5)[A]
Have you had uncomfortable dentures?*	38 (20.9)	25 (43.3)	35 (28.0)
Have you had to avoid eating some foods?*	27 (33.2)	36 (42.4)	41 (35.2)

Values with same superscripts are not significantly different ($P > 0.05$ using Duncan's test).
* Items on 6-item scale.
Subscales:
† Distress.
‡ Worry.
¶ Social functioning.
§ Regular users of diagnostic and preventive care and presented for a prophylaxis.
• Presented to the dental school setting in pain; to undergo emergency endodontic treatment.
∞ Will receive a removable prosthesis.
OH-1, self-report of oral health; SD, standard deviation.

mean change in scores was observed among the denture group for social functioning, feeling depressed, and appearance. Subjects in both the recall and endodontic groups reported improvements which were only minimal to small.

Spilker (11) described health status measures as being discriminative, predictive, and evaluative. Discriminative instruments are used to measure differences between subjects at a point in time when no "gold standard" is available, and these differences can be interpreted as trivial, small, moderate, or large. The aim of predictive instruments is to classify individuals relative to a predefined "gold standard." Evaluative instruments are used to measure longitudinal change within, and between, samples. A major goal of using evaluative instruments is to better understand how dental conditions and subsequent interventions impact quality of life, and use of evaluative instruments is essential to planning health care at the individual and societal level (17).

A major property of an evaluative instrument is its sensitivity to change over time. Locker (24) describes four methods currently used to measure change. The first method is to compare baseline and follow-up measurements. Although this method is simple, it masks within subject change so that positive and negative changes cancel each other out. The second approach is by the use of change scores; obtained by subtracting post-intervention scores from pre-intervention scores. The third approach involves the use of global transition scores reflecting the patients' overall assessment of how their oral health has changed over the time period in question. The final approach is the use of global transition scales derived from a series of global transition statements applied to different dimensions of health. None of these methods is universally accepted. Our approach involved the use of change scores and standardized effect sizes to assess the magnitude of change. In this method the mean change is divided by the SD of the baseline score. Thus, the magnitude of change of individual items of the 6- and 12-item scales in response to dental intervention in this sample could be characterized as minimal to small in the recall and endodontic groups and borderline moderate in the denture group.

Table 4
Adjusted Follow-Up Item and Scale Means by Group (Adjusted for Age, Gender, and Baseline Score)

In the past 3 months, how often have you experienced the following difficulties because of problems with your teeth, mouth, or dentures?	RECALL§ ($n = 22$)	ENDO• ($n = 12$)	DENTURE∞ ($n = 10$)
Summary scales		Mean (SD)	
OH-1	57 (26.8)	56 (30.3)	53 (32.0)
6-item scale	41 (18.7)	44 (17.0)	43 (12.5)
12-item scale	32 (14.6)	34 (18.5)	34 (17.0)
Subscales			
Distress	27 (22.0)	31 (35.4)	26 (18.3)
Worry	48 (15.6)	50 (18.8)	47 (18.0)
Social functioning	12 (17.0)	14 (26.2)	18 (25.3)
Individual items			
Have you been upset?†	30^{AB} (24.5)	40^{B} (32.3)	20^{A} (30.2)
Have you found it difficult to relax?*†	32 (30.0)	40 (35.3)	30 (32.1)
Have you felt depressed?†	17^{A} (23.4)	21^{AB} (38.2)	27^{B} (24.2)
Do you feel nervous or self-conscious?*‡	72 (29.5)	64 (16.5)	84 (25.0)
Have you been worried about dental problems?‡	40 (35.1)	47 (24.2)	45 (40.4)
Have you felt uncomfortable about the appearance of your teeth?‡	34^{A} (28.2)	38^{A} (37.3)	21^{B} (17.5)
Have you avoided going out?*¶	17 (22.3)	24 (32.8)	20 (22.3)
Have you been totally unable to function?¶	7 (12.2)	4 (9.6)	15 (19.3)
Have you had trouble getting along with others?¶	11 (19.3)	14 (22.3)	19 (28.3)
How much pain or distress do you have?*	64 (31.6)	61 (37.6)	57 (27.3)
Have you had uncomfortable dentures?*	39 (32.2)	29 (43.3)	38 (35.2)
Have you had to avoid eating some foods?*	24 (30.1)	37 (44.4)	36 (43.2)

Values with same superscripts are not significantly different ($P > 0.05$ using Duncan's test).
* Items on 6-item scale.
Subscales:
† Distress.
‡ Worry.
¶ Social functioning.
§ Regular users of diagnostic and preventive care and presented for a prophylaxis.
• Presented to the dental school setting in pain; to undergo emergency endodontic treatment.
∞ Will receive a removable prosthesis.
OH-1, self-report of oral health; SD, standard deviation.

This new brief 6- and 12-item instrument is a validated questionnaire, is responsive to differences in clinical status, and has been used in previous studies (16,25). In a study to examine the effects of tooth loss and denture-wearing on quality of life, Jones (25) found that the 6- and 12-item scales differentiated between dentition/denture groupings and that the item assessing avoidance of certain foods discriminated well between dentition groups. A limitation of that study was that it was conducted exclusively in male veterans.

An important limitation of this present study is the small sample size. However, one of the strengths of this study is the diverse sample in terms of age, gender, and ethnic background, and the fact that the same patients were analyzed before and after treatment. The response rate to this study of 70 percent was acceptable and indicated the feasibility of employing a short-form self-completed outcome measure in a dental school setting. Although the age composition, gender, and ethnic representation of subjects who did and did not remain in the study were virtually the same, potential bias because of loss of some participants in the study must be considered when interpreting the findings.

These results were also limited by floor effects (indicates best possible scores) and ceiling effects (indicates worst possible scores); hence, the results of magnitude of changes (effect sizes) following treatment need to be interpreted with caution as changes cannot be reliably estimated for individuals with extreme scores.

A final consideration is the limitations in the use of regression analyses in the development of short-form measures of OQOL measures as underlying assumptions of regression analyses are violated by these types of data. Locker and Allen (26) argue that the method of developing a short-form instrument is not as important as its content and that the items in the questionnaire and its measurement properties need to be appropriate to its purpose, the population to which it is applied, and the context in which it is being used.

The results from this study further support the use of these scales as a brief measure of OQOL in dental school clinical settings. Further research using this new instrument is needed in larger samples and different settings.

Table 5
Effect Size of Scale and Item Means by Group (Change as % Baseline SD)

In the past 3 months, how often have you experienced the following difficulties because of problems with your teeth, mouth, or dentures?	RECALL§ ($n = 22$)	ENDO• ($n = 12$)	DENTURE∞ ($n = 10$)
Summary scales			
OH-1	−3.6 (53.9)	−7.6 (−7.7)	−8.4 (26.7)
6-item scale	−12.7 (66.1)	15.9 (68.9)	10.9 (82.9)
12-item scale	−11.1 (23.7)	4.5 (20.0)	6.3 (71.7)
Subscales			
Distress	−10.6 (24.1)	0 (0)	−2.7 (50.2)
Worry	3 (64.4)	−4.1 (13.9)	−13.8 (92.1)
Social functioning	−1.8 (28.9)	11.1^{AB} (37.0)	36.8^{A} (75.8)
Individual items			
Have you been upset?†	−25.3 (50.7)	−11.7 (39.1)	−38.8 (81.9)
Have you found it difficult to relax?*†	6.5 (31.4)	13.6 (45.4)	15.1 (59.4)
Have you felt depressed?†	−7.1 (34.1)	0^{AB} (0)	24.4^{A} (55.1)
Do you feel nervous or self-conscious?*‡	7.1 (90.1)	−14.2 (47.2)	−17.3 (52.1)
Have you been worried about dental problems?‡	−6.3 (36.7)	6.3 (21.1)	20.9 (57.4)
Have you felt uncomfortable about the appearance of your teeth or dentures?‡	5.7^{B} (34.1)	0^{AB} (0)	26.4 (63.4)
Have you avoided going out?*¶	−3.9 (42.9)	24.6 (81.7)	27.1 (43.6)
Have you been totally unable to function?¶	12.6 (41.8)	0 (0)	58.1 (140.2)
Have you had trouble getting along with others?¶	−9.8 (47.3)	0 (0)	22.6 (71.7)
How much pain or distress do you have?*	−19.6 (72.1)	−6.8 (22.7)	−45.1 (95.1)
Have you had uncomfortable dentures?*	22.8 (45.7)	0 (0)	34.3 (137.7)
Have you had to avoid eating some foods?*	−12.1 (68.7)	20.9 (66.2)	20.9 (114.1)

Values with same superscripts are not significantly different ($P > 0.05$ using Duncan's test).
* Items on 6-item scale.
Subscales:
† Distress.
‡ Worry.
¶ Social functioning.
§ Regular users of diagnostic and preventive care and presented for a prophylaxis.
• Presented to the dental school setting in pain; to undergo emergency endodontic treatment.
∞ Will receive a removable prosthesis.
OH-1, self-report of oral health; SD, standard deviation.

References

1. Locker D. Concepts of oral health, disease and quality of life. In: Slade GD, editor. Measuring oral health and quality of life. Chapel Hill, NC: University of North Carolina, Dental Ecology; 1997. p. 12-23.
2. Slade GD. Measuring oral health related quality of life. Chapel Hill, NC: University of North Carolina, Dental Ecology; 1997.
3. Jones JA, Kressin NR, Spiro A III, Randall CW, Miller DR, Hayes C, Kazis L, Garcia RI. Self-reported and clinical oral health in users of VA health care. J Gerontol: Med Sci. 2001;56A:M55-62.
4. Jones JA, Spiro A, Miller DR, Garcia RI, Kressin NR. Need for dental care in older veterans: assessment of patient-based measures. J Am Geriatr Soc. 2002;50: 163-8.
5. Heydecke G, Locker D, Awad MA, Lund JP, Feine JS. Oral and general health-related quality of life with conventional and implant dentures. Community Dent Oral Epidemiol. 2003;31(3):161-8.
6. Allen PF, McMillan AS, Locker D. An assessment of sensitivity to change of the Oral Health Impact Profile in a clinical trial. Community Dent Oral Epidemiol. 2001;29(3):175-82.
7. Scott BJ, Forgie AH, Davis DM. A study to compare the oral health impact profile and satisfaction before and after having replacement complete dentures constructed by either the copy or the conventional technique. Gerodontology. 2006;23:79-86.
8. Allen PF, Thomason JM, Jepson NJ, Nohl F, Smith DG, Ellis J. A randomized controlled trial of implant retained mandibular overdentures. J Dent Res. 2006;83(6): 547-51.
9. John MT, Reissmann DR, Schierz O, Wassell RW. Oral health-related quality of life in patients with temporomandibular disorders. J Orofac Pain. 2007;21(1):46-54.
10. Locker D, Jokovic A, Clarke M. Assessing the responsiveness of measures of oral health-related quality of life. Community Dent Oral Epidemiol. 2004;32(1):10-8.
11. Spilker B. Quality of life and pharmacoeconomics in clinical trials. 2nd ed. Philadelphia, PA: Lippincott-Raven Publisher; 1996.
12. Allen PF. Assessment of oral health related quality of life. Health Qual Life Outcomes. 2003;1:40-8.
13. Kressin NR, Spiro A III, Bosse R, Garcia R, Kazis L. Assessing oral health-related quality of life: findings from the Normative Aging Study. Med Care. 1996; 34(5):416-2.
14. Atchison KA, Dolan TA. Development of the Geriatric Oral Health Assessment Index. J Dent Educ. 1990;54(11):680-7.
15. Slade GD, Spencer AJ. Development and evaluation of the Oral Health Impact Profile. Community Dent Health. 1994; 11(1):3-11.
16. Kressin NR, Jones JA, Orner MB, Spiro A III. A new brief measure of Oral Quality of Life. Prev Chronic Dis [serial on the Internet]. 2008 [cited 2008 Jun 18];5(2). Available from: http://www.cdc.gov/pcd/issues/2008/apr/06_0147.htm.
17. Patrick DL, Erickson P. Health status and health policy. Quality of life in health care evaluation and resource allocation. New York: Oxford University Press; 1993.
18. Gift HC, Atchison KA. Oral health, health, and health-related quality of life. Med Care. 1995;33(11 Suppl):NS57-77.
19. Gift HC, Atchison KA, Dayton CM. Conceptualizing oral health and oral health-related quality of life. Soc Sci Med. 1997; 44(5):601-8.
20. Hays RD, Hayashi T. Beyond internal consistency reliability: rational and user's guide for the Multitrait Analysis Program

(MAP) on the minicomputer. Behav Res Methods Instrum Comput. 1990;22:167-75.
21. Jones JA, Kressin NR, Spiro A III, Miller D, Kazis L, Garcia R. Comparison of patient-based oral health outcome measures. Qual Life Res. 2004;13(5):975-85.
22. Robinson PG, Gibson B, Khan F, Birnbaum W. A comparison of OHIP 14 and OIDP as interviews and questionnaires. Community Dent Health. 2001;18(3): 144-9.
23. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
24. Locker D. Issues in measuring change in self-perceived oral health status. Community Dent Oral Epidemiol. 1998;267:41-7.
25. Jones JA, Orner MB, Spiro A, Kressin NR. Tooth loss and dentures: patients' perspectives. Int Dent J. 2003;53:327-34.
26. Locker D, Allen PF. Developing short-form measures of oral health-related quality of life. J Public Health Dent. 2002;62(1):13-20.