General Internal Medicine
Boston University School of Medicine
2007 Publications

1. Aggarwal A, Speckman J, **Paasche-Orlow MK**, Roloff KS, **Battaglia T**. The role of numeracy on cancer screening among urban women. Am J Hlth Behav. Sept-Oct 2007;31(S1):57-68.
2. **Alford DP**, **LaBelle C**, **Richardson JM**, **O'Connell JJ**, Hohl CA, **Cheng DM**, **Samet JH**. Treating homeless opioid dependent patients with buprenorphine in an office-based setting. J Gen Intern Med. 2007; 22:171-76.
3. **Alford DP**, Salsitz E, Martin J, Renner JA. Clinical case discussion: Treating opioid dependence with buprenorphine. J Addict Med. 2007;1:73-78
4. Arkes HR, Shaffer VA, **Medow MA**. Patients derogate physicians who use a computer-assisted diagnostic aid. Med Decis Making. Mar-Apr 2007;27(2):189-202.
5. **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. Jan 2007;109(S2):359-67.
6. Boehmer U, Linde R, **Freund KM**. Breast reconstruction following mastectomy for breast cancer: the decisions of sexual minority women. Plast Reconstr Surg. Feb 2007;119(2):464-72.
7. Campos NG, Salomon JA, Servoss JC, Nunes DP, **Samet JH**, Freedberg KA, Goldie SJ. Cost-effectiveness of treatment for Hepatitis C in an urban cohort co-infected with HIV. Am J Med. 2007;120:272-79.
8. **Chen D**, Lew R, **Hershman W**, **Orlander J**. A cross-sectional measurement of medical student empathy. J of Genl Intern Med. 2007;22(10):1434-38.
9. **Cheng DM**, Nunes D, Libman H, Vidaver J, **Alperen JK**, **Saitz R**, **Samet JH**. Impact of Hepatitis C on HIV progression in adults with alcohol problems. Alcohol Clin Exp Res 2007; 31(5):829-36.
10. Chuang CH**, Liebschutz JM**, **Cheng DM**, **Raj A**, **Samet JH**. Substance use during sexual and physical assault in HIV-infected persons. Violence and Victims. 2007;22(2):216-25.
11. Claudio C, Katz RV, Green BL, **Kressin NR**, Wang MQ, Russell SL. (2007). Cancer-screening participation: Comparative willingness of San Juan Puerto Ricans vs. New York City Puerto Ricans. Journal of the National Medical Association. 2007;99(5):542-49.
12. Craft LL, **Freund KM**, Culpepper L, **Perna FM**. Intervention study of exercise for depressive symptoms in women. J Womens Health (Larchmt). 2007 Dec;16(10):1499-509.
13. Cunha-Cruz J, Hujoel PP, **Kressin NR**. Oral health impacts the quality of life of periodontal patients. Journal of Periodontal Research. Apr 2007;42(2):169-76.
14. Cupples LA, Arruda HT, Benjamin EJ, D'Agostino Sr BD, Demissie S, DeStefano AL, Dupuis J, Falls KM, Fox CS, Gottlieb DJ, Govindaraju DR, Guo CY, Heard-Costa NL, Hwang J, Kathiresan S, Kiel DP, Laramie JM, Larson MG, Levy D, Liu Cy, Lunetta KL, Mailman MD, Manning AK, Meigs JB, **Murabito JM**, Newton-Cheh C, O'Connor GT, O'Donnell CJ, Pandey M, Seshadri S, Vasan RS, Wang ZY, Wilk JB, Wolf PA, Yang Q, Atwood LD. The Framingham heart study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. BMC Medical Genetics. 2007;8(S1):1-19.

15. Daeppen JB, **Bertholet N**, Gmel G, Gaume J. Communication during brief intervention, intention to change, and outcome. Sub Abuse. 2007;28(3):43-51.

16. Evans-Molina C, Regan S, **Henault LE**, **Hylek EM**, Schwartz GR. The new Medicare Part D prescription drug benefit: An estimation of its effect on prescription drug costs in a Medicare population with atrial fibrillation. Journal of the American Geriatrics Society. 2007;55(7):1038-1043.

17. Fang MC, Go AS, Chang Y, **Hylek EM**, **Henault LE**, Jensvold NG, Singer DE. Death and disability from warfarin-associated intracranial and extracranial hemorrhages. Am J Med 2007;120:700-5.

18. **Farzanfar R**, **Stevens A**, Vachon L, **Friedman R**, Locke SE. Design and Development of a Mental Health Assessment and Intervention System. Journal of Medical Systems. J of Medical Systems. 2007;31;49-62.

19. **Farzanfar R**, **Frishkopf S**, **Friedman R**, **Ludena K**. Evaluating an automated mental health care system: making meaning of human-computer interaction. Computers in Human Behavior. 2007;23:1167-82.

20. Finucane MM, **Samet JH**, **Horton NJ**. Translational methods in biostatistics: linear mixed effect regression models of alcohol consumption and HIV disease progression over time. Epidemiol Perspect Innov. Sept 2007;19:4(1):8.

21. 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(2):193-7.

22. Garcia D, **Hylek EM**. Stroke prevention in elderly patients with atrial fibrillation. Lancet. 2007;370:460-61.

23. Garcia DA, **Hylek EM**. A risk model predicted major bleeding in older patients with atrial fibrillation receiving warfarin therapy. [Invited Commentary]. ACP J Club. Mar-Apr 2007;146(2):52.

24. Gordon AJ, Sullivan LE, **Alford DP**, Arnsten JH, Gourevitch MN, Kertesz SG, Kunins HV, Merrill JO, **Samet JH**, Fiellin DA. Update in addiction medicine for the generalist. J Gen Intern Med. 2007; 22:1190-94

25. Graham CS, Wells A, Edwards EM, Heeren T, Tumilty S, Stuver SO, **Samet JH**, Nunes D, Horsburgh CR, Koziel MJ. Effect of exposure to injection drugs or alcohol on antigen-specific immune responses in HIV and Hepatitis C coinfection. J Infect Diseases. 2007; 195:847-56.

26. Hart RG, Albers GW, Connolly SJ, Friday G, Gage BF, Go AS, Goldstein LB, Gronseth G, Halperin JL, **Hylek EM**, Lip GYH, Pearce LA, Sherman DG, Singer DE, van Walraven C. Independent predictors of stroke in patients with atrial fibrillation. Neurology. 2007;69:546-54.

27. **Horton NJ**, Kim E, **Saitz R**. A cautionary note regarding count models of alcohol consumption in randomized controlled trials. BMC Medical Research Methodology. 2007;7:1-9.

28. **Hylek EM**. Drug Evaluation: DU-176b, an oral direct factor Xa antagonist. Current Opinion in Investigational Drugs: cardiovascular and Renal Issue. 2007;8:778-83.

29. **Hylek EM**. Does Preadmission anticoagulation therapy reduce stroke fibrillation? [Invited Commentary]. Nature Clinical Practice Cardiovascular Medicine. 2007;4:186-87.

30. **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. May 2007;115(21):2689-96.

31. **Kapoor A**, Page S, LaValley M, Gale DR, Felson DT. Magnetic resonance imaging for diagnosing foot osteomyelitis. Arch Intern Med. 2007;167:125-32.

32. **Kartha A**, Anthony D, Manasseh CS, **Greenwald JL**, Chetty VK, Burgess JF, Culpepper L, Jack BW. Depression is a risk factor for re-hospitalization in medical patients. Prim Care Companion J Clin Psychiatry. 2007;9:256-62.

33. Katz RV, Green BL, **Kressin NR**, Claudio C, Wang MQ, Russell SL. (2007). Willingness of Minorities to Participate in Biomedical Studies: Confirmatory findings from a follow-up study using the Tuskegee Legacy Project Questionnaire. Journal of the National Medical Association: 99(9):1052-60.

34. **Kim TW**, Palepu A, **Cheng DM**, Libman H, **Saitz R**, **Samet JH**. Factors associated with discontinuation of antiretroviral therapy in HIV-infected patients with alcohol problems. AIDS Care. 2007;19(8):1039-47.

35. **Kim TW**, **Samet JH**, **Cheng DM**, Winter MR, Safran DG, **Saitz R**. Primary care quality and addiction severity: a prospective cohort study. Health Services Research. 2007;42(2):755-72.

36. **Kressin NR**, Glickman ME, Peterson ED, Whittle J, Orner MB, Petersen LA. Functional status outcomes among white and African American cardiac patients in an equal access system. Am Heart J. 2007;153:418-25.

37. **Kressin NR**, Saha S, Weaver F, Rubenstein L, Weinberger M. Career and time management strategies for clinical and health services researchers. JGIM. 2007;22(10):1475-78.

38. **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. JGIM. 2007;22(6):768-74.

39. Larson MJ**, Paasche-Orlow M**, **Cheng DM**, **Lloyd-Travaglini C**, **Saitz R**, **Samet JH**. Persistent pain is associated with substance use after detoxification: a prospective cohort analysis. Addiction. 2007;102(5):752-60.

40. **Liebschutz JM**, **Saitz R**, Brower V, Keane TM, **Lloyd-Travaglini C**, **Averbuch T**, **Samet JH**. PTSD in urban primary care: High prevalence and low physician recognition. JGIM. 2007;22(6):719-26.

41. Lunetta KL, D'Agostino RB, Karasik D, Benjamin EJ, Guo CY, Govindaraju R, Kiel DP, Kelly-Hayes M, Massaro JM, Pencina MJ, Seshadri S, **Murabito JM**. Genetic correlates of longevity and selected age-related Phenotypes: a genome-wide association study in the Framingham Study. BMC Medical Genetics. 2007;8(S1):13.

42. McAlearney AS, Chisolm DJ, Schweikhart S, **Medow MA**, Kelleher K. The story behind the story: physician skepticism about relying on clinical information technologies to reduce medical errors. Int J Med Inform. Nov-Dec 2007;76(11-12):836-42.

43. McKinlay JB, Link CL, **Freund KM**, Marceau LD, O'Donnell AB, Lutfey KL. Sources of variation in physician adherence with clinical guidelines: results from a factorial experiment. J Gen Intern Med. Mar 2007;22(3):289-96.

44. Munguia Tapia E, Intille SS, Haskell W, Larson K, **Wright J**, King A, **Friedman R**. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. Proceedings of the International Symposium on Wearable Computers 2007. (pp. 1-4):IEEE Press.

45. **Murabito JM**, Rosenberg CL, Finger D, **Kreger BE**, Levy D, Splansky GL, Antman K, Hwang SJ. A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. BMC Medical Genetics. 2007;8(S1):6.

46. Osborn CY, **Paasche-Orlow MK**, Davis TC, Wolf MS. Health literacy: an overlooked factor in understanding HIV health disparities. Am J Prev Med. Nov 2007;33(5):374-8.

47. **Paasche-Orlow MK**, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav. Sep-Oct 2007;31(S1):19-26.

48. Palfai TP, **Cheng DM**, **Samet JH**, Kraemer KL, Roberts MS, **Saitz R**. Depressive symptoms and subsequent alcohol use & problems: a prospective study of medical inpatients with unhealthy alcohol use. J Stud Alcohol Drugs. 2007;68:673-80.

49. **Raj A**, **Saitz R**, **Cheng DM**, Winter M, **Samet JH**. Associations between alcohol, heroin and cocaine use and high risk sexual behaviors among detoxification patients. Am J Drug Alcohol Abuse. 2007;33:169-78.

50. Rawaf M, **Kressin NR**. Exploring racial and sociodemographic trends in physician behavior, physician trust and their association with blood pressure control. J of Nat Med Assc. 2007;99 (11):1248-54.

51. **Rose AJ**, Backus BM, Gershman ST, Santos P, **Ash AS**, **Battaglia TA**. Predictors of aggressive therapy for nonmetastatic prostate carcinoma in Massachusetts from 1998 to 2002. Med Care. May 2007;45(5):440-7.

52. **Rose AJ**, **Berlowitz DR**, Orner, MB, **Kressin NR**. Understanding Uncontrolled Hypertension: Is it the Patient or the Provider? J of Clinical Hypertension. 2007;9(12):937-43.

53. **Rose AJ**, Sharman J, Ozonoff A, **Henault L**, **Hylek E**. Effectiveness of Warfarin among Patients with Cancer. JGIM. 2007;22(7):997-1002.

54. **Saitz R**. Treatment of alcohol and other drug dependence. Liver Transpl 2007;13:S59-S64.

55. **Saitz R**, **Gaeta J**, **Cheng DM**, **Richardson JM**, Larson MJ, **Samet JH**. Risk of mortality during four years after substance detoxification in urban adults. J Urban Health: Bull NY Acad Sci. 2007;84(2):272–82.

56. **Saitz R**, Palfai TP, **Cheng DM**, **Horton NJ**, **Freedner N**, Dukes K, Kraemer KL, Roberts MS, **Guerriero RT**, **Samet JH**. Brief intervention for medical inpatients with unhealthy alcohol use: A randomized, controlled trial. Ann Intern Med. 2007;146:167-76.

57. **Saitz R**, Palfai TP, **Freedner N**, Winter MR, MacDonald A, Lu J, Ozonoff A, Rosenbloom DL, DeJong W. Screening and brief intervention online for college students: The ihealth study. Alcohol. 2007;42:28-36.

58. **Samet JH**, **Cheng DM**, Libman H, Nunes DP, **Alperen J**, **Saitz R**. Alcohol consumption and HIV disease progression. JAIDS. 2007;46(2):194-9.

59. **Samet JH**, **Walley AY**, **Bridden C**. Illicit drugs, alcohol, and addiction in human immunodeficiency virus. Panminerva Med. 2007;49:67-77.

60. **Sherman BJ**, Gilliland G, Speckman JL, **Freund KM**. The effect of a primary care exercise intervention for rural women. Preventive Medicine. 2007;44:198-201.

61. **Takizawa C**, **Cheng D**, **Samet J**, Winter M, Larson MJ, **Saitz R**. Primary medical care and reductions in HIV risk behaviors in adults with addictions. J Addictive Diseases. 2007;26(3):17-25.

62. **Tsui JI**, **Saitz R**, **Cheng DM**, Nunes D, Libman H, **Alperen JK**, **Samet JH**. Awareness of hepatitis C diagnosis is associated with less alcohol use among persons co-infected with HIV. J Gen Intern Med. 2007;22:822-5.

63. Volandes AE, **Paasche-Orlow MK.** Health literacy, health inequality and a just healthcare system. Am J Bioeth. Nov 2007;7(11):5-10.

64. 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;31:428-35.

65. Wines JD, **Saitz R**, **Horton NJ**, **Lloyd-Travaglini C**, **Samet JH**. Overdose after detoxification: a prospective study. Drug and Alcohol Dependence. 2007;89:161-9.

The Role of Numeracy on Cancer Screening Among Urban Women

Arpita Aggarwal, MD, MSc; Jeanne L. Speckman, MSc
Michael K. Paasche-Orlow, MD, MA, MPH; Kathryn S. Roloff, MA
Tracy A. Battaglia, MD, MPH

***Objectives*: To determine the role of numeracy on cancer screening knowledge and practices among an ethnically diverse population. *Methods*: Women aged 40+ presenting for primary care at an urban academic medical center were surveyed. Numeracy was measured as a dichotomous outcome (numerate: yes/no) using 3 criteria adapted from a validated instrument. Self-report was used to determine if women were up-to-date with breast and colorectal cancer screening. *Results*: Adequate numeracy was associated with increasing age, white race, higher education, and knowledge of breast cancer screening guidelines. No association was found between numeracy and cancer screening practices. *Conclusions*: Adequate numeracy was not a key determinant of cancer screening.**

***Key words*: numeracy, literacy, breast cancer, colorectal cancer, and screening.**

***Am J Health Behav.* 2007;31(Suppl 1):S57-S68**

Early detection of breast and colorectal cancer, the second and third leading causes of cancer death among US women respectively, can save lives, reduce length of treatment, and increase quality of life.[1] In the year 2000, 70% of women 40 and older had undergone a screening mammography as recommended, whereas only 37% of women over 50 had undergone a recommended colorectal screening.[2-4] Despite an overall reduction in mortality from these cancers due to improved treatment options and use of screening, certain women continue to suffer from poor outcomes.[5]

In particular, African American women tend to present with more advanced stage of disease and have higher mortality rates.[6-8] Furthermore, lower socioeconomic status along with specific cultural beliefs and attitudes have been associated with late-stage presentation for both breast and colorectal cancer, and disparities in cancer screening rates may explain some of these disparate outcomes.[9-17,6-11] According to the National Health Interview Survey, mammography screening rates in women over 40 years of age have improved among both whites and blacks.[12] However, black women continue to dem-

Arpita Aggarwal, Assistant Professor of Medicine, Virginia Commonwealth University, Richmond, VA. Jeanne L. Speckman, Research Data Analyst, Health Care Research Unit; Michael K. Paasche-Orlow, Assistant Professor of Medicine; Kathryn S. Roloff, Research Associate, Women's Health Research Unit; Tracy A. Battaglia, Assistant Professor of Medicine, Women's Health Research Unit, Boston University Medical Center, Boston, MA.

Address correspondence to Dr Aggarwal, Virginia Commonwealth University, 1200 East Marshall Street, PO Box 980102, Richmond, VA 23298-0102. E-mail: aaggarwal@mcvh-vcu.edu

onstrate lower rates than compared to white women. This racial disparity is even more pronounced for colorectal cancer: black women are 18% less likely to receive colonoscopy and 39% less likely to receive flexible sigmoidoscopy after controlling for age, sex, income, and access to care.[8]

Meanwhile, a systematic review of US studies on the prevalence of limited health literacy in health care settings revealed that 26% (95% CI 22% to 29%) of American adults had low health literacy and an additional 20% (95% CI 16% to 23%) had marginal health literacy.[13] The prevalence of limited health literacy among American adults presents an important challenge for clinicians, who are expected to facilitate discussion of health risk information especially in preventive care and cancer screening where risk communication may be presented with a series of numbers, rates, and percentages.[14] Low literacy has been found to correlate with less health knowledge, worse self-management skills, less use of preventative tactics, higher rates of hospitalization, and death.[15,16] Individuals with low literacy may find themselves unable to read prescription labels; health education materials; test results, appointment reminders; and other important, but basic, documents.[13,17]

Current organizational guidelines recommend personalized discussion of health risk information as part of the standard of care to help patients make decisions about cancer screening.[3,18] Most of this information is presented to the patient in the form of data collected as percentages and probabilities. The National Center for Education Statistics (NCES) defines *numeracy* as a quantitative domain of literacy. *Numeracy* or *quantitative literacy* is defined as the knowledge and skills needed to understand the fundamental notions of numbers and chance.[19] Numeracy includes the ability to perform calculations and to decipher numbers embedded in text, as well as the ability to handle numbers when writing or filling out forms.

In 2003, the National Adult Health Literacy Survey (NAALS), a nationally representative household survey administered by the National Center for Education Statistics, showed that 22% of all American adults surveyed exhibited the lowest level of numeracy.[19] The prevalence of below-basic numeracy skills is associated with race/ethnicity: 13% of whites, 47% of blacks, and 50% of Hispanics, exhibit below-basic numeracy skills.[19] The disproportionate representation of low numeracy among American adults and the possibility that numeracy may influence cancer screening suggest a potential explanation of racial and ethnic disparities seen in cancer screening.

Though prior research has found numeracy to be a strong predictor of accurately gauging the benefits of mammography, to our knowledge there is no literature addressing the role of numeracy on knowledge of cancer screening guidelines and screening practices.[20] Therefore, we surveyed a racially and ethnically diverse population of urban women to investigate whether patients' numeracy status affects knowing about recommended cancer screening guidelines and being up-to-date with these screening practices.

METHODS

Data Collection

We utilized data from a self-administered cross-sectional survey developed to study racial/ethnic differences in cancer risk perception and screening behavior among urban women. Consecutive women presenting for primary care at 4 ambulatory care sites of an urban academic medical center between August 2004 and July 2005 were eligible. The recruitment sites included 2 hospital-based practices and 2 community health centers. After receiving institutional review board approval, English-speaking research assistants approached eligible women as they presented for scheduled visits and invited them to participate. The survey was offered in English, Spanish or Haitian Creole. After giving written informed consent, participants completed an 85-item written survey in the exam room with a research assistant available to answer participant questions. Approximately 15 minutes were needed to complete the survey, and no incentive was provided.

Eligibility Criteria

Women were eligible for participation if they were 40 years or older and had the ability to speak and read English, Spanish, or Haitian Creole. Women were excluded if they had a personal history of any non-melanoma cancer or had a cog-

nitive impairment that affected mental status.

Study Measures

Numeracy was assessed with 3 criteria adapted from Black and colleagues and scored as a dichotomous outcome (numerate: yes/no).[21] Subjects were considered numerate if they met 3 criteria: (1) basic familiarity with probability, (2) comfort with using probability, and (3) basic familiarity with proportions. The first criterion was assessed by a response to the following question: "Imagine that you flip a coin 1000 times. Out of 1000 flips, how many times do you think the coin comes up heads?" The correct response to this question was 500 times. The second criterion was assessed by posing 3 quantitative risk questions that asked the participant about her perception of her chances of developing breast and colorectal cancer (_X_ in 1000). The correct response was rated as any numerical response to all 3 questions. The third criterion was assessed by comparing the participant's estimates of her lifetime and 5-year risk of getting cancer. A correct response occurred when the participant accurately estimated her lifetime risk to be greater than her 5-year risk. Though the specific questions used for criteria 2 and 3 differed from that used by Black, the concepts were the same. Nonresponse to any of these questions was considered to be an incorrect response. We also analyzed each numeracy question to evaluate the contribution of each criterion on the overall numeracy classification.

Demographic information was self-reported and included age, race/ethnicity, medical insurance, level of education, and income. To determine race and ethnicity, subjects were first asked if they considered themselves to be Hispanic or Latina and then given the following choices for race based on the US census criteria: American Indian or Alaskan Native, Asian, black or African American, Native Hawaiian or Pacific Islander, white, or Other. These questions were paired with a place-of-birth question which created 2 subgroups of non-Hispanic black participants: African American or Caribbean Black. For our analyses racial/ethnic categories were African American black, Caribbean black, Hispanic, white, and Other. Based on a series of insurance questions, we categorized subjects as having either private insurance or nonprivate insurance (which included public, Medicare, none, or other). Participants were asked to report on their family history of breast and colorectal cancer and also whether or not they had a primary care provider (PCP).

Dichotomous outcomes of interest in the study were *knowing about* breast and colorectal cancer screening guidelines and being *up-to-date* with breast and colorectal cancer screening. Subjects were categorized as knowledgeable of cancer screening guidelines if they correctly answered questions about the recommended age to start screening: for an average-risk woman, correct categorical responses were 40-49 years for breast cancer and 50-59 years for colorectal cancer.[22,23] Subjects were considered up-to-date for breast cancer screening if they had a routine mammogram within the last 2 years. For those age 50 years and older, up-to-date for colorectal cancer screening included having a fecal occult blood test in the past year or ever having lower endoscopy (flexible sigmoidoscopy or colonoscopy). As most subjects were between age 50 and 60, ever having had an endoscopy was considered up-to-date with screening.

Statistical Analysis

We used chi-square and Fisher-exact tests for bivariate comparisons. Logistic regression was used for multivariate analyses of the outcomes: numeracy, screening knowledge, and up-to-date with screening. Potential confounders investigated in these models included age, race, education, insurance, income, and site of care. The models of knowledge and screening behavior also investigated the association with having a family history of the disease and having a primary care provider (PCP).

Because income, insurance, and education were collinear, we included only education among these 3 variables in the models of numeracy and knowledge of screening. We felt education had the strongest theoretical relationship with these outcomes. However, in the up-to-date models of cancer screening we used only insurance, as it has been shown that access is often a significant factor in predicting health care use.[24] Six percent of surveys were completed in a non-En-

Table 1
Study Participant Characteristics by Numeracy Status (N=264)

Variable	Total (N=264) %	Inadequate numeracy (N=195) %	Adequate numeracy (N=69)% %	P-value
Age				
40-49	44	76	24	
50-59	29	68	32	
60-69	18	65	35	
70+	9	100	0	0.008
Race				
African American	39	80	20	
White	25	45	55	
Black Caribbean	17	87	13	
Hispanic	12	91	9	
Other NH [a]	6	81	19	<0.0001
Education				
<High school	18	94	6	
High school	24	86	14	
College 1-3 years	32	65	35	
College 3 years	17	45	55	
Missing (N=21)	9	84	16	<0.0001
Income				
<$20,000	29	79	21	
$20,000-50,000	29	69	31	
>$50,000	13	36	64	
Don't know/missing	29	90	10	<0.0001
Insurance				
Private	36	59	41	
Others	64	82	18	<0.0001
Primary Care Provider				
Yes	78	71	29	
No	22	83	17	0.08
Family History of Breast Cancer				
Yes	15	74	26	
No	70	68	26	
Missing	15	100	0	0.9[b]
Family History of Colorectal Cancer				
Yes	8	71	29	
No	84	72	28	
Missing (N=20)	20	95	5	0.9[b]

(conitnued on next page)

Table1
Study Participant Characteristics by Numeracy Status (N=264)

Variable	Total (N=264) %	Inadequate numeracy (N=195) %	Adequate numeracy (N=69)% %	P-value
Perceived Risk for Breast Cancer				
<Average	36	69	31	
Average	41	71	29	
>Average	8	64	36	
Missing (N=40)	15	98	2	0.8 [b]
Perceived Risk for Colorectal Cancer				
<Average	36	73	27	
Average	40	66	34	
>Average	7	67	33	
Missing (N=46)	17	97	33	0.6 [b]
Recruitment Site				
Hospital	47	59	41	
Community health center	53	87	13	<0.0001
Survey Language				
English	94	73	27	
Other	6	94	6	0.08

Note.
a Non-Hispanic
b Subjects with missing data were excluded from the analyses.

glish language, and the survey language was not found to be significantly related to numeracy or any of the primary outcomes; hence this factor was not included in multivariate models. Therefore the final variables in the model of numeracy were age, race, and education. The models of knowledge of cancer screening included numeracy, age, race, education, PCP, and family history of the disease. The models of being up-to-date with cancer screening included numeracy, age, race, insurance, PCP, and family history of the disease.

In the multivariate regressions, subjects who failed to answer the questions about screening knowledge were grouped with those who incorrectly answered this question, and subjects with missing information about being up-to-date with screening were grouped with those who were not up-to-date. Thus all regressions included data on 264 subjects, with the exception of colorectal cancer up-to-date, in which women under age 50 were excluded, leaving data on 152 subjects in the regression analysis.

We performed sensitivity analysis by excluding the subjects who failed to answer all 3 numeracy questions. Breast and colorectal cancer screening models were also run after excluding subjects who failed to answer the questions which determined being up-to-date.

All analyses were 2-sided and were considered to be statistically significant at a P-value of less than 0.05. We used PC-SAS version 8 (Cary, NC) for our analysis.

RESULTS

Study Population

During the study period, 392 women were approached to participate, of whom 282 agreed to participate and completed a

Table 2
Study Participant Response to Numeracy Criteria (N=264)

Numeracy Questions	Incorrect response N (%)
1. Coin-Flip (500/1000)	174 (66)
2. Provide quantitative response to all 3 risk-perception questions	105 (40)
3. Lifetime risk >= 5-year risk	34 (13)

survey (72% response rate). Eighteen women were ineligible based on age or personal history of breast cancer, leaving 264 surveys for analyses. There were 123 (47%) subjects from the 2 hospital-based practices and 141 (53%) from community health centers. Only 16 surveys (6%) were completed in languages other than English, and numeracy did not differ significantly between English and non-English surveys. Therefore, all subsequent analyses combined English and non-English surveys.

Table 1 illustrates the demographic, social, and health characteristics of study subjects by numeracy status. The mean age of subjects was 55 years (SD=10.4) ranging from 40 to 84 years. Based on the distribution, we categorized age to study its association with numeracy. Approximately half of the women surveyed were African American or Caribbean black, 25% were white, 12% were Hispanic, and 6% were of other non-Hispanic minority race. Although 49% reported at least some college education, one third reported an annual income of less than $20,000, and only 36% had private health insurance. The majority of the women (78%) had a primary care physician. Fifteen percent had a first-degree relative with breast cancer and 8% with colorectal cancer. A small percentage of women considered themselves to be at more than average risk for developing breast (8%) or colorectal (7%) cancer.

Numeracy

Twenty-six percent (69/264) of participants were categorized as numerate based on the correct response to all 3 criteria outlined above. Seventy-four percent of participants had inadequate numeracy, and Table 2 illustrates the percent of incorrect response to each criterion for adequate numeracy status. Sixty-six percent (174/264) of participants gave an incorrect response to the coin-flip question, 40% (105/264) were unable to provide answers to all 3 risk-perception questions, and 13% (34/264) incorrectly estimated that their 5-year risk of breast cancer was greater than their lifetime risk. Among the 74% (195/264) of participants who had inadequate numeracy, the majority 92% (174/195) were categorized based on their inability to answer the first coin-flip question alone. Only an additional 8% (21/195) of participants were categorized as having inadequate numeracy based on an incorrect response to the second or third criterion (not providing answer to all risk-perception questions 5/21, lifetime cancer risk 12/21, and 4/21 participants fulfilled both criteria.

In bivariate analyses, the rate of adequate numeracy increased with advancing age but was found to be lowest (0%) among subjects over 70 years of age (Table 1). Adequate numeracy rates were also higher among subjects who were white and who reported higher income, more years of formal education, and private insurance. More of the subjects recruited from community health center sites than hospital sites were innumerate (87% and 59% respectively). In adjusted analyses the odds of having adequate numeracy increased with higher education; those with some college or technical school had increased odds of having adequate numeracy (OR= 2.7, CI 1.04 to 7.2) when compared to subjects with less than high school education. African Americans had a lower odds of having adequate numeracy when compared to whites (OR=0.3, CI 0.1 to 0.6). Because site of care was significantly associated with race, education, income, and insurance, it was not included in any of the multivariate analyses.

Table 3
Knowledge of Cancer Screening Guidelines and Up-to-date with Cancer Screening for Breast and Colorectal Cancer (unadjusted)

Variables	Breast Cancer		Colorectal Cancer	
	Knowledgeable N=81 (32%)	Up-to-date N=191 (72%)	Knowledgeable N=58 (25%)	Up-to-date[c] N=72 (42%)
Numerate				
Yes	48	77	35	51
No	25	71	17	46
	P=0.0003	P= 0.3	**P=0.003**	P= 0.6
Age (years)				
40-49	29	61	19	0
50-59	37	82	33	42
60-69	29	88	21	67
70+	22	70	4	35
	P= 0.5	**P= 0.0008**	**P= 0.02**	**P= 0.002**
Race				
African American	29	68	20	51
White	43	85	34	48
Caribbean Blacks	22	61	13	40
Hispanic	30	82	27	38
Other NH [a]	13	63	0	60
	P= 0.06	**P= 0.01**	**P= 0.007**	P= 0.8
Education				
< High school	21	65	17	44
High school	37	75	19	41
College 1-3 years	31	81	26	52
College >3 years	39	82	32	52
Missing	20	36	8	45
	P= 0.2 [b]	P= 0.1 [b]	P= 0.2 [b]	P= 0.8 [b]
Income				
<$20,000	27	82	17	42
$20,000-50,000	34	70	26	60
>$50,000	42	85	36	52
Missing	26	60	17	36
	P= 0.3 [b]	**P= 0.006 [b]**	P=0.07 [b]	P= 0.1 [b]
Insurance				
Private	38	76	26	56
Other	27	70	20	41
	P=0.07	P= 0.3	P= 0.2	P= 0.06
Primary Care Provider				
Yes	31	80	24	50
No	31	47	14	38
	P= 0.9	**P= 0.0001**	P= 0.09	P= 0.2
Family History				
Yes	23	82	23	51
No	32	74	24	33
Missing	31	56	5	33
	P= 0.3 [b]	P= 0.1 [b]	P=0.2 [b]	P= 0.2 [b]
Perceived Personal Risk of Cancer				
Low	34	71	24	40
Medium	30	79	25	53
High	36	68	11	38
Missing	23	60	15	52
	P= 0.5 [b]	P= 0.14 [b]	P= 0.4 [b]	P= 0.5 [b]
Knowledge of Screening Guidelines				
Yes	N.A	73	N.A.	49
No	N.A.	72	N.A.	47
		P= 0.9		P= 0.8

Note.
a Non-Hispanic
b Subjects with missing data were excluded from the analyses.
c The N for these comparisons is 152, as women under 50 were excluded.

Table 4
Multivariate Analysis of Knowledge of Cancer Screening Guidelines and Up-to-date with Cancer Screening by Numeracy Status

	Knowledge of screening guidelines Odds Ratio (Confidence Interval)		Up-to-date with cancer screening Odds Ratio (Confidence Interval)	
	Breast Cancer [a]	Colorectal Cancer [b]	Breast Cancer [c]	Colorectal Cancer [d]
Numeracy				
Adequate	2.7 (1.4-5.4)	1.6 (0.8-3.5)	0.7 (0.3-1.6)	1.1 (0.5-2.7)
Inadequate	reference	reference	reference	reference

a Logistic regression model predicting knowledge of breast cancer screening guidelines adjusted for numeracy, age, race, education, primary care provider, and family history of breast cancer. N=264
b Logistic regression model predicting knowledge of colorectal cancer screening guidelines adjusting for numeracy, age, race, education, primary care provider, and family history of colorectal cancer. N=264
c Logistic regression model predicting up-to-date with breast cancer screening adjusting for numeracy, age, race, insurance, primary care provider, and family history of breast cancer. N=264
d Logistic regression model predicting up-to-date with colorectal cancer screening adjusting for numeracy, age, race, insurance primary care provider, and family history of colorectal cancer. N for this model 152 as women less than 50 years were excluded.

Knowledge of Cancer Screening Guidelines

A third of our participants were considered knowledgeable of the current recommended age to begin screening mammography whereas only 23% were considered knowledgeable of the current recommended age to start colorectal screening. The majority of the subjects responding incorrectly chose an age younger than the recommended screening: 63% chose under 40 years for mammography; 72% chose under 50 years for colorectal screening.

Table 3 depicts the results of our bivariate analyses for the knowledge outcome. Adequate numeracy was significantly associated with being knowledgeable of both breast (P=0.0003) and colorectal cancer (P=0.003) screening guidelines. Age 50-59 years and white race were also associated with increased knowledge of colorectal cancer screening guidelines. In multivariate analysis (Table 4), we found that adequate numeracy was significantly associated with an increased odds (OR= 2.7, CI 1.4 to 5.4) of being knowledgeable of breast cancer screening but not of colorectal cancer screening guidelines. The odds of being knowledgeable of colorectal cancer screening guidelines, however, do increase with age (OR= 2.0, CI 1.2 to 3.4).

Up-to-date With Cancer Screening

Overall, our subjects reported high screening rates. Seventy-seven percent were up-to-date with mammography screening, and 49% were up-to-date with colorectal cancer screening.

Table 3 also presents bivariate associations of our other primary outcome up-to-date with cancer screening. Breast cancer screening rates were highest among women of 60-69 years (P<0.0008), higher among whites and Hispanics (P<0.01), extremes of income (P<0.006), and among those with a primary care physician (P<0.0001). No association was found between mammography screening and adequate numeracy. We also found no association between knowledge and screening. In multivariate analysis (Table 4), adequate numeracy was not a significant predictor of being up-to-date with mammography screening. Our model predicted the odds of being up-to-date for breast cancer screening to be higher with age (OR= 1.8, CI 1.3 to 2.4) and among subjects with a primary care provider (OR= 5.1, CI 2.6 to 10.2) but lower among the ethnic category Caribbean blacks (OR=

0.2, CI 0.07 to 0.6) when compared to whites. When this model was tested for sensitivity to missing numeracy or outcome data (by excluding those with missing answers) Caribbean black race was no longer statistically different from white race (OR=0.4, CI 0.1-1.8), and other results remained the same.

Bivariate analyses of colorectal cancer screening found higher proportions of subjects of age 50-59 and 60-69 years up-to-date (P<0.002). Adequate numeracy was not associated with being up-to-date with colorectal cancer screening. Again, we found no association between knowledge and screening. In multivariate analysis the odds of being up-to-date were higher with age (OR= 3.0, CI 1.5 to 5.8), among African Americans (OR=3.0, CI 1.03 to 8.9) when compared to whites, and among subjects with private insurance (OR= 3.6, CI 1.5 to 8.7). Having a primary care physician and numeracy were not significantly associated with colorectal cancer screening. There were no differences in model results when those subjects who failed to answer numeracy or screening questions were excluded.

DISCUSSION AND CONCLUSION

Discussion

Our study is the first to explore the role of numeracy on knowledge and cancer screening practices in a population most at risk for poor cancer outcomes. Among this urban group of women with diverse racial/ethnic backgrounds, we found that having adequate numeracy skills was associated with increased knowledge of breast cancer screening guidelines but not with colorectal cancer screening guidelines. We did not find an association between having adequate numeracy and being up-to-date with screening practices, for either breast or colorectal cancer.

It is common for physicians and other health care providers to provide information about rates, percentages, and proportions when discussing health information, especially when evaluating the likelihood of developing cancer. Studies have found that use of numeric concepts is an efficient means of discussing risk with patients. However our study exhibits a stark reality about informed decision making; more than two thirds of our subjects were not able to answer a simple probability question, and 74% of the population had inadequate numeracy by our standards. This raises questions about how to communicate health information to patients. With such a high rate of inadequate numeracy, clinicians need to significantly simplify how they discuss concepts such as probability. We recommend confirming patients' comprehension for all important clinical discussions as a universal precaution against confusion and failed communication.[25]

Previous studies in minority populations suggest that women are aware of mammography and breast cancer but lack specific knowledge about screening guidelines.[26] A study of low-education, low-income Latina women found that 38-45% were aware of recommendations for mammography frequency.[27] Similarly, we found only 32% of women in our study were knowledgeable of breast cancer screening guidelines and 25% of colorectal cancer screening guidelines. Previous studies have found that lack of knowledge about cancer screening was associated with poor screening behavior.[26,27,29] Our study findings did not support this argument. On the contrary, our population reported high screening rates.

Studies have shown associations between low rates of cancer screening, lower educational attainment, and suboptimal health care access with low socioeconomic status.[30-38] Seventy-seven percent of our subjects had mammography screening within the last 2 years, and 49% had colorectal cancer screening—higher than the national averages of 70% and 37% respectively. We found different predictors for breast and colorectal cancer screening in our population. Although having a primary care provider was the strongest predictor for breast cancer screening, having private insurance was the strongest predictor of colorectal cancer screening. These findings support a recent study on mammography decision making which found that free mammograms were not sufficient incentive for never-screened minorities unless also accompanied by a recommendation from a provider or cancer organization.[39] Provider recommendation has also been shown to be an important predictor for colorectal screening in other studies[2] though our findings did not support this finding. These findings may reflect state-funded initiatives. For example, numerous public health programs target breast cancer screening in the Boston area,

including the availability of mammography services free of charge for income-eligible Massachusetts State residents, which may explain why sociodemographic characteristics do not drive screening in this setting.[40] In this respect, our cohort is similar to veterans, who have access to comprehensive preventive health care coverage and have been found to have higher screening rates than nonveteran populations.[41]Similar programs for colorectal cancer screening do not currently exist in Massachusetts, therefore shifting the predictors of screening to more traditional access-driven barriers such as insurance status.

Although the literature is rich with studies exhibiting racial disparity in cancer screening rates, our results did not show any racial or ethnic disparities in the population surveyed.[3,4,6,8-11,42] Again, this may be because our population had fewer barriers to access for care; 85% of the subjects had a primary care physician, and 95% had some kind of insurance, though 45% had only public forms of insurance. This is supported by the findings of Dolan and colleagues, who showed that in the context of equal access to care and similar insurance statuses, African American veterans were more likely to receive colorectal cancer screening than white veterans were.[41] Our study further supports the hypothesis that having both increased access to care and a regular primary care physician plays a significant role in cancer prevention and regular screening.

Our study results are limited to an ethnically diverse urban population with high screening rates and may not be easily generalized to other medical settings. Moreover, by the nature of the study design, the majority of participants have a primary care provider, one of the strongest predictors of compliance with cancer screening. Knowledge of cancer screening recommendations was similar to that of other ethnically diverse urban populations.[27] We were unable to verify the self-reported mammogram and colorectal cancer screening among these women; however, the literature supports high accuracy in self reported data: if asked within 2 years of the screening test, 75% accuracy among those reporting breast cancer screening and 85% accuracy for reported colorectal screening.[43-45] We also acknowledge limitations in our measurements that may be attributable to the fact that this survey was designed primarily to measure risk perception. We utilized awareness of recommended age to begin cancer screening as a proxy for knowledge as this was the only knowledge question included in the study. Though controversy regarding age to begin breast cancer screening does exist, most organizations recommend starting baseline screening at age 40 or higher.[46] Numeracy was assessed utilizing an adaptation of 3 criteria from prior work by Black and colleagues.[21] This measurement was chosen specifically based on its previous use in the evaluation of cancer risk perception. Although our numeracy measurement has not been validated, the demographic associations exhibited in our numeracy analyses do conform to those found on the NAALS with respects to age, race/ethnicity, and education.[19]

Conclusion

This study illustrates that in a multiethnic urban population, adequate numeracy and knowledge of screening guidelines are less important than access to care and physician recommendation in achieving high cancer screening rates. It does not undermine the fact that studies in the past have shown numeracy to be an important tool for health care delivery and improving outcomes in other chronic medical problems like diabetes and hypertension. Nonetheless, the very low rate of adequate numeracy seen in this cohort, even in patients with a high level of educational attainment, should prompt physicians to be conscious of their patients' understanding and interpretation of their recommendations.

Acknowledgments

Dr Aggarwal and Dr Battaglia were both supported by a training award from the American Cancer Society (ACS PTAPM 97-185-04). Dr. Battaglia was supported by 2 career development awards (NIH K12 HD043444-02 and ACS CCCDA 03-228-01). Preliminary results of this study were published in abstract format and subsequently presented as a poster at the annual meeting of the Society for General Internal Medicine in April 2005. ■

REFERENCES

1.Jemal A, Murray T, Ward E, et al. Cancer statistics, 2005. *CA Cancer J Clin.*

2005;55(1):10-30.
2. Seeff LC, Nadel MR, Klabunde CN, et al. Patterns and predictors of colorectal cancer test use in the adult U.S. population. *Cancer.* 2004;100(10):2093-2103.
3. ACS, Facts and Figures (on line). Available at: http://www.cancer.org/docroot/STT/content/STT_1x_Cancer_Facts_Figures_2006.asp. Accessed May 11, 2005.
4. CDC, Trends in the Health of Americans with Special Feature on Adults 55-64 Years (online). Available at: http://www.cdc.gov/nchs/hus.htm. Accessed June 15, 2005.
5. SEER, Surveillance Epidemiology and End Results, Rate/Trend Comparisons (online). Available at: http://statecancerprofiles.cancer.gov/rt_choice.html. Accessed June 15, 2005.
6. Polite BN, Olopade OI. Breast cancer and race: a rising tide does not lift all boats equally. *Perspect Biol Med.* 2005;48(Suppl 1):S166-S175.
7. McMahon LF, Jr., Wolfe RA, Huang S, et al. Racial and gender variation in use of diagnostic colonic procedures in the Michigan Medicare population. *Med Care.* 1999;37(7):712-717.
8. Richards RJ, Reker DM. Racial differences in use of colonoscopy, sigmoidoscopy, and barium enema in Medicare beneficiaries. *Dig Dis Sci.* 2002;47(12):2715-2719.
9. Cooper GS, Koroukian SM. Racial disparities in the use of and indications for colorectal procedures in Medicare beneficiaries. *Cancer.* 2004;100(2):418-424
10. Hegarty V, Burchett BM, Gold DT, et al. Racial differences in use of cancer prevention services among older Americans. *J Am Geriatr Soc.* 2000;48(7):735-740.
11. Swan J, Breen N, Coates RJ, et al. Progress in cancer screening practices in the United States: results from the 2000 National Health Interview Survey. *Cancer.* 2003;97(6):1528-1540.
12. Eley JW, Hill HA, Chen VW, et al. Racial differences in survival from breast cancer. Results of the National Cancer Institute Black/White Cancer Survival Study. *JAMA.* 1994;272(12):947-954.
13. Kalet A, Roberts JC, Fletcher R. How do physicians talk with their patients about risks? *J Gen Intern Med.* 1994;9(7):402-404.
14. Paasche-Orlow MK, Parker RM, Gazmararian JA, et al. The prevalence of limited health literacy. *J Gen Intern Med.* 2005;20(2):175-184.
15. DeWalt DA, Pignone MP. Reading is fundamental: the relationship between literacy and health. *Arch Intern Med.* 2005;165(17):1943-1944.
16. Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional health status among older adults. *Arch Intern Med.* 2005;165(17):1946-1952.
17. Baker DW, Williams MV, Parker RM, et al. Development of a brief test to measure functional health literacy. *Patient Educ Couns.* 1999;38(1):33-42.
18. USPSTF, United States Preventive Services Task Forces (online), Available at: www.preventiveservices.ahrq.gov/. Accessed May 26, 2005.
19. Kutner M, Greenberg E, Jin Y, Paulsen C. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy (NCES 2006-483). U.S. Department of Education. Washington, DC: National Center for Education Statistics, 2006.
20. Schwartz LM, Woloshin S, Black WC, et al. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med.* 1997;127(11):966-972.
21. Black WC, Nease RF, Jr., Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *J Natl Cancer Inst.* 1995; 87(10): 720-731.
22. American Cancer Society, Early Detection (online), Available at: http://www.cancer.org/docroot/PED/ped_2_1_methods.asp?sitearea=PED. Accessed September 15, 2006.
23. Schettino MR, Hernandez-Valero MA, Moguel R, et al. Assessing breast cancer knowledge, beliefs, and misconceptions among Latinas in Houston, Texas. *J Cancer Educ.* 2006; 21(Suppl 1):S42-S46.
24. Hsia J, Kemper E, Kiefe C, et al. The importance of health insurance as a determinant of cancer screening: evidence from the Women's Health Initiative. *Prev Med.* 2000;31(3):261-270.
25. Paasche-Orlow MK, Schillinger D, Greene SM, et al. How health care systems can begin to address the challenge of limited literacy. *J Gen Intern Med.* 2006;21(8):884-887.
26. Peragallo NP, Fox PG, Alba ML. Breast care among Latino immigrant women in the U.S. *Health Care Women Int.* 1998;19(2):165-172.
27. Valdez A, Banerjee K, Ackerson L, et al. Correlates of breast cancer screening among low-income, low-education Latinas. *Prev Med.* 2001;33(5):495-502.
28. Aro AR, de Koning HJ, Absetz P, et al. Two distinct groups of non-attenders in an organized mammography screening program. *Breast Cancer Res Treat.* 2001;70(2):145-153.
29. Burton MV, Warren R, Price D, et al. Psychological predictors of attendance at annual breast screening examinations. *Br J Cancer.* 1998;77(11):2014-2019.
30. Peek ME, Han JH. Disparities in screening mammography. Current status, interventions and implications. *J Gen Intern Med.* 2004;19(2):184-194.
31. Lannin DR, Mathews HF, Mitchell J, et al. Influence of socioeconomic and cultural factors on racial differences in late-stage presentation of breast cancer. *JAMA.* 1998;279(22):1801-1807.
32. Gordon NH, Crowe JP, Brumberg DJ, et al. Socioeconomic factors and race in breast can-

cer recurrence and survival. *Am J Epidemiol.* 1992;135(6):609-618.
33. Berg JW, Ross R, Latourette HB. Economic status and survival of cancer patients. *Cancer.*1977;39(2):467-477.
34. Dayal HH, Power RN, Chiu C. Race and socioeconomic status in survival from breast cancer. *J Chronic Dis.* 1982;35(8):675-683.
35. Freeman HP, Wasfie TJ. Cancer of the breast in poor black women. *Cancer.* 1989;63(12):2562-2569.
36. Neal RD, Allgar VL. Sociodemographic factors and delays in the diagnosis of six cancers: analysis of data from the 'National Survey of NHS Patients: Cancer'. *Br J Cancer.* 2005;92(11):1971-1975.
37. Singh SM, Paszat LF, Li C, et al. Association of socioeconomic status and receipt of colorectal cancer investigations: a population-based retrospective cohort study. *CMAJ.* 2004;171(5):461-465.
38. Tavani A, Fioretti F, Franceschi S, et al. Education, socioeconomic status and risk of cancer of the colon and rectum. *Int J Epidemiol.* 1999;28(3):380-385.
39. Veit CT. Motivating mammography adherence in elderly Latinas: a test of three mathematical models of decision making. *Med Decis Making.* 1999;19(4):448-465.
40. WHN, W.s.H.N. (online). Available at: http://www.womenshealthnetwork.org/index.php. Accessed September 8, 2006.
41. Dolan NC, Ferreira MR, Fitzgibbon ML, et al. Colorectal cancer screening among African-American and white male veterans. *Am J Prev Med.* 2005;28(5):479-482.
42. Macrae FA, St John DJ, Ambikapathy A, et al. Factors affecting compliance in colorectal cancer screening. Results of a study performed in Ballarat. *Med J Aust.* 1986;144(12):621-623.
43. Etzi S, Lane DS, Grimson R. The use of mammography vans by low-income women: the accuracy of self-reports. *Am J Public Health.* 1994;84(1):107-109.
44. McPhee SJ, Nguyen TT, Shema SJ, et al. Validation of recall of breast and cervical cancer screening by women in an ethnically diverse population. *Prev Med.* 2002;35(5):463-473.
45. Madlensky L, McLaughlin J, Goel V. A comparison of self-reported colorectal cancer screening with medical records. *Cancer Epidemiol Biomarkers Prev.* 2003;12(7):656-659.
46. ACS: States with Mammography Screening Coverage Laws (online), Available at: http://www.cancer.org/docroot/CRI/content/CRI_2_6x_Breast_Cancer_Early_Detection.asp?sitearea. Accessed September 7, 2006.

Treating Homeless Opioid Dependent Patients with Buprenorphine in an Office-Based Setting

Daniel P. Alford, MD, MPH[1,2,3], Colleen T. LaBelle[1,3], Jessica M. Richardson[1], James J. O'Connell, MD[4], Carole A. Hohl, MHS[4], Debbie M. Cheng, ScD[1,5], and Jeffrey H. Samet, MD, MA, MPH[1,2,6]

[1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston, MA, USA; [2]Boston University School of Medicine, Boston, MA, USA; [3]Boston Public Health Commission, Boston, MA, USA; [4]Boston Health Care for the Homeless Program, Boston, MA, USA; [5]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; [6]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA.

CONTEXT: Although office-based opioid treatment with buprenorphine (OBOT-B) has been successfully implemented in primary care settings in the US, its use has not been reported in homeless patients.

OBJECTIVE: To characterize the feasibility of OBOT-B in homeless relative to housed patients.

DESIGN: A retrospective record review examining treatment failure, drug use, utilization of substance abuse treatment services, and intensity of clinical support by a nurse care manager (NCM) among homeless and housed patients in an OBOT-B program between August 2003 and October 2004. Treatment failure was defined as elopement before completing medication induction, discharge after medication induction due to ongoing drug use with concurrent nonadherence with intensified treatment, or discharge due to disruptive behavior.

RESULTS: Of 44 homeless and 41 housed patients enrolled over 12 months, homeless patients were more likely to be older, nonwhite, unemployed, infected with HIV and hepatitis C, and report a psychiatric illness. Homeless patients had fewer social supports and more chronic substance abuse histories with a 3- to 6-fold greater number of years of drug use, number of detoxification attempts and percentage with a history of methadone maintenance treatment. The proportion of subjects with treatment failure for the homeless (21%) and housed (22%) did not differ (P=.94). At 12 months, both groups had similar proportions with illicit opioid use [Odds ratio (OR), 0.9 (95% CI, 0.5–1.7) P=.8], utilization of counseling (homeless, 46%; housed, 49%; P=.95), and participation in mutual-help groups (homeless, 25%; housed, 29%; P=.96). At 12 months, 36% of the homeless group was no longer homeless. During the first month of treatment, homeless patients required more clinical support from the NCM than housed patients.

CONCLUSIONS: Despite homeless opioid dependent patients' social instability, greater comorbidities, and more chronic drug use, office-based opioid treatment with buprenorphine was effectively implemented in this population comparable to outcomes in housed patients with respect to treatment failure, illicit opioid use, and utilization of substance abuse treatment.

KEY WORDS: buprenorphine; drug dependence; primary care; homelessness.
DOI: 10.1007/s11606-006-0023-1
 2007;22:171–176

Received January 24, 2006
Revised July 17, 2006
Accepted September 27, 2006
Published online January 17, 2007

INTRODUCTION

Opioid abuse persists as a pervasive public health problem in the United States, both heroin[1] and prescription opioid analgesics.[1,2] Opioid agonist treatment with methadone or buprenorphine is effective for treating opioid dependence.[3–12] With the advent of sublingual buprenorphine for the treatment of opioid dependence, primary care physicians in the United States gained the opportunity to effectively treat opioid-dependent patients in primary medical care settings, commonly referred to as office-based opioid treatment (OBOT).[13,14]

In 2003, the primary care clinic at Boston Medical Center (BMC) implemented an OBOT with buprenorphine (OBOT-B) program employing collaborative care between physicians and a nurse care manager (NCM).[15] All patients in the BMC primary care clinic OBOT-B program were required to have stable housing, as clinical guidelines recommend a stable social environment as an entry criterion for OBOT-B.[16,17] Using social stability as a criterion for OBOT-B precludes homeless persons, a population with a high prevalence of addiction,[18–21] leading to a high risk of illness and death.[22–25]

Unique challenges confront homeless individuals engaging in substance abuse treatment,[26] which likely contribute to their high rates of treatment failure.[27,28] Characteristics of homeless persons are correlated with relapse: lack of social support; unstable living environment; and longer duration of drug dependence.[29,30] However, research has shown that homeless persons' success in substance abuse treatment can increase under supportive circumstances.[30,31] Furthermore, despite limited literature on methadone treatment in homeless populations, published data suggest greater success with

enhanced access to[32] and unconventional methods of treatment (e.g., medical care provision in homeless health care settings).[33] Although homeless opioid dependent individuals may benefit from buprenorphine treatment, clinical guidelines have excluded them.

This retrospective cohort study characterizes the feasibility of office-based opioid treatment with buprenorphine (OBOT-B) in a homeless population. We describe a customized model of OBOT-B care and outcomes in homeless compared to housed patients.

METHODS

Program Description

Model of Care. In August 2003, OBOT-B was implemented in an urban academic medical center for homeless patients in a homeless clinic and for housed patients in a primary care clinic. Both patient groups were treated with collaborative care between qualified primary care physicians (PCPs) and a single nurse care manager (NCM). All patients were referred by their respective clinics and were to receive maintenance buprenorphine treatment for at least 12 months. *Homelessness* was defined as spending 1 or more weeks on the street or in a shelter within the previous 3 months. Upon referral by the PCP, homeless and housed patients were admitted successively or placed on a wait list when the practice was at capacity due to the federally mandated 30-patient-per-practice limit at the time of the study. Our model of care included 4 stages of treatment: (1) Determination of eligibility; (2) Medication induction; (3) Medication stabilization; (4) Treatment maintenance.

Determination of Eligibility. Assessment protocols were similar for homeless and housed groups. OBOT-B, a less structured treatment option compared to methadone maintenance treatment, required a patient assessment for appropriateness. OBOT-B appropriateness was determined by adherence with assessment appointments, mental health stability [i.e., ability to comprehend and consent to OBOT-B protocols, chronic mental illness with ongoing psychiatric care and no acute mental health issues (e.g., suicidal ideation)], absence of painful conditions requiring opioid analgesics, absence of medical contraindications (e.g., pregnancy or liver function tests greater than five times the normal level), and absence of alcohol or other drug co-dependence. Patients with sporadic alcohol or other drug use remained eligible. Patients on methadone maintenance doses greater than 30 mg per day were excluded based on current practice guidelines.[16] If deemed appropriate for OBOT-B, patients received a comprehensive assessment by the NCM that included a detailed medical, psychiatric, and substance abuse history including documentation of DSM-IV diagnosis of opioid dependence. Social history included employment and homelessness status. Admission lab work included a complete blood count, electrolytes, renal and liver function tests, hepatitis A, B, and C serologies, and urine toxicology for opioids (including oxycodone for patients with a history of oxycodone abuse), cocaine, benzodiazepines, barbiturates, and amphetamines. All patients were tested for tuberculosis exposure and were offered HIV testing and counseling. Treatment consents and individualized treatment plans were reviewed and signed by both patient and NCM. At a separate appointment, PCPs reviewed the NCM assessment and treatment plan, performed a complete physical examination, and wrote the initial buprenorphine prescription. Patients were strongly encouraged to utilize other substance abuse services, including addiction counseling and mutual-help groups [e.g., Narcotics Anonymous (NA)].

Medication Induction. Induction, occurring during the first 3 days of treatment, differed for homeless and housed patients. Homeless patients received "office induction," with up to 8 hours observation on day 1 by the NCM for the following: signs of opioid withdrawal, dose administration, and response to medication. Homeless patients left the office with a nighttime dosing protocol to be used if withdrawal recurred and then returned daily for the next 2 days for dose adjustments. In contrast, because all housed patients were reachable by telephone, they received "home induction" by following a 3-day dosing protocol at home. The 3-day protocol gave specific dosing schedules with a maximum daily dose for each day (i.e., max dose, 8, 16, and 24 mg for days 1, 2, and 3, respectively). The NCM called all housed patients at home before the first buprenorphine dose was taken and then at least daily during the 3-day induction. The NCM was available during business hours for drop-in visits and 24 hours a day via cell phone.

Medication Stabilization. Stabilization, occurring during days 4 through 14 of treatment, differed for homeless and housed patients. Homeless patients presented to the clinic daily, except weekends, for observed dosing and assessment while housed patients were seen twice per week but had at least daily NCM phone contact.

Treatment Maintenance. Ongoing monitoring for drug use and treatment adherence, occurring beyond day 14 of treatment, was based on individual patient needs in both groups. Buprenorphine prescriptions were provided in 1 to 4 week amounts determined by the patient's ability to safely secure medications. For example, 1-week amounts for patients without a secure storage place and 2–4 week amounts for patients with a secure storage place. Patients were seen at least monthly by the NCM and at least every 6 months by the PCP. Patients were expected to respond to unscheduled "call-backs" to the clinic for urine toxicology, observed dosing, pill counts, and revisions of treatment plans as needed. The need for "call-backs" was based on the NCM or PCP's clinical suspicion that the patient may have relapsed or may be diverting their medication based on abnormal urine toxicology, reports of lost pills, requests for early refills, or missed appointments. Urine was tested for opioids (including oxycodone for patients with a history of oxycodone abuse) and other drugs (including cocaine, benzodiazepines, barbiturates, and amphetamines) as well as buprenorphine at least every 3 months. Intensified treatment (i.e., substance abuse counseling) was required for patients with ongoing use of opioids, other drugs, or alcohol.

Patient Characteristics

Upon entry into OBOT-B, patients' demographics, employment, homelessness status, involvement of socially supportive individuals, opioid, other drug and alcohol use, and substance abuse, medical and psychiatric histories were recorded.

Table 1. Comparison of Homeless and Housed Patients at Time of Entry into Office-Based Opioid Treatment with Buprenorphine

	Homeless (*N*=44)	Housed (*N*=41)	*P*-value
Demographics			
Male (%)	59	76	.11
Race/ethnicity			<.001
White (%)	41	85	
Hispanic/Latino (%)	34	2	
Black/African American (%)	25	12	
Mean age years (SD)	42 (9.1)	34 (10.4)	<.001
Employed (%)	5	34	<.001
Involvement of social support in care (%)	2	90	<.001
Comorbidity			
Self-reported psychiatric illness (%)	95	54	<.001
HIV-infected (%)	30	5	.003
Hepatitis C-infected (%)	95	44	<.001
Substance abuse history			
Opioid at admission			.001
Heroin (%)	84	63	
Sustained-release oxycodone (%)	0	27	
Methadone maintenance (%)	16	10	
Any methadone maintenance history (%)	59	10	<.001
Median years drug use (range)	15 (5–30)	5 (2–12)	<.001
Median detoxification attempts (range)	18 (5–40)	5 (0–20)	<.001

Outcome Assessments

Time to Treatment Failure. Our primary outcome was time to "treatment failure" defined as any of the following: elopement before completing medication induction (elopement did not occur after induction); discharge after medication induction due to ongoing use of opioids, other drugs or alcohol with concurrent nonadherence with intensified substance abuse treatment; or discharge due to disruptive behavior (e.g., threatening staff, theft of clinic property).

Secondary outcomes included reasons for leaving the OBOT-B program (including treatment failure and successful program departures), illicit opioid and other drug use, number of NCM contacts, utilization of recommended care and social indicators (e.g., homelessness and employment status).

Reasons for Leaving the OBOT-B Program. All patients who left the program before 12 months were categorized into one of the following 2 groups: "treatment failure" (defined above); or "successful program departure." Patients classified as "successful program departure" left the program because they relocated and were transferred to another OBOT program, they needed a more structured treatment setting and were transferred to a methadone maintenance program, or they were fully adherent with treatment for at least 4 months and were approved for a medication taper by both the NCM and PCP.

Illicit Drug Use. Both planned and "call-back" urine toxicology were conducted at least once every 3 months. In each assessment window (i.e., at entry, 3, 6, 9, and 12 months of treatment), the test that was closest, yet prior, to the timepoint was reported, in an attempt to obtain a measure of abstinence that was not biased by the number of urine toxicology tests performed. Urine toxicology tests were mostly unsupervised but measures were taken to try to minimize falsified tests (e.g., testing for buprenorphine metabolites, urine temperature, and creatinine concentration). Urine collections were supervised when the NCM suspected falsified tests due to the following circumstances: patient had a recent history of abnormal urine tests; specimens were cold or diluted; patient demonstrated aberrant behavior (e.g., missed clinic appointments, requested early medication refill). The following 3 outcomes were determined by record review: *number of NCM contacts* (phone calls and clinic visits during each month of treatment); *utilization of recommended care* (involvement in substance abuse counseling and/or mutual-help groups defined by at least weekly attendance for 1 or more months before record review); and *social indicators* [homelessness, employment status (self-reported employment for at least 1 month) and presence of social supports (family member or friend actively involved in the patient's substance abuse rehabilitative progress and in contact with the NCM within the previous 3 months)].

For all patients who left the OBOT-B program (treatment failure and successful program departure), follow-up data post-departure were not collected.

Analysis

Descriptive statistics (e.g., means and proportions) were used to characterize homeless and housed groups. Exploratory, hypothesis-generating tests were then performed to compare process and outcome measures between groups. Chi-square or Fisher's exact tests were used to compare dichotomous outcomes and *t*-tests or Wilcoxon rank sum tests were used to compare continuous outcomes between groups. NCM contacts were described for each group using the mean number of contacts per patient for each month of follow-up. Poisson regression models were constructed to estimate the rate ratio of NCM contacts in the homeless relative to housed groups. Generalized estimating equations (GEE) regression models were used to examine the association between homelessness status and drug use. Data collected from entry, 3, 6, 9, and 12-month timepoints were analyzed. The proportions of treatment failures were estimated using the Kaplan–Meier method and compared between groups using the log-rank test. Successful program departures were censored at their date of departure, as follow-up data were not available beyond that date. Reported *P*-values are two-tailed, and a *P*-value of less than .05 was considered statistically significant. All analyses were run using SAS statistical software.[34] This research was approved by the Institutional Review Board at Boston University Medical Center.

RESULTS

Patient Characteristics

During the 12 months examined, 44 homeless and 41 housed patients were enrolled. The homeless group had fewer males, fewer whites, were older, and less employed compared to the

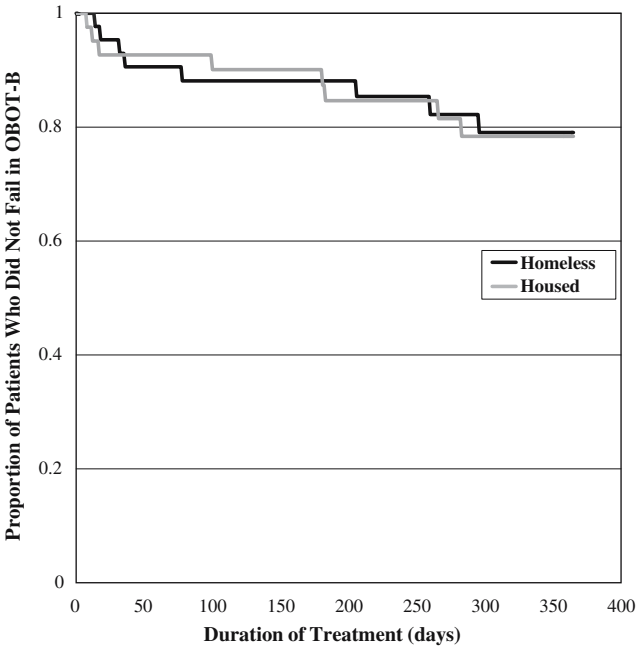


Figure 1. Kaplan–Meier estimates of the proportion of homeless and housed patients who did not fail office-based opioid treatment with buprenorphine. *P*=.94 for the comparison between homeless and housed subjects by the log-rank test.

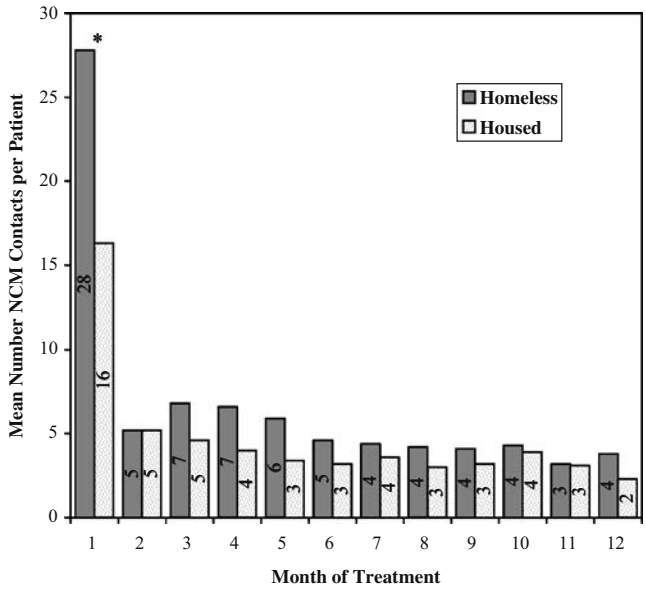


Figure 2. Mean number of monthly nurse care manager (NCM) contacts per homeless and housed patient over 12 months of office-based opioid treatment with buprenorphine. *RR=1.7 (95% CI=1.48–1.95); *P*<.0001 (homeless vs housed in month #1).

housed group. There was a dramatic difference in social supports, essentially nonexistent for the homeless patients but present for almost all housed patients. Similarly, comorbidities were clearly different in the homeless versus the housed group with 95% reporting psychiatric illness versus 54%, 30% HIV-infected versus 5%, and 95% hepatitis C infected versus 44%, respectively. At time of study entry, 84% of the homeless patients were dependent on heroin with none dependent on oxycodone versus 63% heroin dependent and 27% oxycodone dependent in the housed patients. The remaining 16% of the homeless and 10% of housed patients were transferred from methadone maintenance treatment. Homeless patients also had more chronic substance abuse histories with a 3- to 6-fold greater number of years of drug

Table 2. Reasons for Leaving Office-Based Opioid Treatment with Buprenorphine Among Homeless (*n*=20*) and Housed (*n*=17*) Patients

	Homeless *N* (%)	Housed *N* (%)
Treatment failure		
Elopement during induction††	3 (15)	3 (18)
Ongoing drug use and treatment nonadherence§	3 (15)	5 (29)
Disruptive behavior//	2 (10)	0 (0)
Successful program departures		
Transfer to another OBOT-B program	1 (5)	0 (0)
Transfer to methadone maintenance program	5 (25)	4 (24)
Successful taper¶	6 (30)	5 (29)

*Of the 44 homeless and 41 housed subjects who entered the study, 20 homeless and 17 housed subjects did not remain in the program for 12 months.

††No elopement occurred after induction period.

§Ongoing use of opioids or other drugs with concurrent nonadherence of intensified treatment.

//Including threatening staff and theft of clinic property.

¶Treatment adherence including no drug use for at least 4 months, followed by successful tapering off of the program.

Table 3. Outcomes of Homeless (*N*=44) and Housed (*N*=41) Patients after 12-months of Office-Based Opioid Treatment with Buprenorphine

	Homeless *N* (%)	Housed *N* (%)	P-value
Attending counseling†			0.95
Yes	20 (46)	20 (49)	
No	4 (9)	4 (10)	
Unknown	20 (46)	17 (42)	
Attending mutual help groups†			0.96
Yes	11 (25)	12 (29)	
No	13 (30)	12 (29)	
Unknown	20 (46)	17 (42)	
Currently homeless††			0.03
Yes	8 (18)	1 (2)	
No	16 (36)	23 (56)	
Unknown	20 (46)	17 (42)	
Currently employed§			0.07
Yes	17 (39)	23 (56)	
No	7 (16)	1 (2)	
Unknown	20 (46)	17 (42)	
Involvement of social support in care//			0.50
Yes	22 (50)	24 (59)	
No	2 (5)	0 (0)	
Unknown	20 (46)	17 (42)	

†Weekly attendance for 1 or more months.

††Spending 1 or more weeks on the street or in a shelter within the previous 3 months.

§Maintaining employment for 1 or more months.

//Family member or friend actively involved in the patient's substance abuse rehabilitative progress and in contact with the NCM within the previous 3 months.

use, number of detoxification attempts, and percentage with a history of methadone maintenance treatment (Table 1).

Patient Outcomes

The estimated proportion of subjects with treatment failure was 21% for the homeless and 22% for the housed. Homeless and housed patients did not differ in their risk for treatment failure during follow-up (P=.94) (Fig. 1).

Mean duration of OBOT-B retention was 9 months for both homeless and housed groups. The percentage of patients remaining in OBOT-B (i.e., had not failed, transferred or tapered) at the 3-, 6-, 9- and 12-month timepoints was 77%, 73%, 70%, and 55% in the homeless group and 93%, 80%, 68%, and 61% in the housed group. The number of patients who left treatment and the reasons for leaving treatment before 12-months, including the treatment failure and successful program departure groups, appeared similar for both groups (Table 2).

Of the patients remaining in treatment at 3 months, urine samples were positive for illicit opioids in 11% of homeless and 20% of housed patients and at 12 months positive in only 4% in both groups. At entry, urines were positive for "other drugs" in 34% of homeless and 32% of housed patients and of those remaining in treatment at 12 months, positive in only 8% in both groups. Over the 12-month period, no significant associations were found between homelessness status and use of opioids [Odds ratio (OR) 0.9 (95% CI, 0.5–1.7); P=.8] or other drugs [OR 1.7 (95% CI, 0.7–4.3); P=.2).

The mean number of NCM contacts per patient was estimated by homelessness status for each month of follow-up (Fig. 2). During the first month of treatment (induction and stabilization), the rate of NCM contacts per patient in the homeless group was 1.7 times that in the housed group (95% CI=1.48–1.95; P<.0001). At 12 months, 46% of homeless and 49% of housed patients reported at least weekly receipt of substance abuse counseling and 25% of homeless and 29% of housed patients reported at least weekly mutual-help group attendance (Table 3). At 12 months, of the 44 homeless patients originally enrolled, 16 (36%) no longer met the criteria for homelessness. Employment increased for both homeless and housed patients that remained in the program for 12 months.

DISCUSSION

Office-based opioid treatment with buprenorphine (OBOT-B) was effectively implemented in homeless patients with outcomes comparable to housed patients. When compared to housed patients, homeless patients appeared to have a similar proportion of treatment successes (i.e., retention in treatment, successful program departure, decreased drug use, and utilization of counseling and mutual help groups), despite having more chronic substance abuse histories (e.g., greater number of years of drug use, detoxification attempts, and history of methadone maintenance treatment), more medical and psychiatric comorbidities, and less social support. Employment and housing increased substantially among homeless patients. Homeless patients required more clinical support by the Nurse Care Manager (NCM) than housed patients during the first month of treatment. However, notwithstanding the increased support required, feasibility of OBOT-B care to homeless patients was demonstrated and outcomes appeared comparable to a housed group.

These findings conflict with previous reports of poorer addiction treatment outcomes in homeless populations,[27,28,32] but are consistent with studies in which homeless-specific interventions improved addiction outcomes. Milby et al. randomized 176 homeless substance abusers to either usual or enhanced care. Enhanced care entailed daily psychoeducational groups and weekly individual counseling, plus abstinence-contingent employment and housing starting at month 4 of treatment. Subjects receiving enhanced care had significantly lower rates of alcohol and drug use during the 12 months of follow-up.[31] Kertesz et al. found an association between post-detoxification stabilization programs and time to substance use post-discharge from a residential detoxification program for homeless persons with addictions.[30]

This study provides further evidence that addiction treatment in homeless populations can yield effective results. It suggests that more than the usual resources, such as availability of a NCM, may be required for this population when pursuing substance abuse treatment.

The availability of a NCM also allowed us to implement a home-based medication induction protocol, which is not currently recommended by national guidelines, through daily phone contact between the NCM and housed patients.

Several limitations to this study should be considered. While data were collected prospectively during clinical care using a carefully designed medical record, the study was retrospective. In addition, follow-up data were not available once patients departed the program. However, of patients who left the program earlier than 12 months, the reasons for leaving (i.e., treatment failure vs successful program departure) appeared similar between homeless and housed groups. To assess the value of an OBOT-B program in homeless persons, one ideally would compare outcomes with homeless persons requesting substance abuse treatment who did not get access to buprenorphine. However, this study was not intended to retest the efficacy of buprenorphine treatment, but rather, evaluate the feasibility of delivering this known effective treatment in a homeless population. Thus, the comparison group in this study was non-homeless persons receiving buprenorphine. An experienced, skilled NCM played an essential role in caring for patients in the programs described. Generalizability of such a model may depend on skills of such a key individual. Generalizability may also depend on the level of collaboration among practitioners within the microsystem in which the care is delivered.[35] The provider collaboration component of the OBOT-B program was not explicitly measured in this study. Finally, although the total number of patients in this study is small, a consequence of the federal law limiting the number of patients receiving buprenorphine in a single clinical practice, the findings are robust and statistically significant.

In conclusion, office-based opioid treatment with buprenorphine (OBOT-B) can be effective in homeless patients. In this study, despite homeless patients' greater comorbidities, less social support, and more chronic substance use histories, their outcomes in terms of treatment failure and illicit drug use were comparable to housed patients. Social benefits, specifically gaining access to housing and employment, occurred in a

surprisingly high percentage of the homeless OBOT-B patients. Beyond individual clinical benefit, potentially unmeasured public health benefits regarding risk behaviors and health care utilization need to be further assessed. Using appropriate models of care, buprenorphine treatment for opioid dependence among homeless persons should be implemented.

Acknowledgements: *Preliminary results of this study were presented at the 28th annual meeting of the Society of General Internal Medicine in New Orleans on May 11–14, 2005, the 67th annual national meeting of the College on Problems of Drug Dependence in Orlando, FL on June 18–23, 2005, and at the 29th annual meeting of the Association for Medical Education and Research in Substance Abuse in Bethesda, MD on October 27–29, 2005. Data management was provided by Michael Winter MPH and Jacqueline Ashba MA, MPH, at the Data Coordinating Center, Boston University School of Public Health, Boston, MA. Support for these programs and their evaluation was provided by the Massachusetts Department of Public Health: Bureau of Substance Abuse Services and HIV/AIDS Bureau.*

Potential Financial Conflicts 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, 91 E. Concord Street, Suite 200, Boston, MA 02118, USA (e-mail: dalford@bu.edu).*

REFERENCES

1. US Department of Health and Human Services. Substance Abuse and Mental Health Services Administration (SAMHSA). Results from the 2004 National Survey on Drug Use & Health. Available at http://oas.samhsa.gov/NSDUH/2k4NSDUH/2k4results/2k4results.htm#ch2. Accessed July 11, 2006.
2. US Department of Health and Human Services. Substance Abuse and Mental Health Services Administration (SAMHSA). The DASIS Report: Non-Heroin Opiate Admissions, 2003. Results from the Drug and Alcohol Services Information System (DASIS). 2003. Available at http://oas.samhsa.gov/2k6/opiatesTX/opiatesTX.cfm. Accessed July 11, 2006.
3. National Consensus Development Panel on Effective Medical Treatment of Opiate Addiction. Effective medical treatment of opiate addiction. *JAMA.* 1998;280:1936–43.
4. **Johnson RE, Jaffe JH, and Fudala PJ.** A controlled trial of buprenorphine treatment for opioid dependence. *JAMA.* 1992;267:2750–55.
5. **Ling W, Wesson DR, Charuvastra C, and Klett CJ.** A controlled trial comparing buprenorphine and methadone maintenance in opioid dependence. *Arch Gen Psychiatry.* 1996;53:401–7.
6. **Ball JC and Ross A.** The Effectiveness of Methadone Maintenance Treatment. New York, NY: Springer—Berlin Heidelberg New York; 1991.
7. **Joseph H, Stancliff S, and Langrod J.** Methadone maintenance treatment (MMT): a review of historical and clinical issues. *Mt Sinai J Med.* 2000;67:347–64.
8. **O'Connor PG and Fiellin DA.** Pharmacologic treatment of heroin-dependent patients. *Ann Intern Med.* 2000;133:40–54.
9. Institute of Medicine. *Federal regulation of methadone treatment.* Washington, DC: National Academy of Medicine; 1995.
10. **Johnson RE, Chutuape MA, Strain EC, Walsh SL, Stitzer ML, and Bigelow GE.** A comparison of levomethadyl acetate, buprenorphine, and methadone for opioid dependence. *N Engl J Med.* 2000;343:1290–7.
11. **Fudala PJ, Bridge TP, Herbert S, et al.** Office-based treatment of opiate addiction with a sublingual-tablet formulation of buprenorphine and naloxone. *N Engl J Med.* 2003;349:949–58.
12. **Kakko J, Svanborg KD, Kreek MJ, and Heilig M.** 1-year retention and social function after buprenorphine-assisted relapse prevention treatment for heroin dependence in Sweden: a randomised, placebo-controlled trial. *Lancet.* 2003;361:662–8.
13. **Fiellin DA and O'Connor PG.** New federal initiatives to enhance the medical treatment of opioid dependence. *Ann Intern Med.* 2002;137:688–92.
14. **Stein MD, Cioe P, and Friedmann PD.** Buprenorphine retention in primary care. *J Gen Intern Med.* 2005;20:1038–41.
15. **Alford DP, Saitz R, LaBelle CT, and Samet JH.** Buprenorphine initiation and maintenance in primary care: a successful interdisciplinary approach. *J Gen Intern Med.* 2004;19(suppl):103.
16. Center for Substance Abuse Treatment. Clinical Guidelines for the Use of Buprenorphine in the Treatment of Opioid Addiction, Treatment Improvement Protocol (TIP) Series 40. DHHS Publication No. (SMA) 04-3939. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2004.
17. Center for Substance Abuse Treatment. Medication-Assisted Treatment for Opioid Addiction in Opioid Treatment Programs, Treatment Improvement Protocol (TIP) Series 43. DHHS Publication No. (SMA) 05-4048. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2005.
18. **Burt MR, Aron LY, Douglas T, Valente J, Lee E, and Iwen B.** Homelessness: programs and the people they serve. Technical report of findings of the National Survey of Homeless Assistance Providers and Clients. 1999. Interagency Council on the Homeless.
19. **Fischer PJ and Breakey WR.** The epidemiology of alcohol, drug, and mental disorders among homeless persons. *Am J Psychol.* 1991; 46:1115–28.
20. **Stahler GJ and Cohen E.** Homelessness and substance abuse in the 1990s. *Contemp Drug Probl.* 1995;22:169–92.
21. **Koegel P, Sullivan G, Burnam A, Morton SC, and Wenzel S.** Utilization of mental health and substance abuse services among homeless adults in Los Angeles. *Med Care.* 1999;37:306–17.
22. **Hibbs JR, Benner L, Klugman L, et al.** Mortality in a cohort of homeless adults in Philadelphia. *N Engl J Med.* 1994;331:304–9.
23. **Hwang SW, Lebow JM, Bierer MF, O'Connell JJ, Orav J, and Brennan TA.** Risk factors for death in homeless adults in Boston. *Arch Int Med.* 1998;158:1454–60.
24. **Wright JD and Weber E.** Homelessness and health. New York, NY: McGraw-Hill, Inc; 1987.
25. **Seal KH, Kral AH, Gee L, et al.** Predictors and prevention of nonfatal overdose among street-recruited injection heroin users in the San Francisco Bay Area, 1998–1999. *Am J Public Health.* 2001;91:1842–6.
26. **Blanchon T, Boissonnas A, Vareseon I, and Vidal-Trecan G.** Homelessness and high-dosage buprenorphine misuse. *Subst Use Misuse.* 2003;38:429–42.
27. **Thomas L, Kelly M, and Cousineau M.** Alcoholism and substance abuse. In: Brickner PW, Scharer LK, Conanan BA, Savarese M, and Scanlan BC, eds. Under the safety net: The health and social welfare of the homeless in the United States. New York, NY: W.W. Norton & Co.; 1990;204–14.
28. **Song JY, Safaeian M, Strathdee SA, Vlahov D, and Celentano DD.** The prevalence of homelessness among injection drug users with and without HIV infection. *J Urban Health.* 2000;77:678–87.
29. **Galanter M, Dermatis H, Glickman L, et al.** Network therapy: decreased secondary opioid use during buprenorphine maintenance. *J Subst Abuse Treat.* 2004;26:313–8.
30. **Kertesz SG, Horton NJ, Friedmann PD, Saitz R, and Samet JH.** Slowing the revolving door: stabilization programs reduce homeless persons' substance use after detoxification. *J Subst Abuse Treat.* 2003;24:197–207.
31. **Milby JB, Schumacher JE, Raczynski JM, et al.** Sufficient conditions for effective treatment of substance abusing homeless persons. *Drug Alcohol Depend.* 1996;43:39–47.
32. **Shah NG, Celentano DD, Vlahov D, et al.** Correlates of enrollment in methadone maintenance treatment programs differ by HIV-serostatus. *AIDS.* 2000;14:2035–43.
33. **Mistral W and Velleman R.** Are practice nurses an underused resource for managing patients having problems with illicit drugs? *J Subst Use.* 1999;4:82–7.
34. SAS Institute Inc. SAS/STAT Software: Changes and Enhancements, Release 8.2. Cary, NC: SAS Institute Inc.; 2001.
35. **Berwick DM.** A user's manual for the IOM's 'quality chasm' report. *Health Aff.* 2002;21:80–90.

Clinical Case Discussion: Treating Opioid Dependence with Buprenorphine

Daniel P. Alford, MD, MPH, Edwin A. Salsitz, MD, FASAM, Judith Martin, MD, John A. Renner, MD

Key Words: opioid, opioid dependence, buprenorphine

(*J Addict Med* 2007;1: 73–78)

This clinical case conference discusses a woman who presented with opioid dependence and requested buprenorphine treatment in an office-based setting. Three expert clinicians shared their thoughts about the patient's assessment and management.

CASE DESCRIPTION

Monday

You received a frantic phone call from a nurse colleague because her 26-year-old daughter, Susan, was addicted to heroin and relapsed immediately after being discharged from a 5-day inpatient detoxification program. Susan was eager to try treatment with buprenorphine but was unable to find a physician accepting new patients. You arranged for Susan to see you in your office the following day.

Tuesday

Susan's opioid abuse started in high school when she was 17 years old with weekend use of oxycodone. After a few months, she began using daily when she realized that on days without oxycodone she would become sick. After 1 year, she switched to intranasal heroin because it was much cheaper and more readily available. Within 6 months, she and her closest friend began injecting heroin. Her longest drug-free time was 2 months, approximately 3 years ago, when she moved to live with her father. She attributed her sobriety to being away from her drug-using friends. After a fight with her father, she moved back to live with her mother. She started using heroin immediately and had been using daily since that time. Although she shared needles in the past, she was currently enrolled in a needle exchange program. Within the past 2 years, she had completed 8 detoxification programs and always relapsed within 1 to 2 days after discharge. She stopped attending Narcotics Anonymous (NA) meetings because hearing other's "war stories" was not helpful. She had considered methadone maintenance treatment but was worried that methadone "eats your bones and teeth, and makes you fat." She also feared that her work hours as a hotel clerk would conflict with the dosing hours of methadone clinics. Moreover, her mother was "antimethadone" as a result of seeing methadone patients loitering around the hospital "selling drugs and nodding off." Susan had tried snorting cocaine many years ago but did not like the way she felt. She had taken Klonopin, obtained from friends, in the past to help with "anxiety" and trouble sleeping. She smoked cigarettes and denied other drug or alcohol use. She had no primary health care and had not seen a psychiatrist.

Susan dropped out of community college after one semester and had held various clerical jobs. She was living with her mother and had a boyfriend who was in recovery from alcoholism and who attended Alcoholics Anonymous (AA) daily.

Susan's physical examination was normal except for track marks on both arms. Her laboratory tests were positive for hepatitis B exposure and immunity and chronically active hepatitis C (genotype 1) with normal liver function tests. She was negative for HIV infection. Susan was not interested in hepatitis C treatment because she felt fine and knew friends who got sick while taking interferon. Her urine toxicology was positive for opiates and benzodiazepines. She had last used heroin and Klonopin the previous night.

You spend 20 minutes educating Susan about buprenorphine treatment, the need for concurrent psychosocial treatment (counseling and/or 12-step meetings), and the importance of avoiding illicit benzodiazepines. Susan left your office with a prescription for a 3-day supply of Suboxone (buprenorphine/naloxone) and was instructed to return the following day with all her tablets and in mild-to-moderate opioid withdrawal. She agreed to reconnect with her previous addiction counselor.

Wednesday

Susan returned at 8 AM feeling "dope sick" and was extremely anxious. She last used heroin 12 hours ago and was yawning, had runny nose and eyes, and dilated pupils. She

From the Department of Medicine (DPA), Boston University School of Medicine, Office Based Opioid Treatment Program, Boston Medical Center, Boston, MA; Office-Based Opioid Therapy, Department of Medicine (ES), Beth Israel Medical Center, New York, NY; The 14th Street Clinic and East Bay Community Recovery Project (JM), Oakland, CA; and Department of Psychiatry (JAR), Boston University School of Medicine, VA Boston Healthcare System, Boston, MA.

Received April 3, 2007; accepted April 4, 2007.

Send correspondence and reprint requests to Dr. Daniel P. Alford, Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, 91 East Concord Street, Suite 200, Boston, MA 02118. e-mail: dan.alford@bmc.org

ISSN: 1921-0629/07/0102-0073

took her first dose of 4 mg sublingual of Suboxone and felt better within 30 minutes. After 2 hours of observation, she felt fine and was instructed to take an additional 4 mg later in the day if needed. She was given a daily dosing schedule for the next 2 days and told to return on Friday.

Friday

Susan returned stating that she felt "great and completely normal" while taking 16 mg. For the first time in 2 years she was no longer thinking about using drugs. She agreed to continue taking 16 mg daily, to make an appointment with her previous addiction counselor, and to start attending NA meetings.

Next Few Weeks

You saw Susan twice and she continued to do well. She denied withdrawal symptoms or drug craving, and she said that she had hung up on her drug dealer. She had changed her phone number to stop her drug-using friends from calling her. She was thinking about going back to school the next semester. Susan's mother called to thank you for "saving" her daughter's life. Susan had not called her addiction counselor and had not started attending 12-step meetings but promised to do so before her next appointment. She repeatedly exclaimed that Suboxone was a "miracle drug." She offered to speak at physician trainings if it would increase the number of doctors offering buprenorphine treatment.

Next Few Months

Susan continued to remain compliant with her nursing and medical appointments and maintained abstinence from illicit drugs as confirmed by negative urine drug tests; however, she had not started going to counseling or 12-step meetings.

During the second month, Susan arrived 1 hour late for her appointment and complained of increased insomnia, headaches, and anxiety. She had spoken to her former addiction counselor who attributed her symptoms to the naloxone in Suboxone and suggested she ask for Subutex, the "buprenorphine-only" tablet. Her counselor emphasized the importance of "getting off" buprenorphine as soon as possible because it was only a "temporary crutch" that she should not need. Susan also requested an increase in her dose to 24 mg to help her sleep at night. Her urine was positive for opiates and benzodiazepines and negative for buprenorphine. She had stopped her Suboxone for a few days to party "one last time" with friends. She felt guilty about her "slip" and was now "back on track." She easily restarted Suboxone on her own. She was still not attending 12-step meetings because "they are against buprenorphine." She had broken up with her boyfriend after he relapsed and was actively drinking. You recommended continuing Suboxone 16 mg, emphasized the importance of attending addiction counseling with someone familiar with buprenorphine treatment, and required a psychiatric evaluation of her anxiety symptoms. The following day Susan's mother called because she was concerned that Susan was starting to associate with her drug-using friends and wanted to know the results of her drug tests.

DISCUSSION

Edwin Salsitz, MD

This case represents a common clinical presentation of the disease of opioid addiction. Her initial illicit opioid exposure occurred at age 17 years with the prescription opioid, oxycodone. Unfortunately there has been a dramatic rise in prescription opioid abuse among teenagers. The drugs are obtained from medicine cabinets, friends, dealers, and often the Internet. After taking oxycodone for a few months, she developed physical dependence, switched to heroin intranasal, and then intravenous use. The intravenous use exposed her to Hepatitis B and C, but not HIV.

We are not given information about Susan's life experiences before age 17 years. A life narrative is helpful to assess individual vulnerability to addiction. A genetic predisposition to addiction is common and is considered to account for 50% of the vulnerability. Especially in women, sexual trauma as an environmental risk factor is common. In some studies of addicted women, up to 80% report a history of sexual trauma/violence. Of those, a significant number develop posttraumatic stress and anxiety disorder. Susan seems to have an anxiety disorder, which had not been formally addressed. In addition, patients with a history of anxiety or depression are more likely to be placed on prescription opioids.[1] It may not be possible to obtain such sensitive, traumatic history on day 1, but on follow-up visits these issues should be explored.

She had a 2-month period of abstinence from opioids. I would ask her how she felt without opioids: Was she anxious? Did she function well? The answers to these questions may provide important clues as to the need for, and benefit from, opioid agonist therapy.

Like many patients, with opioid addiction, Susan has had multiple episodes of withdrawal treatment (detoxification). Unfortunately, as has been often documented, relapse is the rule, rather than the exception.[2] She was an excellent candidate for office-based opioid therapy—had family support, had a job, and was motivated. Her logistic concerns about attending a methadone clinic are reasonable. Persistent myths about methadone destroying teeth and bones unfortunately confuse many potential patients. The stigma surrounding methadone treatment is tragic, but real. However, Susan was a good candidate for sublingual buprenorphine/naloxone in an office-based setting.

The buprenorphine induction and early stabilization followed published recommendations[3] and went smoothly. Importantly she changed some of her behaviors (hung up telephone on drug dealer, changed friends, making plans to get back to school). Opioid agonist therapy, when successful, enables a patient to rehabilitate their lives. She found the drug to be a "miracle" but was reluctant to access psychosocial treatment. Many patients think that the medication alone can "cure" their problems; however, it is well documented that medication in conjunction with psychosocial therapy produces the best outcomes.[4] Patients with opioid addiction may have social phobia, making participation in mutual help groups difficult. Some mutual help groups are opposed to medication-assisted treatment; however, buprenorphine

anonymous groups have developed, which will fill a treatment gap. Some patients prefer and do well with individual psychotherapy.

Unfortunately, but not uncommonly, Susan's addiction counselor emphasized the importance of "getting" off the buprenorphine rapidly and used the term "crutch" in referring to buprenorphine treatment. Although these are widely held beliefs, there is no scientific support for their validity. The evidence clearly shows that the longer a patient remains on opioid agonist therapy, the better they do in all parameters of their lives. Neuroimaging studies show normalization and stabilization of key neurobiological functions in those patients who remain on opioid agonist therapy.[5] Therefore, it was predictable that when Susan stopped taking buprenorphine, she relapsed. Because addiction, like all chronic diseases, is a relapsing, progressive, remitting, incurable disorder, this should not be surprising nor viewed as a reason to stop treatment. It usually takes some time to achieve remission.

The current recommendation is to use Suboxone, except in pregnancy. At times there is an indication to switch from Suboxone to Subutex. Patients who develop headaches, cannot tolerate the taste of Suboxone, experience nausea after Suboxone, or develop a rash can be considered for a switch to Subutex. Ideally these patients would be stable with low risk for diversion or misuse of the Subutex. In Susan's case, switching to Subutex was not indicated, because she had stopped taking Suboxone and relapsed.

She has acquired chronic hepatitis C infection, most likely from injection drug use. Most injection drug users who share needles become infected with hepatitis C. Needle exchange programs have proven to be effective in reducing transmission of infectious diseases, without increasing rates of injection drug use.[6] Because her liver enzymes are normal, she will most likely have minimal disease on her liver biopsy. Genotype 1 is the genotype with the lowest response rate to currently available treatment with interferon and ribavirin (50% chance of sustained viral response). Because treatment has significant side effects, it is probably prudent to delay treatment for now while Susan deals with consolidating her stability with the addictive disease.

Although Susan's mother is your colleague, and a health care provider, unless you have gotten specific written consent to discuss Susan's case with her, you are prohibited from doing so. These are issues that should be discussed and clarified early in treatment. For example, you will want to have an ongoing discussion with her psychiatrist and counselor. You must obtain permission for each individual with whom you will communicate. These rules are governed under CFR 42.

In summary, Susan represents a typical patient with the disease of opioid addiction. Preaddiction history may provide clues about her vulnerability to opioid addiction. Abstinence-based therapies were not effective, and she claimed that she felt "normal" while taking buprenorphine. She is an excellent candidate for buprenorphine maintenance treatment, along with effective psychosocial treatment. The buprenorphine treatment may be efficacious for her anxiety. She should use the stability provided by the buprenorphine to further her education, vocational status, underlying psychologic issues, and personal life. There should be no rush or goal to discontinue buprenorphine maintenance.

Judith Martin, MD

This is not an uncommon presentation for young adult patients. Distressed parents or relatives call, and they often bring (sometimes drag) the patient to the doctor's office. In such cases, the initial visit often will take a long time, because there will be more than one person present who will want time with you. This is an unusually flexible scheduling feat; many places would not have such a next-day appointment available. This addiction history has many familiar features. Although the patient is young, she has used daily for 9 years. If relapse to drug use is used as the measure of success or failure, there have already been many "failed" detoxification attempts. It may be helpful to review these repeated relapses with Susan to gain some therapeutically useful clues. Does she see these treatment episodes as personal failures or as a deficiency in the medical system? And to come up with a treatment plan for Susan, we need to hear more about what she wants to achieve in treatment. Is it relief of withdrawal and craving? Does she aspire to more career development or degrees? Does she want to marry her boyfriend and have children? Maybe she just wants to get her mother off her back. What does she feel she has accomplished so far?

The history of escalation of risk from oral ingestion of prescription opioids to intranasal heroin and eventually to injected heroin is common, especially in cities where heroin is cheap and plentiful. Susan's history of repeated relapse after inpatient care might be a relative contraindication to office-based treatment with buprenorphine, but it is worth a try if such care has never included maintenance pharmacotherapy. The daily medication might stabilize her opioid use to the point that she can sustain ongoing engagement in treatment, and outpatient counseling and office visits could be successful.

Psychiatric co-occurring disorders may contribute to relapse. When addicted patients present psychiatric symptoms it may be difficult to distinguish the primary condition. Anxiety in the context of repeated opioid withdrawal episodes is difficult to distinguish from craving or opioid deficit. The history does not clarify details about the anxiety or whether there is a family history of addiction or psychiatric problems. The approach of treating the withdrawal and craving for opioids first and re-evaluating the anxiety if still present later seems reasonable and in this case seemed to support a withdrawal-mediated source for her anxiety. The patient's well-being, which resulted during and after induction to sublingual buprenorphine, suggests that the anxiety was opioid withdrawal. Patients often find it hard to believe that their cravings are controlled with medication that does not alter their consciousness. It is not unusual to hear the word "miracle" in medication-assisted treatment. However, patients often focus on drug use alone and remain unaware of life changes that are needed to maintain the gains that they have made in treatment. Susan had begun by changing some of her social life, and she said that she was

making plans to integrate ongoing recovery support into her life.

Some of the details of the case suggest that the counselor does not understand medication-assisted treatment and may have been undermining Susan's gains, calling it a "crutch." It may be useful to have regular contact between physician and counselor to discuss Susan's progress and to be sure the treatment plan reflects Susan's purpose. Although there is no requirement for counselor training in office-based buprenorphine treatment, there is a CSAT online course available at http://www.danyalearningcenter.org/courseprofile.asp?cid=7.[7]

It is not clear whether the boyfriend, described as a recovering and then relapsed alcoholic, is a support for Susan's recovery. Being "recovery savvy" may help, but on the other hand there could be some judgment against the use of medications. It may benefit both of them to meet with the physician together to explain how opioid pharmacotherapy works.

Use of syringe and needle exchange may have protected her against HIV but not against hepatitis C. Because of the high concentration of virus in the bloodstream, hepatitis C prevention includes not sharing cookers, rinse water, cotton, in addition to not sharing the needle and syringe. Anecdotal reports suggest that nasal mucosal cuts or tears from shared straws used to inhale drugs also may transmit hepatitis C. The fact that her liver enzymes are normal does not reveal the stage of this illness. There is a lot of information with which Susan needs to choose what to do. She already has heard of interferon and fears the side effects. Does she also know that not everyone needs treatment and that interferon and ribavirin treatment frequently leads to sustained remission? Does she know that even small amounts of alcohol can contribute to disease progression? She is already immune to hepatitis B but could consider hepatitis A vaccination as a protective measure. Because adherence to treatment of 1 year is recommended for genotype 1, we would expect that stabilizing her drug abuse behavior would be a good move, and enhance her ability to undergo treatment if necessary. Treatment of hepatitis could be beneficial even if she continues to use illicit substances. The decision to treat hepatitis C takes many factors into account, and ideally includes a liver biopsy to establish the level of disease progression.

Stigma and misunderstanding about methadone treatment are common among patients, as well as among medical and nursing colleagues, in fact even in addiction treatment programs in which they are euphemistically couched in "philosophy of care" jargon. Typically, methadone maintenance treatment (MMT) is offered in an outpatient clinic setting that includes daily observed doses of medication, mandated counseling, and random drug testing. This treatment has been repeatedly shown to save lives, reduce seroconversion to HIV, reduce illicit drug use, and increase "socially productive activity."[8] A significant portion of the reduced mortality is reduction in opioid overdoses. Most clinics have resources available to address medical and psychiatric comorbidities. There is limited ability to use evidence-based patient placement criteria, such as the ASAM Patient Placement Criteria, in methadone maintenance for several reasons. First, the treatment is heavily regulated to control diversion, so daily visits are mandatory for many months. This schedule may make the treatment less accessible or desirable for working patients, or for those with home or school obligations, or for persons who live far from the clinic site. Second, even if placement criteria requires a higher level of care, such as inpatient, or residential, or intensive hours of outpatient care for the more severely affected patients, referrals to such placement may be hampered by "no methadone patients" policy of such programs. For example, should a patient need inpatient care for detoxification from alcohol or benzodiazepines, they might be refused admission because they are enrolled in methadone treatment. Susan's mother's experience as a nurse of seeing MMT patients "hanging around and nodding off" illustrates the lack of a continuum of services for MMT patients who may still be using synergistically sedating substances, such as benzodiazepines. Although such patients need a higher level of care, they are forced to choose between the methadone clinic and other treatments, rather than to be able to access treatment according to placement need.

Susan's laundry list of reasons to avoid methadone (bones, teeth, weight) is part of long-standing urban legends that stigmatize methadone.[9] These statements are not supported in the medical literature. The physician's list of side effects of methadone is more likely to include constipation (caused by methadone duration of action compared with short-acting abusable opioids) and the logistical difficulties of being tied to a highly regulated clinic.

John Renner, MD

My initial concern has to do with boundary issues, particularly accepting the child of a colleague into treatment. This would be considered problematic in psychiatry, although the issue is probably less clearly defined in other areas of medicine. I would have considered facilitating referral to another provider but would not have taken the patient into my own practice.

My second concern is proceeding with an appointment when a relative, rather than the patient, has made the initial contact. This often blurs the issue about who is motivated for treatment: the parent or the child? My recommendation would have been to say that the patient must call (me, my secretary, or the nurse) to schedule the intake. This does not solve the problem of minimal motivation, but it sends a message about who has to take responsibility for the recovery process.

Her history and circumstances suggest that she would be a good candidate for buprenorphine. Despite her misinformation and prejudices about methadone, she has been able to function in the community and daily attendance at a methadone clinic might make that more difficult. I would not be particularly concerned about her cocaine and benzodiazepine use, other than suggesting that she see a psychiatrist and/or be screened for an anxiety disorder. Such an evaluation should not be performed until she has stabilized on buprenorphine, at least 3 weeks into treatment.

Given the length of time (9 years) and severity of her addiction (8 failed detoxifications), I would be concerned that she needs more treatment than buprenorphine alone. I would make psychosocial treatment a requirement, rather than just "educating" her on the need for other services. At a minimum, this would include a psychiatric evaluation and NA/AA or individual or group counseling. Hopefully, a therapist skilled in 12-step facilitation could help her work through her resistance to 12-step programs.

Regarding the induction to buprenorphine, I think the initial plan was appropriate. It is not clear whether the Thursday dose of 16 mg was higher than recommended, but I do not think it unreasonable given her history. The exact frequency of initial visits is not stated, and I am concerned that several weeks have gone by without any visits to NA or the counselor. Not monitoring that situation more closely reinforces the idea that this was not an important part of the treatment plan and reinforces her idea that buprenorphine alone is the "miracle." Her excellent initial response to pharmacotherapy is not unusual in these circumstances, but it may obscure the need for other therapy to ensure long-term success. Her offer to help with buprenorphine training courses could set a dangerous precedent. It is too early to be so sure of success; it reinforces the idea that she is "special" and that she has a special relationship with her physician—all of which can be risky. Buying into that idea certainly undercuts your ability to enforce the requirement that she participate in other treatment activities.

Her complaints in the second month are suggestive of problems with anxiety and/or depression, certainly not side effects of naloxone. It is of course impossible to separate these complaints from symptoms related to her recurrent drug use. This case highlights the problem of working with counselors who are not educated about buprenorphine and the need for long-term pharmacotherapy. This is a setup for ongoing problems.

I am comfortable with keeping the dose at 16 mg because she did not report relapsing as a result of craving or withdrawal symptoms. At this point, I would be more insistent about getting the psychiatric evaluation, a new counselor, and returning to NA/AA. I also would increase the frequency of medication visits and drug testing. I would require all of this to continue treatment, explaining again the risks of pharmacotherapy alone—doing it her way did not work. It is always a judgment call with new patients about whether to require additional treatment. It is easy to try medications alone (as I often do); it then becomes harder to require counseling when the patient seems to be doing well. This case points out the risks but also the reality that you may only have the leverage to require added services after there is some type of lapse or relapse.

The mother's call and request for information is clearly a "trap," and violating the patient's confidentiality would be a serious mistake. The initial boundary issue resurfaces and makes it more difficult to manage. The correct response may be to recommend a family meeting, if not family therapy. However, this becomes problematic if one of the family members is a colleague.

Summary

Three experts have shared their thoughts regarding the assessment and management of this opioid-dependent person requesting office-based buprenorphine treatment. There was agreement that this patient represented a common clinical scenario with addiction starting with prescription opioids and progressing to intravenous opioid use. All agreed that the patient would be a good candidate for buprenorphine maintenance because of her opioid-dependence diagnosis, her failed medication-free treatment history, her family supports, and her ability to function in the community. However, Drs. Martin and Renner expressed some concern about her appropriateness for office-based treatment based on her repeated relapse after inpatient care and an unclear level of motivation (mother making the initial phone call). Dr. Martin emphasized the need for understanding the patient's treatment goals to better understand her treatment motivation.

Although the patient was actively using nonprescribed benzodiazepines, none of the experts considered this a contraindication to buprenorphine treatment. Because of overdoses in Europe with concomitant misuse of intravenous buprenorphine and high potency benzodiazepines, guidelines have recommend not prescribing buprenorphine to patients with uncontrolled benzodiazepine use. However, although most clinicians would consider concurrent benzodiazepine abuse a relative and not absolute contraindication, all would treat cautiously and try to avoid prescribing Subutex to such patients.

All suggested a need for additional clinical assessment to obtain potentially "therapeutically useful clues," by better understanding the patient's predisposition to addiction, previous experiences during detoxification, and periods of abstinence. All agreed for the need, albeit difficult, to evaluate whether the patient's anxiety was drug-induced or a primary disorder. The consensus was reassessment of her anxiety after being stabilized on buprenorphine. Dr. Salsitz noted that her anxiety may compromise her ability to engage in group or 12-step facilitated treatment and that she may require individual counseling.

All 3 experts addressed the unfortunate, but all too common, realities of methadone maintenance stigma and misinformation by patients, the medical community, and sadly, the substance abuse treatment field. The counselor's misinformation about buprenorphine treatment undermined the patient's care. With more than 40 years of literature describing high rates of relapse after detoxification and improved outcomes with maintenance medication, the suggestion of medication discontinuation was clinically inappropriate. All agreed that a switch from Suboxone (buprenorphine/naloxone) to Subutex (buprenorphine alone) was not indicated. Suboxone was designed as a more diversion-proof medication, whereas Subutex, is vulnerable to abuse.[10]

The importance of concurrent psychosocial treatment was addressed by all 3 experts. In fact, Dr. Renner would have required psychosocial treatment as a condition for starting this patient on buprenorphine. The patient's perceived "cure" (freedom of withdrawal and craving symptoms) likely lead to decreased motivation to engage in psychosocial treatment. In my experience, this "honeymoon period" of bu-

prenorphine treatment often is followed by boredom (eg, no friends, no job, no school), an abundance of free time, and increased risk for relapse.

Both Drs. Salsitz and Martin agreed that the patient's decision to delay hepatitis C evaluation and treatment was appropriate during early recovery. However, Dr. Martin discussed the need for patient education (eg, alcohol avoidance) and immunization against hepatitis A and B. Finally, the issue of the mother's involvement in her daughter's care was interesting. Although the nonpsychiatrist experts (Drs. Salsitz and Martin) did not express concern for treating a colleague's child, the psychiatrist expert (Dr. Renner) did. All the experts emphasized the importance of not violating the patient's confidentiality by disclosing information to the mother (eg, drug test results) and suggest that a family meeting may be beneficial.

REFERENCES

1. Sullivan MD, Edlund MJ, Zhang L, et al. Association between mental health disorders, problem drug use, and regular prescription opioid use. *Arch Intern Med.* 2006;166:2097–2093.
2. O'Connor PG. Methods of detoxification and their role in treating patients with opioid dependence. *JAMA.* 2005;294:961–963.
3. Center for Substance Abuse Treatment. Clinical Guidelines for the Use of Buprenorphine in the Treatment of Opioid Addiction. Treatment Improvement Protocol (TIP) Series 40. DHHS Publication No. (SMA) 04-3939. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2004.
4. McLellan AT, Arndt IO, Metzger DS, et al. The effects of psychosocial services in substance abuse treatment. *JAMA.* 1993;269:1953–1959.
5. Kaufman MJ, Pollack MH, Villafuerte RA, et al. Cerebral phosphorus metabolite abnormalities in opiate-dependent polydrug abusers in methadone maintenance. *Psychiatry Res.* 1999;90:143–152.
6. Watters JK, Estilo MJ, Clark EL, et al. Syringe and needle exchange as HIV/AIDS prevention for injection drug users. *JAMA.* 1994;271.
7. CSAT Online Course. Available at: http://www.danyalearningcenter-.org/courseprofile.asp?cid=7. Accessed April 2, 2007.
8. Kuehn BM. Methadone treatment marks 40 years. *JAMA.* 2005;294: 887–888.
9. Hunt DE, Lipton DS, Goldsmith DS, et al. "It takes your heart": the image of methadone maintenance in the addict world and its effect on recruitment into treatment. *Int J Addict.* 1985;20:1751–1771.
10. Parfitt T. Designer drug Subutex takes it toll in Tbilisi. *Lancet.* 2006; 368:273–274.

Patients Derogate Physicians Who Use a Computer-Assisted Diagnostic Aid

Hal R. Arkes, PhD, Victoria A. Shaffer, PhD, Mitchell A. Medow, PhD, MD

Objective. *To ascertain whether a physician who uses a computer-assisted diagnostic support system (DSS) would be rated less capable than a physician who does not.* ***Method.*** *Students assumed the role of a patient with a possible ankle fracture (experiment 1) or a possible deep vein thrombosis (experiment 2). They read a scenario that described an interaction with a physician who used no DSS, one who used an unspecified DSS, or one who used a DSS developed at a prestigious medical center. Participants were then asked to rate the interaction on 5 criteria, the most important of which was the diagnostic ability of the physician. In experiment 3, 74 patients in the waiting room of a clinic were randomly assigned to the same 3 types of groups as used in experiment 1. In experiment 4, 131 3rd- and 4th-year medical students read a scenario of a physician-patient interaction and were randomly assigned to 1 of 4 groups: the physician used no DSS, heeded the recommendation of a DSS, defied a recommendation of a DSS by treating in a less aggressive manner, or defied a recommendation of a DSS by treating in a more aggressive manner.* ***Results.*** *The participants always deemed the physician who used no decision aid to have the highest diagnostic ability.* ***Conclusion.*** *Patients may surmise that a physician who uses a DSS is not as capable as a physician who makes the diagnosis with no assistance from a DSS.* ***Key words:*** *decision support techniques; diagnosis computer assisted; patient satisfaction.* ***(Med Decis Making 2007;27:189–202)***

A large number of computer-based diagnostic support systems (DSSs) have been developed during the past 30 years.[1] DSSs have been heralded as offering a way to decrease various types of errors,[2] foster the implementation of evidence-based medicine,[3] and reduce inappropriate admissions and costs.[4]

There are strong a priori reasons for believing that computer-based DSSs would improve diagnostic accuracy. Dawes and his colleagues[5] reviewed approximately 100 studies, the overwhelming majority of which showed that actuarial predictions were superior to the unaided diagnostician. Actuarial judgments are characterized by combining evidence using a formula such as a regression equation. Of course, combining pieces of evidence in such a manner is easily done with a simple computer program. As a prototypic actuarial method, a computer-based DSS would thus enjoy an advantage over unaided diagnosticians, consistent with the evidence summarized by Dawes and colleagues.[5]

The research reviewed by Dawes and colleagues[5] pertained to a wide variety of domains with a substantial number emanating from the domain of psychology. However, specifically within the domain of medicine, a large number of studies have found that computer-based DSSs perform better than physicians in a wide variety of diagnostic contexts.[6–11] In an early study, de Dombal and colleagues[7] showed that the use of a computer program resulted in significantly more accurate diagnoses of acute appendicitis. Similarly, Corey and Merenstein[6] showed that use of a predictive index for acute cardiac ischemia resulted in far more accurate classification of patients than occurred when

Received 10 February 2005 from the Center for Health Outcomes, Policy, & Evaluation Studies and the Department of Psychology, The Ohio State University, Columbus, Ohio (HRA); the Department of Psychology, Wichita State University, Wichita, Kansas (VAS); and the Division of General Internal Medicine, College of Medicine, The Ohio State University, Columbus, Ohio (MAM). Financial support for this study was provided by a grant from the Program in Decision, Risk, and Management Science at the National Science Foundation. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. Revision accepted for publication 6 September 2006.

Address correspondence to Hal R. Arkes, PhD, Center for Health Outcomes, Policy, and Evaluation Studies, The Ohio State University, 466 Cunz Hall, 1841 Millikin Road, Columbus, OH 43210-1229; e-mail: arkes.1@osu.edu.

DOI: 10.1177/0272989X06297391

physicians did not use the index. Ridderikhoff and van Herk[8] found that a DSS used in a general practice more than doubled the diagnostic accuracy of unaided physicians. In one of the experiments to be presented in this article, we presented respondents with a scenario describing a sore throat. Several prediction rules have been developed to help physicians diagnose strep throat.[12–14] Another of our scenarios pertains to a possible ankle fracture. A well-known scoring rule has been developed in this domain as well.[15]

Reviews by Kaplan[1] and Hunt and colleagues[16] confirm that although many studies verify the superiority of DSSs in the diagnostic process, some studies do not.[17] However, there is unanimity with regard to 1 characteristic of DSSs: they are grossly underused.[1] To cite 1 example of underutilization, the acute ischemic heart disease predictive instrument put in place by Corey and Merenstein[6] reduced the false-positive diagnosis rate from 71% to 0%. Following the use of the aid in randomized controlled trials, physicians were free to use the aid or not. Utilization during this latter phase was only 2.8%. Other examples of underutilization abound.[18–22]

A number of reasons have been proposed to explain why diagnosticians might not want to use computer-assisted DSSs or any other actuarial system such as a practice guideline.[20,23,24] These reasons include the reluctance to use "cookbook medicine," inertia of previous practice, and lack of agreement with the output of the DSS.[19,20] One additional and potentially important reason why physicians might be reluctant to use computer-based DSSs was suggested in a short report by Cruikshank.[25] Using only 3 British physicians in his study, Cruikshank[25] assessed patients' opinions of their physicians before and after a computer-assisted diagnostic tool was introduced. Even though such a decision aid was designed to reduce diagnostic errors, patients were markedly less positive about the thoroughness, cleverness, decisiveness, and thoughtfulness of their physician after the decision aid was introduced. A physician who values his or her reputation may be understandably reluctant to use a computer-based DSS if its use has this detrimental effect. However, this study was published in 1984, when the use of computers was drastically less prevalent than it is today. Therefore, the wariness patients might have had concerning the use of a novel machine might not be present in today's computer-saturated environment. Note that we are not investigating patients' attitudes toward physicians' use of computerized medical records, which have generally not been negative.[26,27] Instead, we are specifically interested in whether patients would derogate the diagnostic ability of a physician who uses a computer-based DSS compared to the rating given by a patient to a physician who does not use such an aid.

EXPERIMENT 1

Method

Participants

In total, 347 undergraduate students at The Ohio State University participated in the experiment in partial fulfillment of a course requirement. The participants were randomly assigned to 1 of 3 experimental groups: no aid (n = 108), aid (n = 127), and prestigious aid (n = 112).

Materials

The experiment employed 3 short scenarios that described an interaction between a doctor and her patient, a brief questionnaire containing the target-dependent variables, and a numeracy scale that contained 3 questions. The questions were the "General Numeracy Items" from Lipkus and colleagues,[28] which were in turn adapted from Schwartz and colleagues.[29] We hypothesized that facility with numerical concepts might foster greater acceptance of DSSs.

All stimulus materials were presented, and all data were collected via computer.

Procedure

Each participant read a 1-page scenario in which he or she was asked to take the role of the patient. In the scenario, the patient had an ankle injury and went to see a primary care physician, who provided a diagnosis. In the course of the diagnostic process, the doctor used no decision aid, an unspecified decision aid, or a prestigious decision aid. After reading the scenario, the participants were asked to rate the following 5 criteria: thoroughness of examination, length of wait, diagnostic ability of the physician, professionalism of the physician, and overall satisfaction with the examination. Participants were asked to respond to each question separately using 5 different Likert-type scales; each scale ranged from 1 to 7. The more negative evaluations corresponded to the left side of the scale and the more positive evaluations with the right. Each point on the scale was given both a numeric value and a descriptive label. After they responded to the dependent variables, the participants were asked to respond to the numeracy scale (see Appendix A for the scenarios and numeracy scale).

Table 1 Experiment 1 Group Means

Dependent Variable	Group		
	No Aid	Aid	Prestigious Aid
Diagnostic ability	4.69[a]	3.70[b]	3.63[b]
Overall satisfaction	4.71[a]	4.23[b]	4.15[b]
Professionalism	5.20[a]	4.60[b]	4.50[b]
Thoroughness of examination	5.27[a]	4.80[ac]	4.70[bc]
Length of wait	3.17[a]	3.37[a]	3.18[a]

Note: Within a row, means not sharing a common superscript differ significantly by a Kruskal-Wallis test ($P < 0.05$).

Results

A numeracy score was calculated for each participant by tallying the number of correct items from the 3-item scale. The average numeracy score was 1.83.

For all dependent variables except "length of wait," the physician who used no decision aid was given the highest evaluation. See Table 1 for the means of all dependent variables. Significant Pearson correlations existed between the 5 dependent variables, ranging from .302 to .781.

The 5 dependent variables were subjected to a multivariate analysis of covariance (MANCOVA); the covariate was numeracy score, and the single factor was experimental condition: no aid, aid, and prestigious aid. The numeracy score was not a significant covariate for any of the 5 dependent variables. Numeracy data were also collected in the subsequent 3 experiments; however, numeracy was not a significant covariate in any study. Therefore, the numeracy data will not be addressed again in this article.

Significant F values were obtained for the factor of experimental condition in 4 of the 5 dependent variables: thoroughness of examination, diagnostic ability, professionalism, and overall satisfaction. The evaluation of a patient's length of wait did not differ significantly between the 3 experimental groups. However, responses in the no-aid condition were significantly less variable for 4 of the dependent variables (thoroughness of examination, diagnostic ability, professionalism, and overall satisfaction). Therefore, the parametric analysis of variance (ANOVA) was rejected in favor of the nonparametric Kruskal-Wallis test, which provides an omnibus test for differences in ranks. Kruskal-Wallis tests yielded significant differences ($P < 0.05$) in thoroughness of examination, $\chi^2(2) = 6.70$; diagnostic ability, $\chi^2(2) = 33.90$; professionalism, $\chi^2(2) = 15.82$; and overall satisfaction, $\chi^2(2) = 8.86$. Subsequent post hoc comparisons indicated that the no-aid group was rated significantly higher than the prestigious-aid group for all 4 dependent variables ($P < 0.05$). For 3 of these 4 variables, the no-aid group was rated significantly higher than the aid group ($P < 0.05$), with "thoroughness of the examination" not reaching significance. In contrast, the aid and prestigious-aid groups did not differ significantly on any of the 4 dependent variables.

Discussion

The data from our 1st experiment suggest that patients' perceptions of their physicians are influenced by the use of decision aids during routine visits. The participants' evaluation of the physician's diagnostic ability, the thoroughness of the examination, the professionalism of the physician, and their overall evaluation of the physician were affected by our experimental conditions. The physician who did not use a decision aid was rated more favorably on all 4 dimensions by our participants than the physicians who used an unspecified decision aid or a highly prestigious decision aid. Numeracy scores, found to be a factor in other studies about medical judgments,[28,29] did not significantly affect the evaluations given in this experiment.

Based on the results of Cruikshank,[25] we expected group differences on the dependent variable pertaining to the diagnostic ability of the physician. We did not anticipate that the use of a computer-assisted DSS would also influence such factors as the thoroughness of the examination because the groups did not differ in the amount of data collected by the physician or offered by the patient. Nevertheless, group differences did occur on 4 of the 5 factors.

Perhaps our conclusions were specific to the type of injury; maybe participants believed that a potential ankle fracture was such a common occurrence that physicians should be able to treat the injury without the use of a diagnostic aid. Therefore, we decided to replicate our findings with a 2nd, but potentially more serious, medical condition—deep vein thrombosis (DVT). Because DVT is potentially more dangerous than an ankle fracture, we thought that it would be advisable to ascertain if the same general pattern of findings would occur in this new situation. In addition, because the use of a "prestigious" aid did not affect diagnostic ability, only 2 groups (no aid and aid) were used with this population.

EXPERIMENT 2

Method

Participants

In total, 128 undergraduate students at The Ohio State University participated in the experiment in partial fulfillment of a course requirement. The participants were randomly assigned to 1 of 2 experimental groups: no aid ($n = 55$) and aid ($n = 73$).

Materials

The experiment employed 2 short scenarios that described an interaction between a doctor and her patient and a brief questionnaire containing the target dependent variables. All stimulus materials were presented, and all data were collected via computer.

Procedure

Each participant read a 1-page scenario in which he or she was asked to take the role of the patient. In the scenario, the patient had a leg injury and was at risk for DVT. The patient went to see a primary care physician, who provided a diagnosis. In the course of the diagnostic process, the doctor used either no decision aid or an unspecified decision aid (see Appendix B for the text of the scenario). After reading the scenario, the participants were asked to rate the following 6 criteria: perceived difficulty of diagnosis, thoroughness of examination, length of wait, diagnostic ability of the physician, professionalism of the physician, and overall satisfaction with the examination. Participants were asked to respond to each question separately using 6 different Likert-type scales constructed in a manner identical to those in experiment 1.

Results

Responses to diagnostic ability and thoroughness of examination were significantly more variable in the aid condition. Therefore, the appropriate statistical correction in the degrees of freedom was made for t tests involving these variables. In addition, significant positive Pearson correlations existed between 5 of the dependent variables (thoroughness of examination, length of wait, diagnostic ability of the physician, professionalism of the physician, and overall satisfaction with the examination); correlations ranged from .342 to .820. Perceived difficulty of diagnosis was negatively correlated with professionalism ($r = -.202$, $P < 0.05$). See Table 2 for the means of all dependent variables.

Table 2 Experiment 2 Group Means

	Group	
Dependent Variable	**No Aid**	**Aid**
Diagnostic ability[a]	4.91	4.05
Perceived difficulty	3.98	3.88
Overall satisfaction	4.64	4.15
Professionalism[a]	5.44	4.78
Thoroughness of examination[a]	5.55	5.11
Length of wait	3.02	3.00

a. According to a t test, the 2 groups differed on this dependent variable ($P < 0.05$).

Again, participants in the aid condition gave significantly lower ratings to the target dependent variable, diagnostic ability, $t(125.93) = 3.53$, $P < 0.01$, Cohen $d = .54$. In addition, participants in the aid condition also deemed the physicians to be less professional, $t(126) = 2.62$, $P < 0.01$, Cohen $d = .43$, and less thorough, $t(125.95) = 1.743$, $P < 0.05$, Cohen $d = .27$. There were no significant differences between the no-aid and aid conditions on perceived difficulty, overall satisfaction, and length of wait.

Discussion

The data from experiment 2 replicated the findings from experiment 1 using a more serious medical condition. The evaluation of a physician was again influenced by the use of a diagnostic aid.

To improve the generality of our conclusions, we decided to replicate our study with a nonstudent population and a scenario that used a different malady. We wanted to ascertain whether the results found in experiment 1 would generalize to patients who were seeking medical care.

EXPERIMENT 3

Method

Participants

In total, 74 patients from the University Health Connections clinic, an urgent and primary care service for faculty and staff at The Ohio State University, participated in this experiment. Each patient was paid $25 for participation. The participants were randomly assigned to 1 of 3 experimental groups: no aid ($n = 24$), aid ($n = 27$), and prestigious aid ($n = 22$).

Materials

The experiment employed 3 short scenarios that described an interaction between a doctor and his patient and a brief questionnaire containing the target dependent variables. All data were collected using paper-and-pencil materials in the lobby of the University Health Connections clinic. Appendix C contains the scenarios.

Procedure

When participants checked in at the receptionist desk at the University Health Connections clinic to receive medical treatment, they were approached by either a receptionist or a researcher and asked to participate in the experiment. Those patients who gave their consent were given a packet that contained the patient-doctor scenario and the questionnaire containing the 5 dependent variables. The dependent variables (thoroughness of examination, evaluation of length of wait, diagnostic ability, professionalism, and overall satisfaction with the examination) were identical to those in experiment 1. However, the scenario in experiment 1, which described an ankle injury, was replaced with a scenario in which the patient had a persistent cough. Again, the participant was asked to take the role of the patient who went to see his or her primary care physician and was subsequently provided with a diagnosis. In the course of the diagnostic process, the doctor employed no decision aid, an unspecified decision aid, or a prestigious decision aid. After reading the scenario, the participants were asked to rate the physician on the 5 criteria. Participants were asked to respond to each question separately using 5 different Likert-type scales constructed in the same manner as those used in the prior 2 studies.

Results

The 5 dependent variables (thoroughness of examination, evaluation of length of wait, diagnostic ability, professionalism, and overall satisfaction) were subjected to a MANOVA with the single independent variable of experimental condition: no aid, aid, and prestigious aid. See Table 3 for the means of all variables. A significant F value was obtained for the target variable, diagnostic ability, $F(2, 68) = 4.94$, $P < 0.05$, $\eta^2 = .12$. Tukey post hoc tests revealed significant differences between the aid and no-aid groups, $P < 0.05$. Patients at the University Health Connections clinic awarded the highest rating for diagnostic ability to the physician who used no decision aid. There were no significant differences in experimental condition for the remaining 4 dependent variables: thoroughness of examination, evaluation of length of wait, professionalism, and overall satisfaction. Significant Pearson correlations were obtained between pairs of dependent variables ranging from .261 to .862.

Table 3 Experiment 3 Group Means

	Group		
Dependent Variable	No Aid	Aid	Prestigious Aid
Diagnostic ability[a]	4.71	3.52	4.41
Overall satisfaction	5.00	4.19	4.64
Professionalism	4.79	4.04	4.55
Thoroughness of examination	5.04	5.22	5.50
Length of wait	3.46	3.44	3.58

a. According to a Tukey post hoc test, the aid and no-aid groups differed significantly on this dependent variable ($P < 0.05$).

Discussion

The diagnostic ability of the physician was again rated significantly higher when she used no computer-assisted DSS compared to the rating given when she did use a computer-assisted DSS. This result replicated experiment 1 and experiment 2. Unlike experiment 1, however, the use of a prestigious aid—developed at the Mayo Clinic—mitigated the patients' derogation of the physician; no significant difference was detected between the no-aid and prestigious-aid groups.

The other dependent variables did not differ significantly as a function of DSS use. A likely reason for this variation between the 2 experiments is that there were far fewer participants in experiment 3 than in both experiment 1 and experiment 2, which meant lower power to detect significant differences; in this experiment, the power for thoroughness of examination, professionalism of physician, and overall satisfaction ranged from .181 to .429. Furthermore, the difference between the ratings given to the overall satisfaction and thoroughness of the examination dependent variables by the aid and no-aid groups was actually greater in experiment 3 than in experiments 1 and 2.

To expand the generality of our findings, we decided to use medical students in experiment 4 rather than undergraduate students (experiments 1 and 2) or patients (experiment 3). We thought it was important to ascertain if those who were training to be physicians had the same opinion as laypersons with regard to doctors' use of decision aids.

EXPERIMENT 4

Method

Participants

In total, 131 3rd- and 4th-year medical students (students in their clinical years) at The Ohio State University participated in this experiment.

Materials

The experiment employed both the ankle and DVT scenarios that described an interaction between a doctor and a patient plus a brief questionnaire containing the target-dependent variables.

Procedure

An e-mail was sent to all 3rd- and 4th-year medical students at The Ohio State University soliciting their participation in this study. All participants were promised and paid $25. Of the 403 eligible medical students, 131 (33%) responded to our invitation to participate.

Approximately half of the participants read a modified version of the ankle fracture scenario used in experiment 1 ($n = 68$). The other participants read the DVT scenario ($n = 63$).

Some of the groups in this experiment differed from those used in the prior 3 experiments. In addition to the usual no-aid group ($n = 33$) were 3 new ones. In the 1st new group, the physician used a decision aid and explicitly heeded the recommendation of the aid ($n = 30$). In the 2nd group, the physician used a decision aid, explicitly defied the recommendation of the aid, and treated the patient in a less aggressive manner than the recommendation ($n = 38$). In the final new group, the physician used a decision aid, explicitly defied the recommendation of the aid, and treated the patient in a more aggressive manner than the recommendation ($n = 30$). We wanted to contrast these final 2 groups because of the results of prior research that suggests that physicians prefer overtreatment to undertreatment,[30,31] and medical student raters might therefore be more favorably disposed toward a physician who treats more aggressively—but not less aggressively—than a decision aid might recommend.

The same dependent variables were used as in the prior 3 studies with the addition of 2 questions. First, all participants were asked, "How do you think that other physicians would regard you if you did use a computer-assisted diagnostic aid in making a diagnosis in a case such as this one?" Respondents could answer by marking a 7-point scale whose labels varied from *extremely negative* to *extremely positive*. Second, all participants were asked, "If you did use such a decision aid in making a diagnosis and an adverse medical outcome occurred, do you think using the aid would make you more vulnerable or more protected to a claim of malpractice?" Respondents could answer by marking a 7-point scale whose labels varied from *extremely vulnerable* to *extremely protected*.

Results

The 8 dependent variables (malpractice vulnerability, opinion of colleagues, perceived difficulty of diagnosis, thoroughness of examination, evaluation of length of wait, diagnostic ability, professionalism, and overall satisfaction) were subjected to a MANOVA with 2 factors: scenario (ankle or DVT) and experimental group (no aid, heed aid, defy aid by treating less aggressively, defy aid by treating more aggressively).

As was the case in experiment 2, only the diagnostic ability of the physician was significantly affected by the aid manipulation (no aid, aid heeded, aid defied less aggressively, and aid defied more aggressively), $F(3, 128) = 3.71$, $P < 0.05$, $\eta^2 = .08$. See Table 4 for means. The physician who used no DSS was given the highest rating (4.70), whereas the physician who used the aid but defied the output of the aid by treating in a less aggressive manner was given the lowest rating (3.84). A significant difference (using Tukey post hoc procedure) was found between the no-aid group and the "defy—less aggressive" group ($P < 0.05$).

Although it was not of central interest, there was a significant main effect of scenario for several dependent variables: perceived difficulty, $F(1, 130) = 86.90$, $P < 0.05$; thoroughness of examination, $F(1, 130) = 30.09$, $P < 0.05$; wait time, $F(1, 130) = 10.83$, $P < 0.05$; diagnostic ability, $F(1, 130) = 14.91$, $P < 0.05$; and professionalism, $F(1, 130) = 5.00$, $P < 0.05$. The ankle scenario received significantly greater ratings of wait time and perceived difficulty of diagnosis. Physicians in the DVT scenario were perceived as more thorough and professional, and they were given higher ratings of diagnostic ability.

Although no group differences were found on the question pertaining to vulnerability to malpractice, it is interesting to note that the medical students provided an average rating of 4.21, which was slightly toward the "protected" end of the scale.

Table 4 Experiment 4 Group Means

Dependent Variable	Group			
	No Aid	Aid Heeded	Aid Defy—Less Aggressive	Aid Defy—More Aggressive
Physician perception	3.30	3.57	3.45	3.87
Malpractice vulnerability	4.15	4.13	4.08	4.53
Perceived difficulty	4.24	4.40	4.50	4.30
Diagnostic ability[a]	4.70	4.07	3.84	4.10
Overall satisfaction	4.55	4.47	3.95	3.93
Professionalism	4.85	4.73	4.66	4.43
Thoroughness of examination	4.30	4.33	4.42	4.37
Length of wait	3.27	3.50	3.37	3.10

a. According to a Tukey post hoc test, the no-aid and aid defy—less aggressive groups differed significantly on this dependent variable ($P < 0.05$).

Discussion

We again found that the physician who used no decision aid was rated the highest in diagnostic ability. The physician who used an aid but defied it by treating in a less aggressive manner was deemed to have significantly less diagnostic ability. This mirrors the results of Pezzo and Pezzo,[32] who found in their experiment 2 that a physician who defied the recommendation of a decision aid was considered to be less competent than one who heeded its recommendation. Pezzo and Pezzo[32] found no significant difference in rated competence between those physicians who defied an aid and those who used no aid, a difference we did find in our study. However, a major difference between the procedure in their study and ours was that in their study, an adverse outcome occurred. As we suspected, the physician in our study who defied the aid and treated in a more aggressive manner was rated higher than the physician who defied the aid and treated in a less aggressive manner, but this difference was not significant.

GENERAL DISCUSSION

In both experiment 1 and experiment 2, we found that respondents derogated the diagnostic ability of physicians who used a computer-based DSS. In experiment 1, we also found evidence that other characteristics of the physician, such as professionalism, were also rated lower if a DSS was used. Unfortunately, patients' derogation of physicians who use a decision aid might be a defensible response, even if the aid does promote accurate judgment performance. According to social psychology's discounting principle,[33] whenever an effect has 2 possible causes, a perceiver tends to emphasize one cause and discount the other. Thus, if a correct diagnosis is made, and its accuracy can be attributed to either the skill of the physician or the output of the computer-assisted decision aid, it may be understandable why the perceiver may attribute some component of the accuracy to the aid rather than attributing all of it to the person. Consistent with this analysis, Pezzo and Pezzo[32] found that compared to a situation in which a physician did not use a decision aid, if a physician did use the decision aid, he or she was rated less positive following a good outcome and less negative following a bad one. Raters apparently attributed to the aid some of the credit or blame for the outcome, thus reducing the magnitude of the impact of the outcome on their evaluation of the physician. An unfortunate consequence of this attribution is that it may make decision makers less willing to use the decision aid for fear of appearing less capable should the outcome be a typical positive one. Because patients want their physician to be extremely knowledgeable (if not omniscient), the physician who appears to rely on a DSS may be risking the patient's otherwise high regard. The physician prudent enough to employ a helpful aid should be thanked—not derogated. We hasten to point out that this attributional complexity is germane not only to the medical arena but to virtually all other domains in which professional judgment is required. The discounting principle may be involved whenever a decision maker's ability and the decision aid's precision are each possible causes of a professional judgment.

We were surprised in experiment 1 that the independent variable significantly influenced not only the rated diagnostic ability of the physician—the dependent variable on which we anticipated the greatest impact—but 3 of the 4 additional dependent variables too. Some of these additional factors, such as the thoroughness of the examination, would not seem to be directly related to the use of a diagnostic aid. However, another social psychology principle, the "halo effect," might apply here.[34] This effect pertains to the phenomenon in which an evaluation of 1 particular salient trait of a person influences the evaluation of other traits of that same person. Thus, if one thinks that a physician is a good diagnostician, one is likely to believe in addition that the physician has

other positive characteristics—very thorough, for example—even if there exists no relevant evidence bearing on these other factors.

Possible Theoretical Frameworks

Loss of the Aura of Omniscience

Kaplan[35] suggests that doctors' negative attitudes toward and underutilization of decision aids might both be fueled by the fear that the use of such aids might diminish their professional status. Our data suggest that this fear might be well founded in that patients rate the physician who uses an aid as a less competent diagnostician than a physician who acts without such assistance. Abundant research suggests that people think they can do better than a decision aid. For example, Arkes and colleagues[36] demonstrated that baseball experts were reluctant to use a decision rule despite the fact that it would have improved their performance significantly on a baseball-related judgment task. The experts were confident that they could do well without the helpful rule. Whitecotton[37] showed that there was a significant negative correlation between the confidence that professional financial analysts manifested toward the forecasting task that confronted them and their willingness to use a decision aid to perform the task. If we assume that those who hire the experts are as confident in the experts' performance as the experts themselves are, this would explain why patients have a relatively negative view toward expert diagnosticians who do not seem self-assured enough to eschew a decision aid.

Intuition v. Analysis

Insight and creativity are often associated with intuitive thinking. We admire such attributes. Analysis, on the other hand, seems tedious and opaque to a layperson. Latham and Whyte[38] and Whyte and Latham[39] investigated the relative persuasiveness of advice provided to managers either by an unaided expert or by an expert who had performed a utility analysis to support his or her recommendation. Although the advice was identical, the recommended course of action was less likely to be supported by the managers if it were predicated on the utility analysis. The fact that the merits of the utility analysis were explained by an expert who was an "internationally recognized authority" did not seem to impress the managers in a positive way.

We hypothesize that patient-raters might have an opinion similar to that of the managers in the aforementioned research. The unaided physician is analogous to the unaided expert who uses only intuition. The physician who uses a DSS is analogous to the expert who uses a decision analysis. In both situations, the person who uses his or her own intuition benefits, either by having the advice heeded or by being given higher evaluations. One difference in the 2 domains is that in the utility analysis study by Whyte and Latham,[39] the support of a prestigious expert did not seem to help. In our experiment 3, we did find that when patients learned that the aid had been developed at the prestigious Mayo Clinic, the derogation usually visited upon the physician who used a decision aid was largely mitigated.

Several authors have suggested reasons why people trust intuition more than analysis.[40–42] Among the most prominent is the suspicion that some important intuitions are performed without awareness and therefore cannot be expressed in analytic terms; this would therefore render them unavailable for inclusion in a decision aid. Whether this suspicion is warranted is a matter of debate.[43,44]

Case-Specific Data v. Base Rates

DSS tools gain their power from being based on a large amount of prior information—namely, the base rates of various diseases and the likelihood ratios that describe the probability of having a particular symptom given the presence of each disease. An abundance of psychological research compels the conclusions that most people do not 1) think that base rates are relevant,[45] 2) use them appropriately when they are presented,[46] or 3) consider them to be as probative as individuating information.[47] We want our physician to see us as individuals. Each of us is unique, and each of us wants to be treated accordingly. If a DSS lumps each of us with all the other people whose data go into the base rates or likelihood ratios, we feel minimized or even disregarded.[40] A physician who uses a DSS that offends us in this manner is going to be derogated compared to a physician who appears to appreciate our individuality, even though the latter physician may not be as accurate a diagnostician as the former.

The Issue of Generality

A number of factors that varied between and within the 4 experiments did not qualify the general finding that physicians who used no decision aid were deemed to be more capable than at least some physicians who did.

In experiments 1, 2, and 4, the pronouns used in the scenario would lead one to conclude that the

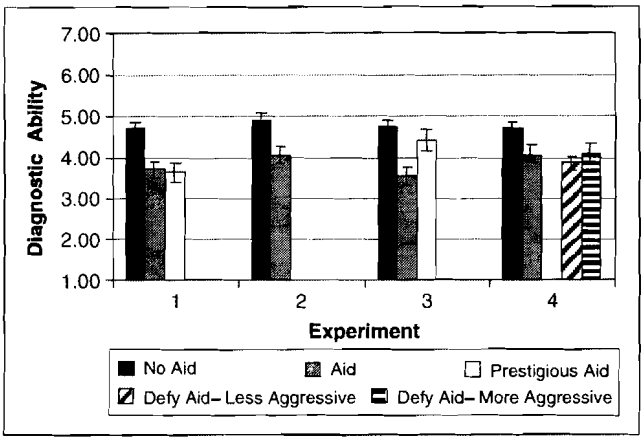


Figure 1 Ratings of diagnostic ability across experiments.

physician was a female. In experiment 3, the physician was a male. The principal result was the same in all studies.

In experiment 2, the medical condition was more serious than that in experiments 1 and 3. The principal results were the same in all 3 studies.

In experiments 1 and 2, the participants were younger than those in experiments 3 and 4. The principal results were the same.

In experiment 4, the participants were in training to be physicians, whereas the participants in the other studies were laypersons. Again, the principal result was the same.

We conclude that our principal finding has substantial generality. The common pattern of results in all 4 experiments is depicted in Figure 1.

External Validity

Reviewers of a prior draft of this article have questioned whether various aspects of our scenarios faithfully depict a typical doctor-patient interaction. For example, would a patient normally be aware of the fact that the physician has used a decision aid? In order for us to manipulate this independent variable, the participant in our research who was taking the role of the patient had to be made aware of whether the physician was using a decision aid. Thus, to investigate the influence of this factor, we had to make the presence of the aid quite explicit. From personal experience, we know that patients sometimes are aware of the fact that the physician has consulted a reference book, a colleague, or a journal article. We wanted to investigate what would be the patient's attitude if a computer-assisted aid were consulted instead.

Another closely related issue pertains to the context in which the aid was used. Perhaps a patient's attitude toward the physician who used a DSS would vary depending on whether the aid was used to make the initial diagnosis, corroborate a physician's own diagnosis, amalgamate a large number of complex laboratory tests, make the final diagnosis, or even override a physician's intuition. Of course, all of these are potentially important qualifications of our main finding. In our experiment 1, our scenario included the following: "she is using a computer program to decide whether to order an X-ray of your ankle." In experiment 3, we used the following: "he explains that to help make the diagnosis, he is using a computer program." Compared to its helping or consultative role in experiment 3, the DSS seems to be playing a more direct role in making the diagnosis in experiment 1. In both cases, the physician who used no decision aid was rated as more competent than the physician who did use an aid. Future research might wish to diminish the role of the decision aid even further to ascertain if the effect remains when the aid's role becomes truly nominal.

Another feature of our scenarios is that the use of a DSS may have delayed or frustrated a person who expected the physician to respond immediately with the recommended course of action. In experiment 1, the patient in the scenario had already waited 45 minutes, and the additional wait while the DSS was being used may have been particularly aversive. This extra delay may have been a source of some of the displeasure directed toward the physician who used a diagnostic aid. To the extent the use of an aid does not result in any delay or does not follow a wait for an already tardy and harried health care provider, patients may be less negative toward the physician who employs a DSS.

The results of these studies pose a problem for a physician who believes that computer-assisted DSSs improve diagnostic accuracy, as several studies suggest.[6–11] Using such decision aids may indeed increase diagnostic accuracy, but they might also simultaneously decrease the patient's opinion of the physician's diagnostic ability.

APPENDIX A
Experiment 1 Ankle Scenarios

NO-AID SCENARIO

On Saturday afternoon, during an informal game of softball at the local park, you hurt your left ankle. You jumped up to catch a line drive, and when you

landed, your ankle turned in. You fell to the ground and were unable to get up or walk because of the pain. Your teammates, one of whom is a physical therapist, helped you to the side and got some ice for your ankle from the concession stand. Your friend gave you a ride home and helped you to your couch. She recommended that you see your physician tomorrow. Until then, you kept your ankle elevated, used ice, and took some ibuprofen for the pain.

Early the next morning, at 7:30 AM, you called your physician's office. The recorded message said they do not take calls for appointments for another hour and a half. You called back promptly 90 minutes later and got an appointment later that afternoon. You managed to go to work, using an ornamental walking stick you brought back from a trip.

In the afternoon, you left work early to get to your appointment. After a 30-minute wait, a nurse takes you to an examination room. The nurse asks you what the problem is, and you respond that you injured your left ankle yesterday playing softball. The nurse takes your temperature (98.7°), measures your blood pressure (122/78), takes your pulse (78 beats per minute), measures your respiratory rate (16 breaths per minute), and asks you to step on the scale. The nurse writes this information in the chart and then leaves. You want to ask her if you should get an X-ray to help move things along, but she left before you had a chance.

About 15 minutes later, the doctor comes into the room and asks you a number of questions:

1. When did this happen?
   Last evening.
2. What were you doing?
   Playing softball.
3. Describe the accident.
   When I landed, my ankle turned in. I could not walk on it. My friends helped me to the side and got some ice for it. The ankle swelled up, and there seems to be a bruise over the outer part of it.
4. Can you walk on it now?
   No, I cannot put my full weight on it. I must use this walking stick.
5. Have you ever injured your ankle before?
   Yes, but only minor twists.

In addition, the doctor asks you some more questions:

Do you have any drug allergies? *No.*
Do you have any other major health problems? *No.*

The doctor uses her stethoscope to listen to your lungs as you breathe deeply and to your heart while you lie back quietly. She then examines your ankles: your left ankle is puffy. There is a visible black and blue mark on the outside, over the bone. When she pushes over the bone on the outside of the ankle, it is tender. She moves it side to side and back and forth. It is beginning to throb. She asks you to walk on it without the walking stick. You are unable to. The doctor asks you to sit down again.

[Additional text for aid scenario: The doctor then turns to a computer in the room, explaining that she is using a computer program to decide whether to order an X-ray of your ankle. According to this decision aid, you should have an X-ray of your ankle to see if it is fractured. So she orders an X-ray of your ankle.] [Additional text for prestigious-aid scenario: The doctor then turns to a computer in the room, explaining that she is using a computer program developed at the prestigious Mayo Clinic, one of the nation's premier medical facilities. She explains that she is using their computer program to decide whether to order an X-ray of your ankle. According to this decision aid, you should have an X-ray of your ankle to see if it is fractured. So she orders an X-ray of your ankle.] [Concluding sentence in the no-aid scenario: Your doctor explains to you that she is concerned that you might have fractured your ankle during the injury. So she orders an X-ray of your ankle.]

NUMERACY SCALE

1. Imagine that we rolled a fair, 6-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?
2. In the BIG BUCKS LOTTERY, the chance of winning a $10.00 prize is 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket to BIG BUCKS?
3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES wins a car?

APPENDIX B
Experiment 2 Deep Venous Thrombosis Scenarios

In this study, we are interested in what factors lead people to like or dislike their interactions with medical personnel. We are going to present you with a scenario describing a patient-physician interaction. Please take

the role of a patient, read the story, and then, using your own medical knowledge and experience, give us your candid opinion of this physician.

After 5 days at home with a painful, swollen right leg, you decide to go to the doctor's office. You wait in the waiting room for about 30 minutes with approximately a half-dozen other people. Some of them are coughing, sneezing, and blowing their noses; you find the atmosphere very noisy.

After the 30-minute wait, a nurse takes you to an examination room. The nurse asks you what the problem is, and you respond that you've had a very swollen and sore leg for 5 days, after spraining your ankle skiing. The nurse takes your temperature (98.7°), measures your blood pressure (122/78), takes your pulse (78 beats per minute), measures your respiratory rate (16 breaths per minute), and asks you to step on the scale so you can be weighed. The nurse writes this information in the chart and then leaves you alone in the room. About 15 minutes later, the doctor comes into the room.

The doctor asks you a large number of questions:

- Why did you come to the doctor's office today? *You repeat that your leg has been painful and swollen for 5 days after you sprained your ankle skiing.*
- Can you walk on the leg? *You could after the accident, but for the last 3 days, you have stayed off of your leg. In fact, the ankle feels much better, but the leg itself hurts.*
- Has anyone examined the leg since the injury? *After the accident, the local emergency department examined and X-rayed the leg. The emergency department physician told you that it was just a sprain—you did not tear any ligaments and did not fracture any bones.*
- Do you have paralysis, paresis (weakness), or recent cast immobilization of the lower extremities? *No, you used an Ace wrap on the ankle for a couple of days.*
- Have you recently been bedridden for more than 3 days? *No, you have not stayed in bed but have been lying on the couch with your leg elevated most of the time since the accident. You have been getting up to go to the bathroom, make yourself meals, and stretch out occasionally.*
- Have you recently traveled anywhere, sitting still for 4 or more hours? *Yes, the trip home from the ski resort took 6 hours.*
- Do you have active cancer, are you being treated for cancer, have you been treated for cancer within the previous 6 months, or are you now having palliative care? *No.*
- Have you had major surgery within 4 weeks? *No.*
- Have you ever had a blood clot? *No.*
- Has anyone in your family ever had a blood clot? *No.*
- Have you been running any fevers or had shaking chills? *No.*
- Have you had any recent bug bites, cuts, or scrapes on the leg? *No.*
- Do you have a cough? *No.*
- Have you coughed up any blood? *No.*
- Can you feel your heart pounding? *No.*
- Are you short of breath? *No.*

In addition, the doctor asks you some more questions:

- Do you have any drug allergies? *No.*
- Do you have any other major health problems? *No.*
- Do you bleed easily? *No.*
- Have you ever had a stroke or an ulcer? *No.*

The doctor uses her stethoscope to listen to your heart and your lungs as you breathe deeply. She examines your legs: your right leg is swollen to just above the knee and is tender to gentle squeezing throughout. It is warm to the touch but not red. The right calf is 1.5 inches in diameter larger than your left leg. If you push on the leg, you can see the indentation after you remove your finger. The veins on your leg are not evident.

Your doctor explains to you that she is very concerned that you have a blood clot in your leg (a "deep venous thrombosis," she calls it). This can be dangerous, even fatal. So, she wants to check some blood test and send you immediately for a test (a "duplex ultrasound") to see if there is a clot in your right leg. She says she will be right back, after she arranges for the test.

Twenty minutes later, the nurse returns to the room. First, the nurse draws some blood. Then the nurse tells you that the test has been ordered. The nurse then takes you over to the test in a wheelchair and tells you that someone will come and pick you up when the test is over.

The test takes 20 minutes, but then you wait an hour before someone brings you back to the clinic. In the clinic, you wait for another 15 minutes before the doctor comes in to see you. She sits down and examines the test results. After a brief pause, she tells you that these tests do not show a blood clot in the leg.

The doctor says that you should take ibuprofen and use hot/cold packs for the pain.

[NO DECISION AID] However, the doctor says that you are at intermediate risk for a deep venous

thrombosis, and so the duplex ultrasound should be done again in 1 week. You make an appointment 1 week from now.

[COMPUTER DECISION AID] The doctor then turns to a computer in the room, explaining to you that she is using a computer program to decide what to do next. According to this decision aid, you are at intermediate risk for a deep venous thrombosis, and the "advice" from the decision aid is that the duplex ultrasound should be done again in 1 week. So you make an appointment 1 week from now.

Before you leave, the doctor proceeds to explain to you how a blood clot is treated, just in case you do eventually develop a clot. You would be admitted to the hospital for 4 to 7 days. First an intravenous (IV) blood thinner will be started, and then a pill form of a blood thinner will be started. When your blood is thin enough, the IV thinner will be stopped and you can go home on the pill. You will probably remain on the pill for 6 months. After hearing this explanation, you leave the doctor's office.

APPENDIX C
Experiment 3 Cough Scenarios

NO-AID SCENARIO

After 5 days with a particularly bad sore throat, you decide to go to the doctor. You wait in the waiting room for about 30 minutes with approximately a half-dozen other people. Some of them are coughing, sneezing, and blowing their nose.

After the 30-minute wait, a nurse takes you to an examination room. The nurse asks you what the problem is, and you respond that you've had a very sore throat for 5 days. It's really hard for you to speak to the nurse because your throat is so sore. Your voice is extremely hoarse. The nurse takes your temperature (99.5°), measures your blood pressure (120/80), and asks you to step on the scale. The nurse writes this information in the chart and then leaves. About 15 minutes later, the doctor comes into the room.

The doctor asks you a large number of questions concerning the following:

1. Whether you have been coughing up blood (*no*)
2. What color the material is that you do cough up (*yellow*)
3. When the sore throat started (*5 days ago*)
4. Whether you have any chest pain (*no*)
5. Whether you are on any medications at the current time (*no*)
6. Whether the sore throat is more painful when you swallow (*yes*)
7. Whether you've had a lot of nasal congestion (*yes*)

The doctor also uses his stethoscope to listen to your heart and then to listen to your lungs as you breathe deeply.

He then asks more questions:

1. Do you have allergies? (*yes, ragweed*)
2. Do you have asthma? (*no*)
3. Have you ever had pneumonia? (*no*)

Although the doctor did not ask you about the topic, you volunteer the information that during the last several years, you've been treated for several urinary and kidney infections.

He then feels the lymph nodes under your jaw in order to determine if they are swollen. (They are.) He looks in your ears and asks if you have had any earaches in the last few days. You say that you have had a couple of earaches.

The doctor then says that he will order some laboratory tests. First, he takes a swab and brushes it against the back of your throat. He says that he wants to get a culture in case there is a strep infection. He says that he wants to get a chest X-ray because he wants to get a better idea of the condition of your lungs. He says that he detected some fluid when he listened to your lungs with the stethoscope. He says that the nurse will return to take a blood sample. He says that the laboratory tests will be back tomorrow and that you should stop by tomorrow early in the afternoon. He writes a prescription, which he tells you to get filled right away. The nurse comes in and draws some blood from your left arm. Then you go get the chest X-ray. When you finish with the X-rays, you go to the pharmacy to get the prescription filled. You go home, take the medicine, and spend the rest of the day in bed.

The following day, you stop back in the early afternoon for your 1 PM appointment. You wait about 30 minutes, and then you are called into the examination room. The nurse again takes your temperature (99.5°) and blood pressure (120/80). The doctor comes in 10 minutes later. He sits down and examines the laboratory test results, which have just come in. He has not examined the results before your appointment. He says that you don't have a strep infection. He then looks at the chest X-ray and looks again at the various laboratory results, including the results from the blood tests. [The following sentence is only in the no-aid scenario: He spends a minute or two mulling over this information.] [Additional text for aid scenario: He

then opens a laptop computer and types in the results of the various laboratory tests. He also types in the information you provided yesterday concerning your coughing, earaches, etc. He explains that to help make the diagnosis, he is using a computer program [The following clause is inserted only in the prestigious-aid scenario: developed at the prestigious Mayo Clinic, one of the nation's premier medical facilities]. To run the program, he says, he merely types in the information you've provided. The computer program then assigns a likelihood to each of several possible diagnoses. He says that based on his examination of you yesterday, he thought at that time that you might have pneumonia, but based on the output of the computer program, he now is pretty sure that you have acute bronchitis.] He says that based on his examination of you yesterday, he thought at that time that you might have pneumonia, but based on the laboratory tests and X-ray, he now is pretty sure that you have acute bronchitis. He tells you to continue to take the medication he prescribed yesterday, and in addition, he writes you a 2nd prescription for a cough suppressant. He tells you to come back in 5 days if there is no improvement. He says that you will not be entirely well for a couple of weeks; the cough will last quite a long time.

REFERENCES

1. Kaplan B. Evaluating informatics applications: clinical decision support systems literature review. Int J Med Inform. 2001; 64:15–37.

2. Balas EA. Information systems can prevent errors and improve quality. J Am Med Inform Assoc. 2001;8:398–9.

3. 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;8:527–34.

4. Pozen MW, D'Agostino RB, Mitchell JB, et al. The usefulness of a predictive instrument to reduce inappropriate admissions to the coronary care unit. Ann Intern Med. 1980;92:238–42.

5. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. Science. 1989;243:1668–74.

6. Corey GA, Merenstein JH. Applying the acute ischemic heart disease predictive instrument. J Fam Pract. 1987;25:127–33.

7. de Dombal FT, Dallos V, McAdam WAF. Can computer aided teaching packages improve clinical care in patients with acute abdominal pain? BMJ. 1991;310:1495–7.

8. Ridderikhoff J, van Herk E. A diagnostic support system in general practice: is it feasible? Int J Med Inform. 1997;45:133–43.

9. Friedman DP, Elstein AS, Wolf FM. Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. JAMA. 1999;282:1851–5.

10. Getty DJ, Pickett RM, D'Orsi CJ, Swets JA. Enhanced interpretation of diagnostic images. Invest Radiol. 1988;23:240–52.

11. Chase CR, Vacek PM, Shinozaki T, Giard AM, Ashikaga T. Medical information management: improving the transfer of research results to presurgical evaluation. Med Care. 1983;21:410–24.

12. Ebell MH, Smith MA, Barry HC, Ives K, Carey M. Does this patient have strep throat? JAMA. 2000;284:2912–8.

13. Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. Med Decis Making. 1981;1:239–46.

14. Walsh BT, Bookheim WW, Johnson RC, Tompkins RK. Recognition of *Streptococcal pharyngitis* in adults. Arch Intern Med. 1975;135:1493–7.

15. Stiell I. Ottawa ankle rules. Can Fam Physician. 1996;42: 478–80.

16. Hunt DL, Haynes B, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes. JAMA. 1998;280:1339–46.

17. Kassirer JP. A report card on computer-assisted diagnosis—the grade C. New Engl J Med. 1994;330:1824–5.

18. Rocha BHSC, Christenson JC, Evans RS, Gardner RM. Clinicians' response to computerized detection of infections. J Am Med Inform Assoc. 2001;8:117–25.

19. Overhage JM, Tierney WM, McDonald CJ. Computer reminders to implement preventive care guidelines for hospitalized patients. Arch Intern Med. 1996;156:1551–6.

20. Keefe B, Subramanian U, Tierney WM, et al. Provider response to computer-based care suggestions for chronic heart failure. Med Care. 2005;43:461–5.

21. Graham ID, Stiell IG, Laupacis A, et al. Awareness and use of the Ottawa Ankle and Knee Rules in 5 countries: can publication alone be enough to change practice? Ann Emerg Med. 2001; 37:259–66.

22. Hilton NZ, Simmons JL. The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. Law Hum Beh. 2001;25:393–408.

23. Cabana MD, Rand CS, Powe NR. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA. 1999;282:1458–65.

24. Meehl PE. Causes and effects of my disturbing little book. J Pers Assess. 1986;50:370–5.

25. Cruikshank PJ. Patient ratings of doctors using computers. Soc Sci Med. 1985;21:615–22.

26. Garrison GM, Bernard ME. 21st century health care: the effect of computer use by physicians on patient satisfaction at a family medicine clinic. Fam Med. 2002;34:362–8.

27. Solomon GL, Dechter M. Are patients pleased with computer use in the examination room? J Fam Pract. 1995;41:242–4.

28. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making. 2001;21:37–44.

29. Schwartz LM, Woloshin S, Black WC, Welch GH. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127:966–71.

30. Ayanian JZ, Berwick DM. Do physicians have a bias toward action? A classic study revisited. Med Decis Making. 1991;11:154–8.

31. Bakwin H. Pseudodoxia pediatrica. New Engl J Med. 1945; 232:691–7.

32. Pezzo MV, Pezzo SP. Physician evaluation after medical errors: does having a computer decision aid help or hurt in hindsight? Med Decis Making. 2006;26:48–56.

33. Kelley HH. The process of causal attribution. Am Psychol. 1973;28:107–28.

34. Thorndike EL. A constant error on psychological rating. J Appl Psychol. 1920;4:25–9.

35. Kaplan B. Culture counts: how institutional values affect computer use. MD Comput. 2000;17:23–6.

36. Arkes HR, Dawes RM, Christensen C. Factors influencing the use of a decision rule in a probabilistic task. Organ Behav Hum Dec. 1986;37:93–110.

37. Whitecotton SM. The effects of experience and confidence on decision aid reliance: a causal model. Behav Res Account. 1996; 8:194–216.

38. Latham GP, Whyte G. The futility of utility analysis. Pers Psychol. 1994;47:31–46.

39. Whyte G, Latham GP. The futility of utility analysis revisited: when even an expert fails. Pers Psychol. 1997;50:601–10.

40. Dawes RM. The robust beauty of improper linear models. Am Psychol. 1979;34:571–82.

41. Meehl PE. Causes and effects of my disturbing little book. J Pers Assess. 1986;50:370–5.

42. Yates JF, Veinott ES, Patalano AL. Hard decisions, bad decisions: on decision quality and decision aiding. In: Schneider SL, Shanteau J, eds. Emerging Perspectives on Judgment and Decision Research. New York: Cambridge University Press; 2003. p 13–63.

43. Hammond KR. Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice. New York: Oxford University Press; 1996.

44. Hogarth RM. Educating Intuition. Chicago: University of Chicago Press; 2001.

45. Arkes HR. Impediments to accurate clinical judgment and possible ways to minimize their impact. J Consult Clin Psych. 1981;49: 323–30.

46. Arkes HR, Rothbart M. Memory organization, retrieval and contingency judgments. J Pers Soc Psychol. 1985;49:598–606.

47. Kahneman D, Tversky A. On the psychology of prediction. Psychol Rev. 1973;80:237–51.

Exploring Models to Eliminate Cancer Disparities Among African American and Latino Populations: Research and Community Solutions

Supplement to ***Cancer***

Improving Follow-Up to Abnormal Breast Cancer Screening in an Urban Population

A Patient Navigation Intervention

Tracy A. Battaglia, MD, MPH[1,2]
Kathryn Roloff, BS[3]
Michael A. Posner, PhD[4]
Karen M. Freund, MD, MPH[1,2]

[1] Women's Health Unit, Boston University School of Medicine, Boston, Massachusetts.

[2] Women's Health Interdisciplinary Research Center, Boston University School of Medicine, Boston, Massachusetts.

[3] Department of Organizational Behavior, Harvard Business School, Boston, Massachusetts.

[4] Department of Mathematical Sciences, Villanova University, Villanova, Pennsylvania.

*Presented at Exploring Models to Eliminate Cancer Disparities Among African American and Latino Populations: Research and Community Solutions, Atlanta, GA, April 21–22, 2005.

Supported in part by grants from Building Interdisciplinary Research Careers in Women's Health (BIRCWH) Faculty Scholar Award ([K12-HD43444]), ACS Career Development Award ([CCCDA-03-228-01]) and the Avon Foundation.

We acknowledge the daily efforts of our Patient Navigator, Wanda Turner, for her steadfast dedication and commitment to the women we serve. In addition, we thank Emily Looney and Shreya Patel for their contributions to the final dataset and manuscript preparation.

Address for reprints: Tracy A. Battaglia, MD, MPH, Women's Health Unit, Section of General Internal Medicine, Evans Department of Medicine, Boston University School of Medicine, Doctors Office Building, 720 Harrison Avenue, Suite 1108, Boston, MA 02118, USA; Fax: (617) 638-8026; E-mail: tracy.battaglia@bmc.org

Received June 30, 2006; revision received September 19, 2006; accepted September 21, 2006.

Delays in follow-up after cancer screening contribute to racial/ethnic disparities in cancer outcomes. We evaluated a patient navigator intervention among inner-city women with breast abnormalities. A full-time patient navigator supported patients using the care management model. Female patients 18 years and above, referred to an urban, hospital-based, diagnostic breast health practice from January to June 2000 (preintervention) and November 2001 to February 2003 (intervention), were studied. Timely follow-up was defined as arrival to diagnostic evaluation within 120 days from the date the original appointment was scheduled. Data were collected via computerized registration, medical records, and patient interview. Bivariate and multivariate logistic regression analyses were conducted, comparing preintervention and intervention groups, with propensity score analysis and time trend analysis to address the limitations of the pre–post design. 314 patients were scheduled preintervention; 1018, during the intervention. Overall, mean age was 44 years; 40% black, 36% non-Hispanic white, 14% Hispanic, 4% Asian, 5% other; 15% required an interpreter; 68% had no or only public insurance. Forty-four percent of referrals originated from a community health center, 34% from a hospital-based practice. During the intervention, 78% had timely follow-up versus 64% preintervention ($P < .0001$). In adjusted analyses, women in the intervention group had 39% greater odds of having timely follow-up (95% CI, 1.01–1.9). Timely follow-up in the adjusted model was associated with older age ($P = .0003$), having private insurance ($P = .006$), having an abnormal mammogram ($P = .0001$), and being referred from a hospital-based practice, as compared to a community health center ($P = .003$). Our data suggest a benefit of patient navigators in reducing delay in breast cancer care for poor and minority populations. ***Cancer* 2007;109(2 Suppl);359–67.**

KEYWORDS: breast neoplasm, mass screening, case management, urban population, minority groups, medically underserved areas.

The unequal burden of cancer is highlighted among lower socioeconomic status and racial/ethnic minority women who suffer higher mortality from breast cancer compared with their more affluent non-Hispanic white counterparts. Age standardized death rates among African American women with breast cancer exceed those of non-Hispanic white women. Similarly, residents of poorer counties have higher death rates from breast cancer than do residents of more affluent counties.[1] Advanced stage at diagnosis, which contri-

DOI 10.1002/cncr.22354
Published online 22 November 2006 in Wiley InterScience (www.interscience.wiley.com).

butes to poorer outcomes, is found more frequently in these populations. Recent SEER statistics show that the proportion of women diagnosed with regional- and distant-stage breast cancer continues to be higher among African Americans and Hispanics than among non-Hispanic whites. The same is true when one compares high- versus low-poverty census tracts; the proportion diagnosed with distant stages is higher in high-poverty census tracts.[1]

Improvements in access and equity in mammography screening alone have not translated into survival improvements for these disadvantaged women.[1] Although racial/ethnic differences in mammography utilization rates have been found to explain some of the observed differences in stage at diagnosis,[2] several observational studies have not found this to be a significant explanatory factor.[3–5] Potential reduction in morbidity and mortality through breast cancer screening will never be realized without timely and efficient follow-up care once an abnormality has been detected. Investigating the disparities in follow-up after abnormal breast screening serves as the next step in addressing breast cancer health disparities. Although measures of follow-up lack precision,[6] numerous studies have documented that racial/ethnic minority women often suffer from the longest delays at this point.[5,7–11]

Patient navigation, a type of care management that encompasses a wide-range of advocacy and coordination activities, has been proposed to address known barriers to the delivery of high-quality cancer care in underserved populations.[12–14] A recent national case study review of emerging programs to reduce racial/ethnic disparities in cancer found that almost all adopted a variant of patient navigation.[15] Despite increasing interest along with both federal and private resource allocation,[16,17] published evidence of the benefit of navigation programs is limited.[18]

We evaluated a hospital-based patient navigator intervention among inner-city minority women with breast abnormalities. The main objective of the intervention is to improve the rate of timely diagnostic follow-up in comparison to a group of historical controls, and to identify characteristics of patients who are most at risk for loss to follow-up.

MATERIALS AND METHODS

This pre–post intervention study was conducted at a hospital-based diagnostic breast health practice at a major academic medical center in Boston, Massachusetts. The specialty practice accepts referrals for evaluation of any breast health issue, including screening abnormalities, suspicious symptoms, or elevated cancer risk. Patients served include those who receive their primary care at the academic medical center and those from over 20 affiliated community health centers throughout Boston. Together with its affiliated health centers, the academic institution serves as the major safety-net hospital in the region. All income eligible, uninsured women can receive services without charge. This study was approved by our Institutional Review Board. Written informed consent was waived since the intervention was implemented as a new standard of care for all patients. The final dataset used in these analyses was devoid of all patient identifiers.

Study Sample

The sample for this study included all women with scheduled visits at the diagnostic breast health practice from January through June 2000 (preintervention, $N = 314$) and November 2001 through February 2003 (intervention, $N = 1018$). We included consecutive female patients >18 years referred for evaluation during the study period, as referral criteria to the practice were consistent throughout this time period. Individual women were only included once throughout the study, so that preintervention and intervention subjects are unique women. For women seen more than once during this timeframe, we evaluated their initial scheduled visit only, using subsequent visits for the sole purpose of evaluating follow-up within 120 days.

Preintervention

As a result of initial observations of patient care data showing high rates of failure to arrive for scheduled evaluation, a formal program evaluation was conceived. At the time of preintervention data collection, the standard protocol for these patients used existing secretarial staff to attempt telephone contact without clinical oversight. In the ensuing 16 months prior to program implementation, this protocol continued without other institutional or practice changes while a needs assessment was undertaken to inform the intervention. Findings from structured interviews with patients, breast health providers/support staff, and referring providers identified lack of coordination of care and communication as major barriers to diagnostic evaluation. Program planning, hiring, and navigator training occurred during this time period.

Intervention

The patient navigator intervention was guided by the principles of *Care Management.*[19] Services provided by the navigator focused around 4 key activities: 1) case

identification, 2) identification of individual barriers to care, 3) implementation of a care plan, and 4) tracking through completion. Criteria for hiring of the patient navigator included experience caring for a diverse patient population and knowledge of the existing local health systems. The patient navigator was trained to coordinate care for each patient referred for diagnostic evaluation. Training included written triage and follow-up protocols as well as monthly barriers-focused cultural competence training at the local department of public health. Initial contact with patients began after appointments were scheduled, 1 week before their scheduled visit. Telephone outreach with all patients, using interpreters for non-English speaking women, was attempted to confirm appointments, provide information about the visit, and learn about any individual barriers to arriving for that appointment. The navigator utilized available resources to address those barriers. A key function of the navigator included advocating for the patient through rapid communication with breast health providers, referring providers, and with other specialty sites, including radiology, surgery, and pathology. In accordance with the 4th key activity, the patient navigator, with daily assistance and weekly oversight by the study coordinator, tracked information on patient demographics, reason for referral, site and provider who referred the patient, diagnostic evaluation conducted, and outcomes of evaluation. Dates from time of initial referral through arrival to first scheduled visit were collected.

Data Collection

Preintervention data collection was conducted via retrospective chart review for women referred to the diagnostic breast evaluation practice from January through June 2000. The administrative registration database was reviewed for the date that the first diagnostic appointment was scheduled, demographic data, and the date at which the patient first arrived for evaluation. If the patient's race was not listed in the registration database, but their birthplace and native language were listed, race was extrapolated from birthplace and native language. For example, if a woman was born in Haiti and spoke Haitian-Creole, she was categorized as black. The clinical record was abstracted for reason for visit based on the following categories: abnormal mammography (BI-RADS category 0, 3, 4, 5), abnormal clinical breast exam, or other (including suspicious symptoms or elevated cancer risk). To ensure accuracy of data, all preintervention data were reviewed by a second chart abstractor.

Intervention data were collected prospectively by the navigator beginning 1 week before the scheduled diagnostic visit at the time schedules were reviewed for case identification and telephone outreach. Following preintervention data collection protocol, the administrative registration database was reviewed for the date that the diagnostic appointment was scheduled and demographic data while the clinical record was abstracted for reason for visit. The date the subjects first arrived to a diagnostic evaluation was documented prospectively by the navigator. To ensure completeness and accuracy of data, all intervention data were reviewed by a one of the authors on a weekly basis (TB, KF). The final analytic file was a limited dataset devoid of key patient identifiers.

The main outcome in this intervention study was the dichotomous variable *timely follow-up* (yes, no). Subjects were considered to have timely follow-up if they arrived to a diagnostic evaluation visit within 120 days from the date the original appointment was scheduled. Although an ideal outcome would be time to actual diagnosis or resolution of the abnormality, limited data collected during the preintervention period prohibited this comparison. Since no gold standard exists to determine what constitutes timely diagnostic follow-up,[6] we operationalized the concept based on: 1) literature that suggests diagnostic and treatment delays of 3–6 months may impact survival[20]; and 2) review of our data, which indicated that beyond 120 days no appreciable additional follow-up was achieved. Furthermore, repeating our analyses using 90 days as the definition of timely follow-up did not change our findings. To determine 120 day follow-up rates for patients who were seen within the last 120 days for the specified study time period, records beyond the study end date were examined.

Statistical Analyses

Baseline demographic data, including age, race, reason for consultation, site of referral, need for interpreter services, and insurance status, were compared between subjects presenting for care during the preintervention and intervention period. Bivariate comparisons using χ^2 tests of independence tested for differences in timely follow-up between the 2 groups overall, and within specific demographic variables. Logistic regression was conducted to assess the proportion of benefit that can be attributed to the navigator intervention, once demographic variables were taken into account, and 95% confidence intervals were calculated. All analyses were done using PC-SAS version 8.02.[21]

Because baseline differences were noted between the preintervention and intervention group, 2 subsequent analyses were performed to control for differ-

TABLE 1
Characteristics of Patients Before and After Patient Navigation Intervention

Subject characteristic	Total (N = 1332)	Preintervention, (N = 314)	Intervention (N = 1018)	P*
Age, y				
18–39	522 (39)	149 (47)	373 (37)	.0008
40–64	689 (52)	147 (47)	542 (53)	
≥65	121 (9)	18 (6)	103 (10)	
Race				
White	480 (36)	107 (34)	373 (37)	.03
Black	534 (40)	143 (46)	391 (38)	
Hispanic	190 (14)	44 (14)	146 (14)	
Asian	59 (4)	13 (4)	46 (5)	
Other	69 (5)	7 (2)	62 (6)	
Insurance				
Private	423 (32)	82 (26)	341 (34)	.01
Public/none	909 (68)	232 (74)	677 (67)	
Interpreter				
Needed	197 (15)	25 (8)	172 (17)	<.0001
Reason for visit				
Breast mass	630 (47)	117 (37)	513 (50)	<.0001
Abnormal mammogram	158 (12)	29 (9)	129 (13)	
Other	544 (41)	168 (54)	376 (37)	
Source of referral				
CHC	582 (44)	96 (31)	486 (48)	<.0001
Hospital	448 (34)	78 (25)	370 (36)	
Private	105 (8)	13 (4)	92 (9)	
Unknown/other	197 (15)	127 (40)	70 (7)	

CHC indicates community health center. All values given are in number (percentages).
*Comparing preintervention versus intervention periods.

ences in case-mix in the 2 study groups. First, a propensity score analysis was conducted to adjust for differences between the study periods. This methodology selects a random subset of subjects in each study period, so that baseline characteristics of the 2 groups are similar. One intervention subject was chosen per preintervention subject. After demonstrating that the groups were similar in the their baseline characteristics (age, race/ethnicity, insurance, reason for referral, and source of referral), we conducted a multiple logistic regression on the reduced dataset predicting timely follow-up, again adjusting for age, race/ethnicity, insurance, reason for referral, and referral source. Second, time was considered as a covariate in the model to control for longitudinal effects not related to the intervention. Time was calculated as the month of the study when the visit was scheduled. For example, a women who scheduled her appointment during the first month of the study (January 2000) would have a value of time of 1, while a women who was in the intervention period might have a value of time of 25, representing that she scheduled her appointment in January 2002. Three models were considered: including time to the main model including covariates; including an interaction effect between time and intervention in the main model; and considering only the effects of time and intervention on timely follow-up.[22]

RESULTS

During the study period, 2044 scheduled patient visits were identified, which corresponded to 1381 individual patients. Of those patients, 21 (1.5%) men were excluded, 10 (0.7%) visits for female teens under age 18 were excluded, and 14 (1.0%) were excluded due to missing dates. The final analytic sample consisted of 1332 individual women with scheduled visits (314 preintervention, 1018 during the intervention period).

Baseline characteristics by study group are shown in Table 1. The majority of women scheduled for evaluation were under age 65 (91%), with over half (52%) between 40 and 64 years of age. The majority of women were of minority race (40% black, 14% Hispanic, 4% Asian), while 36% were non-Hispanic white. Fifteen percent required a language interpreter during their visit, and most (68%) had no insurance or some type of public health insurance (Medicaid, Medicare only, or uncompensated care coverage). Over half the women were referred for

TABLE 2
Factors Associated With Timely Follow-Up by Logistic Regression and Propensity Score Analysis: Patient Navigation Intervention

Variable	Logistic Regression Odds Ratio*	95% CI	Propensity Score Odds Ratio*	95% CI
Patient navigation intervention	1.4	1.01–1.9	1.7	1.2–2.6
Age, y				
18–39	0.7	0.5–0.9	0.4	0.3–0.6
40–64	1.0	xxx	1.0	xxx
≥65	1.9	1.1–3.4	2.3	0.8–6.4
Insurance				
Private (vs. public)	1.5	1.1–2.1	2.2	1.3–3.7
Race				
White	1.0	xxx	1.0	xxx
Black	0.8	0.6–1.0	0.9	0.6–1.5
Hispanic	1.2	0.8–1.8	2.0	1.0–3.9
Asian	1.5	0.7–3.0	2.1	0.6–7.8
Other	0.6	0.3–1.1	0.4	0.1–1.2
Reason for visit				
Breast mass	0.5	0.3–0.9	0.9	0.4–1.8
Abnormal mammogram	1.0	xxx	1.0	xxx
Other	0.3	0.2–0.5	0.6	0.3–1.1
Source of referral				
CHC	1.0	Xxx	1.0	xxx
Hospital	1.4	1.0–2.0	0.8	0.5–1.3
Private	1.5	0.8–2.7	1.6	0.4–5.7
Unknown/other	0.7	0.5–1.0	0.5	0.3–0.8

CHC indicates community health center.

* Analyses adjusted for all variables listed in table.

evaluation of a screening abnormality (47% abnormal exam, 12% abnormal mammogram). Forty-one percent of women in the "other" category which included breast pain, increased breast cancer risk based on family or personal history of breast cancer, or nipple discharge, respectively. The majority of women were referred from a Community Health Center (CHC; 44%) or a hospital-based practice site (34%).

Table 1 also demonstrates that subjects who presented for evaluation during the intervention period differed significantly from preintervention subjects in most demographic characteristics. Intervention subjects were more likely to be older, white, to have private health insurance coverage, to require an interpreter, to be referred for a screening abnormality and to have been referred from an affiliated CHC.

Overall, 64% of subjects referred for diagnostic evaluation had timely follow-up during the preintervention period ($N = 314$) compared with 78% of women during the intervention period ($N = 1018$) ($P < .0001$, unadjusted OR: 2.0 [95% CI, 1.5–2.6]). Table 2 presents the results of our adjusted analysis. Controlling for age, race, insurance status, reason for referral, and source of referral, women in the intervention group had a 39% greater odds of having timely follow-up (OR = 1.39 [95% CI, 1.01–1.91]). Compared with women aged 40–64, women over 65 years of age were more likely to have timely follow-up (OR of 1.9; 95% CI, 1.1–3.4), while those aged 18–39 were less likely to have timely follow-up (OR of 0.7; 95% CI, 0.5–0.9). Women with private health insurance were more likely to have timely follow-up, compared with those with only public or no health insurance (OR of 1.5; 95% CI, 1.1–2.1). Compared with women referred for evaluation of an abnormal screening mammogram, timely follow-up was less likely among women referred for evaluation of a breast mass (OR of 0.5; 95% CI, 0.3–0.9) or other breast abnormality (OR of 0.3; 95% CI 0.2–0.5). Compared with women referred from community health centers, timely follow-up was more likely among those referred from hospital-based practice sites (OR of 1.4; 95% CI, 1.0–2.0).

Table 2 also shows the results of our propensity score analysis to address the differences in case-mix. Our propensity score analysis selected a subset of 284 preintervention and 284 intervention subjects with similar baseline characteristics, using the same 5 covariates as the main analytic model to calculate a propensity of being in the intervention group. Within each quintile of propensity score, an even

number of preintervention and intervention women were chosen, so that the resulting subsample would be matched on baseline covariates. All covariates were associated with the intervention before sampling, while only source of referral remained associated with the intervention postsampling (*P*-values were .79 for age, .34 for insurance, .41 for race, .0004 for source, and .13 for source). Comparison of these 2 matched subgroups did not change the direction or significance of the intervention effect. Using these reduced data, 65% of preintervention versus 76% of intervention subjects had timely follow-up ($P = .008$). The odds ratio (OR) for the postsampling intervention was 1.7 (95% CI, 1.2–2.6). The only difference noted in the propensity score analysis is a change in effect due to source of referral, specifically from hospital versus CHC, where the effect size changed direction from 1.4 to 0.8. These results can be found in Table 2.

As a second sensitivity analysis, we considered the trend over time. Three models were considered using the time variable. We first added time to the final multiple logistic regression reported in Table 2. In this model, controlling for other covariates, time was not significant ($P = .57$). Second, a time-intervention interaction was considered. Once again, this term was not significant ($P = .41$). Lastly, in recognition of the fact that the effect of the time variable might be confounded within the other covariates, which have changed over time, a model including only time and intervention was evaluated to predict timely follow-up. In this model, once again, the trend over time being associated with timely follow-up was not significant ($P = .34$). These results lead us to conclude that any change over time has been successfully addressed by the modeling on the other covariates.

DISCUSSION

Patient, provider, cultural, and system level factors have been identified as barriers to the provision of effective cancer diagnostic and treatment services in an equitable and timely manner,[23–37] thus leading to health disparities.[38] This is the first publication, to our knowledge, to provide evaluation of a barrier-focused, patient navigation program within an academic medical center, using outcomes data on the group of patients receiving service for breast screening abnormalities. Our data suggest that patient navigation improved rates of timely diagnostic follow-up for abnormal breast cancer screening among a racially diverse group of urban women. We determined that nearly all subgroups of women benefited from the intervention even after adjusting for selection bias. The vast majority of our study population were women from either an ethnic minority population, and/or from a low income group, thus representing the women at greatest risk of poor outcomes from cancer.

Our navigation intervention is modeled after the care management model, with 4 key components to navigation.[19] These include a mechanism for case identification, a process of identifying barriers to follow-up for each individual case, a plan for addressing barriers, and a system of tracking. Each of these components is critical to both accomplish effective navigation, but also to track the outcomes of individual women included in our system. Our system utilized hand entered systems for all aspects of care and tracking, thus demonstrating a method which is easily generalizable to many systems that care for underserved and minority populations.

Care management, used in nursing and social work since the mid 1800s, has been employed for those at increased risk for adverse outcomes and excessive healthcare utilization.[39] The concept has evolved over the years to target various chronic diseases, most notably diabetes,[40] where evidence has led to clinical recommendations for such interventions by medical organizations. There has been growing interest in the use of patient navigators as a mechanism to reduce cancer disparities through care management since the concept was first introduced in the early 1990s.[18] Since then, navigator programs have been implemented across the country with support from private foundations[13] and more recently the federal government[16,41] after 2 bills in Congress[42,43] proposed support for navigation programs to address cancer disparity in underserved populations. To date, however, the data on the effectiveness of patient navigation interventions has been anecdotal or uncontrolled.[18] Although evidence does exist to demonstrate the effectiveness of individual components of our model, such as telephone outreach or tracking,[44–46] intervention studies to improve diagnostic follow-up using patient navigation encompassing the full care management model only exist for women with abnormal Papanicolaou (Pap) tests.[47,48] This marks the first report of an evaluation to investigate the benefit of a patient navigator in an academic medical center serving an urban, minority patient population with screening breast abnormalities.

We evaluated *timely follow-up*, defined as the time from when the first diagnostic breast appointment is scheduled until actual follow-up evaluation was initiated, in a racially diverse group of inner-city women referred for diagnostic breast evaluation at an urban academic medical center. Our study did not permit an evaluation of timeliness to actual diag-

nosis. Rather, the intermediate outcome of time to initiation of diagnostic evaluation was chosen based on the limited data collected in the preintervention period. Currently, there lacks clear and widely accepted definitions of what constitutes timely follow-up after a breast abnormality is detected,[6] which limits our ability to compare our findings with existing literature. We demonstrated a 15% improvement in timely follow-up after program implementation. This increase is consistent with a previously published report of a similar care management intervention for inner-city minority women with abnormal Pap tests.[48]

We identified characteristics of those patients who are most at risk for loss to follow-up. Significant predictors of timely follow-up include older age, having a referral, which originated on-site, having private health insurance, or being referred for evaluation of an abnormal screening mammogram. Women referred from community health centers benefited from the intervention, but not to the same degree as other women. Although transportation or convenience factors could have been operant, our system of care provides regularly scheduled shuttle rides to and from all our affiliated health centers. Rather, the etiology may lie in other patient characteristics not measured in our study, such as fear, mistrust, lack of appreciation of the potential seriousness of the problem, inability to get time off from work, child care issues, etc.

Our model employed a medical assistant with some clinical experience recruited to the navigator position. Although no standard guidelines exist, patient navigation typically utilizes a person who is not the provider of direct healthcare for coordination and implementing care. The literature to date has described navigation across a broad spectrum of experience, including trained lay navigators, to those with advance nursing degrees.[12] Lack of clarity exists for what constitutes patient navigation, as it has multiple components, although a several studies sought to evaluate programs with some components of patient navigation.[46,47] Our data cannot address which level of training is most effective or most cost effective or which components of navigation are critical to successful outcomes. Our data precluded an analysis of the specific intervention components. Future studies should focus on developing measurements of the 4 principles of care management utilized in this study to determine which components have the greatest impact on target populations.

Limitations of our study include the evaluation of the intervention in only one health care system. However, the study had a large sample size and a diverse population referred from multiple sites. Nonetheless, our system has some unique features that may encompass a more responsive system of care for underserved urban communities, limiting generalizability to other health care systems, even with similar underserved communities. Unique features include a managed Medicaid plan, which covers the greater proportion of the otherwise uninsured, and an uncompensated care pool, to support those without any health insurance access otherwise. Furthermore, extensive interpreter services and transportation strategies are in place to allow patients to travel easily from community health centers to the medical center.

Another limitation of this study is the pre–post design, which is vulnerable to secular trends as the explanation of the findings. Although our pre–post design introduces the possibility of historical bias, the above-mentioned resources were in place throughout the study duration. Furthermore, there were no major institutional, city- or statewide breast cancer diagnostic programs implemented during the study period to introduce secular trends as a potential explanation for our findings. A propensity score analysis to adjust for known baseline differences in the 2 groups found the same magnitude of effect of the intervention. A time trend analysis was also conducted and identified no significant time trend. For these reasons, it is less likely that the observed differences in case-mix in the 2 study periods are attributable to changes in the healthcare system that are being confused with a navigator effect.

Our data suggest a benefit of patient navigators for addressing follow-up after abnormal cancer screening. Patient navigation shows promise as the missing link between available cancer care services and delivery to vulnerable populations. Future funding of patient navigation is critical to continue more rigorous evaluation efforts, especially to address which type of navigator and what components of navigation are most effective. Furthermore, understanding the work design issues and how to coordinate patient navigator efforts across specialists and disease conditions is especially timely, as multiple venues of the health care system begin to employ navigators. These data will be critical in policy decisions on incorporating support through insurers to accomplish cancer care management for vulnerable populations.

REFERENCES

1. Ries LE, Eisner MP, Kosary CL, et al., eds. SEER Cancer Statistics Review, 1975–2000. Bethesda, MD: National Cancer Institute; 2003.

2. McCarthy EP, Burns RB, Coughlin SS, et al. Mammography use helps to explain differences in breast cancer stage at diagnosis between older black and white women. *Ann Intern Med*. 1998;128:729–736.
3. Jacobellis J, Cutter G. Mammography screening and difference in stage of disease by race/ethnicity. *Am J Public Health*. 2002;92:1144–1150.
4. Jones BA, Kasl SV, Curnen MG, Owens PH, Dubrow R. Can mammography screening explain the racial difference in stage at diagnosis of breast cancer? *Cancer*. 1995;75:2103–2113.
5. Jones BA, Patterson EA, Calvocoressi L. Mammography screening in African American women: Evaluating the research. *Cancer*. 2003;97 (Suppl 1):258–272.
6. Bastani R, Yabroff KR, Myers RE, Glenn B. Interventions to improve follow-up of abnormal findings in cancer screening. *Cancer*. 2004;101 (Suppl 5):1188–1200.
7. Jones BA, Dailey A, Calvocoressi L, Reams K, Kasl SV, Lee CH, Hsu H. Inadequate follow-up of abnormal screening mammogram: Findings from the race differences in screening mammography process study. *Cancer Causes Control*. 2005; 16:809–821.
8. Caplan LS, May DS, Richardson LC. Time to diagnosis and treatment of breast cancer: Results from the National Breast and Cervical Cancer Early Detection Program, 1991–1995. *Am J Public Health*. 2000;90:130–134.
9. Chang SW, Kerlikowske K, Napoles-Springer A, Posner SF, Sickles EA, Perez-Stable EJ. Racial differences in timeliness of follow-up after abnormal screening mammography. *Cancer*. 1996;78:1395–1402.
10. Kerlikowske K. Timeliness of follow-up after abnormal screening mammography. *Breast Cancer Res Treat*. 1996;40:53–64.
11. Webber PA, Fox P, Zhang X, Pond M. An examination of differential follow-up rates in breast cancer screening. *J Community Health*. 1996;21:123–132.
12. Centers for Medicare and Medicaid Services. Evidence report and evidence-based recommendations: Cancer prevention and treatment demonstration for ethnic and racial minorities. Bethesda, MD: United States Department of Health and Human Services; 2003.
13. Dohan D, Schrag D. Using navigators to improve care of underserved patients. *Cancer*. 2005;104:848–855.
14. Healthcare Association of New York State (HANYS). Breast health patient navigator resource kit; 2002. Available at www.hanys.org. Accessed August 15, 2005.
15. Capitman JA, Bhalotra S, Ruwe M. Cancer and Elders of Color: Opportunities for Reducing Health Disparities. Aldershot, Hampshire, England: Ashgate; 2005.
16. National Institutes of Health. Patient navigation research program; 2004. Available from http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-05-019.html. Accessed August 2005.
17. Avon Foundation Breast Cancer Crusade. Available at www.avoncompany.com/women/avoncrusade/2005_BC_funding_guidelines.pdf. Accessed August 28, 2005.
18. 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.
19. Longest BB, Young GJ. Coordination and Communication. In: Shortell SK, Kaluzny AD, eds. Health Care Management: Organization Design and Behavior, 4th edn. Albany, NY: Delmar Publishers; 2000:237–275.
20. Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: A systematic review. *Lancet*. 1999;353:1119–1126.
21. SAS Institute. SAS release 8.02. Cary, NC: SAS Institute; 1999–2001.
22. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther*. 2002; 27:299–309.
23. Benard VB, Lee NC, Piper M, Richardson L. Race-specific results of Papanicolaou testing and the rate of cervical neoplasia in the National Breast and Cervical Cancer Early Detection Program, 1991–1998 (United States). *Cancer Causes Control*. 2001;12:61–68.
24. Bennetts A, Irwig L, Oldenburg B, et al. PEAPS-Q: A questionnaire to measure the psychosocial effects of having an abnormal pap smear. Psychosocial effects of abnormal Pap smears questionnaire. *J Clin Epidemiol*. 1995;48:1235–1243.
25. Brett J, Austoker J. Women who are recalled for further investigation for breast screening: Psychological consequences 3 years after recall and factors affecting reattendance. *J Public Health Med*. 2001;23:292–300.
26. Carey P, Gjerdingen DK. Follow-up of abnormal Papanicolaou smears among women of different races. *J Fam Pract*. 1993;37:583–587.
27. Crane LA. Social support and adherence behavior among women with abnormal Pap smears. *J Cancer Educ*. 1996; 11:164–173.
28. Deane KA, Degner LF. Information needs, uncertainty, and anxiety in women who had a breast biopsy with benign outcome. *Cancer Nurs*. 1998;21:117–126.
29. Haas J, Kaplan C, McMillan A, Esserman LJ. Does timely assessment affect the anxiety associated with an abnormal mammogram result? *J Womens Health Gend Based Med*. 2001;10:599–605.
30. Hunt LM, De Voogd KB, Soucy MD, Longworth JC. Exploring loss to follow-up: Abnormal Pap screening in Hispanic patients. *Cancer Pract*. 2002;10:122–129.
31. Lauver D, Rubin M. Women's concerns about abnormal Papanicolaou test results. *J Obstet Gynecol Neonatal Nurs*. 1991;20:154–159.
32. Lauver DR, Kruse K, Baggot A. Women's uncertainties, coping, and moods regarding abnormal Papanicolaou results. *J Womens Health Gend Based Med*. 1999;8:1103–1112.
33. McCarthy BD, Yood MU, Janz NK, Boohaker EA, Ward RE, Johnson CC. Evaluation of factors potentially associated with inadequate follow-up of mammographic abnormalities. *Cancer*. 1996;77:2070–2076.
34. McKee MD, Caban A, Burton W, Mulvihill M. Women's knowledge and experience of atypical Pap results in a high risk community. *Women Health*. 2002;36:19–31.
35. Rojas M, Mandelblatt J, Cagney K, Kerner J, Freeman H. Barriers to follow-up of abnormal screening mammograms among low-income minority women. Cancer Control Center of Harlem. *Ethn Health*. 1996;1:221–228.
36. Shinn E, Basen-Engquist K, Le T, et al. Distress after an abnormal Pap smear result: Scale development and psychometric validation. *Prev Med*. 2004;39:404–412.
37. Zapka JG, Puleo E, Taplin SH, et al. Processes of care in cervical and breast cancer screening and follow-up the importance of communication. *Prev Med*. 2004;39:81–90.
38. Institute of Medicine. Crossing the Quality Chiasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.

39. Ward MD, Rieve JA. The role of case management in disease management. In: Todd WE, Nash E, eds. Disease Management: A Systems Approach to Improving Patient Outcomes. Chicago, IL: American Hospital Publishing; 1997: 235–259.
40. Norris SL, Nichols PJ, Caspersen CJ, et al., and the Task Force on Community Preventive Services. The effectiveness of disease and case management for people with diabetes. A systematic review. *Am J Prev Med*. 2002;22 (Suppl 4):15–38.
41. Centers for Medicare and Medicaid Services. Cancer Prevention and Treatment Demonstration for Ethnic and Racial Minorities 2005. Available at www.cms.hhs.gov/researchers/demos/CPTD/FRN12_23_04.pdf.
42. United States Senate. House Committee Report. Patient Navigator Outreach and Chronic Disease Prevention Act of 2005, 109th Congress, 1st session, 2005.
43. United States Senate.House Committee Report. Patient Navigator Outreach and Chronic Disease Prevention Act of 2003, 108th Congress, 1st session, 2003.
44. Champion V, Maraj M, Hui S, Perkins AJ, Tierney W, Menon U, Skinner CS. Comparisons of tailored interventions to increase mammography screening in nonadherent older women. *Prev Med*. 2003:36;150–158.
45. Olivotto IA, Borugian MJ, Kan L, et al. Improving the time to diagnosis after an abnormal screening mammogram. *Can J Public Health*. 2001;92:366–371.
46. Manfredi C, Lacey L, Warnecke R. Results of an intervention to improve compliance with referrals for evaluation of suspected malignancies at neighborhood public health centers. *Am J Public Health*. 1990;80:85–87.
47. Engelstad LP, Stewart SL, Nguyen BH, Bedeian KL, Rubin MM, Pasick RJ, Hiatt RA. Abnormal Pap smear follow-up in a high-risk population. *Cancer Epi Bio Prev*. 2001:10;1015–1020.
48. Ell K, Vourlekis B, Muderspach L, et al. Abnormal cervical screen follow-up among low income Latinas: Project safe. *J Womens Health Gend Based Med*. 2002;11:639–651.

Breast Reconstruction following Mastectomy for Breast Cancer: The Decisions of Sexual Minority Women

Ulrike Boehmer, Ph.D.
Rhonda Linde, Ph.D.
Karen M. Freund, M.D., M.P.H.

Boston and Bedford, Mass.

Background: Prior research on decision-making for reconstructive surgery after mastectomy has not addressed the specific considerations of sexual minority women (women who partner with women, and lesbian or bisexual identified women). The purpose of this study is to explore which issues sexual minority women considered when making decisions on reconstructive surgery and to understand the influence and perspectives of these women's most important support persons.

Methods: Study participants were recruited through targeted community-based sampling. The authors conducted individual semistructured interviews with 15 sexual minority women who had been treated with mastectomy after breast cancer diagnosis and 12 support persons who were identified by these women as their most important source of support. Using qualitative data analysis software, transcribed interviews were analyzed. Through constant comparison methods, themes related to the decision on and experiences and satisfaction with reconstructive choice were identified from the narrative data.

Results: The considerations of women who decided for or against reconstruction are rooted in a value system and body image shaped by their sexual minority identity. Women who chose reconstruction experienced difficulties and regrets, whereas women without reconstruction adjusted well after time. Partners of sexual minority women matched the level of satisfaction with reconstructive choice achieved by the women themselves.

Conclusion: Providers who treat sexual minority women might benefit from knowing about issues important to this population to provide more comprehensive care. (*Plast. Reconstr. Surg.* 119: 464, 2007.)

The majority of women treated with mastectomy after breast cancer diagnosis forego reconstructive surgery.[1,2] There is inconsistent evidence that benefits are derived from undergoing reconstruction.[3–7] Determining what influences patients to elect reconstruction has been the focus of a number of studies.[8–14] Women's decision-making on reconstruction is still not entirely understood, but intrapersonal and interpersonal factors such as body image, self-concept, total self-image, and women's partner relationship have been identified as influences on women's reconstructive choices.[7,14,15] Women who chose breast reconstruction reported wanting to get rid of a breast prosthesis, freedom to wear different types of clothing, regaining their femininity, and a desire to feel whole again, whereas women with breast prostheses reported reconstruction not to be essential for their physical or emotional well-being, not wanting a foreign object in their body.[13] Although women weighed these issues when deciding on reconstructive surgery, studies also concluded that the two groups of women do not significantly differ regarding body image or positive self-concept.[3,8,9] However, body image is a multidimensional concept and has been defined differently by researchers.[16] The complexity of this construct continues to generate hypotheses to ex-

From the Department of Health Services, Boston University School of Public Health; Center for Health Quality, Outcomes, and Economic Research; Fenway Institute, Fenway Community Health; and Women's Health Unit, Evans Department of Medicine and Women's Health Interdisciplinary Research Center, Boston University School of Medicine.
Received for publication June 15, 2005; accepted August 24, 2005.
The views expressed in this article are those of the authors and do not necessarily represent the views of the Massachusetts Department of Public Health, the American Cancer Society, or the Department of Veterans Affairs.

DOI: 10.1097/01.prs.0000246402.79334.3b

plain the rejection of reconstruction. These include the following: women who make this choice have a more resilient body image, are less invested in their physical appearance, or are in more denial about their body image; however, the empirical evidence for these hypotheses is mostly lacking.[7]

Reconstructive surgery or the desire for reconstruction has also been linked to demographic characteristics, in that older and poorer women undergo significantly fewer reconstructions,[1,15] and African American and Asian women undergo reconstructive surgery less often than white women.[17] We are not aware of any study considering the influence of sexual orientation on electing reconstructive surgery, although one study of breast cancer patients found that lesbians have a better body image compared with heterosexual patients.[18] The evidence is inconsistent that women's body image, self-concept, and perceptions of attractiveness differ by sexual orientation in the general population.[19–21] To date, a prevailing heterosexual framework in studies on reconstructive surgery considered only women's male partners' perceptions of reconstruction, whereas sexual minority women's partners or social support persons have been neglected. Surgeons would benefit from understanding factors that lead to surgical decisions so that patients can be appropriately counseled and reconstruction options discussed. Body image is central for decisions on this surgical choice,[22] and we hypothesize that sexual minority women consider different dimensions of body image in making decisions about reconstruction. If this is true, knowledge of sexual orientation as part of the medical history may benefit surgeons and patients in their discussion of management options.

PATIENTS AND METHODS

We conducted a retrospective qualitative study of sexual minority women with breast cancer and their support persons. Qualitative methods are uniquely suited to this type of exploratory inquiry, especially where traditional pencil and paper surveys have not been tested or validated in the unique population in question.[20] We relied on community-based purposive sampling, which has been widely used to overcome the challenges of recruiting members of vulnerable or "hidden" populations into research studies,[23,24] and is well-suited for a comprehensive exploration of sexual minority women's experiences with breast cancer. Recruitment has been further enhanced through the use of snowball sampling, where participants are asked to refer others who may be willing to participate in the study.

We obtained institutional human subjects approval for this study. Participants received $20 for their participation. After consent was obtained, we conducted one tape-recorded, in-depth, semi-structured interview that lasted on average 90 minutes. The narrative interviews were conducted to investigate the following broad domains: diagnosis, treatment course, treatment decision-making, sources of social support, and adjustment. We asked women to identify their "trusted other," which we defined as their most important support person with respect to their cancer care, someone other than their treating physician.[25] Having a support person was not required for study entry, and the gender and sexual orientation of the support person was unrestricted. We contacted the support person, obtained consent, and conducted a narrative interview that covered the same domains as the interview with the women themselves, to elicit the support person's perspective and level of involvement. We also collected demographic and cancer-related information (stage and date of diagnosis) on participants through a self-administered questionnaire.

Women with breast cancer were eligible for this study if they met the following eligibility criteria: (1) sexual minority status, (2) a diagnosis of nonrecurrent and nonmetastatic breast cancer, (3) having completed invasive treatment (surgery, radiation, chemotherapy) for breast cancer within the past 5 years, and (4) English proficiency. We defined sexual minority status as stating a lesbian or bisexual identity. We included women who reported partnering with women in an attempt to be inclusive of women who might feel uncomfortable embracing a lesbian or bisexual identity. Women who were currently using hormonal treatment, such as tamoxifen, were included in this study.

Measures

History of the disease and treatment were derived from both the questionnaire and interviews. The questionnaire inquired about the demographics of participants. Race was categorized into white, African American, Hispanic, Asian, American Indian, mixed race, or other. Income was determined by providing respondents with the following ranges: less than $20,000, $10,000 increments to $99,999, and over $100,000. Other questions assessed age, education, health insurance, and employment. In the absence of an agreed-on standardized measure of sexual orientation,[26] we

defined our population using two of three dimensions (identity and behavior, but not desire) of sexual orientation. Respondents were asked to report the sexual identity or the relationship behavior that best described them, allowing for the following responses: lesbian, bisexual, partner with women, heterosexual, or other. Disclosure to others in their social network was assessed by asking participants to report whether sexual orientation was disclosed (1), disclosed and openly talked about (2), or not disclosed (0) to each of the following 12 person groups: children, grandchildren, nieces and nephews, parents, grandparents, siblings other blood relatives, heterosexual friends, colleagues at work, one's boss, neighbors, or heterosexuals in general. The possible range on this disclosure scale was 0 to 24, whereby higher scores indicated a greater level of disclosure.

Analysis

Summary statistics, including means or percentage distribution, were calculated to describe participants' demographic and medical characteristics (Tables 1 and 2). The audio-recorded interview data were transcribed verbatim and then analyzed according to grounded theory methods of qualitative analysis.[27] Grounded theory methodology offers a comprehensive and systematic framework for developing a theory of understanding about participants' experiences that is iteratively assessed from the narrative data. The analysis summarized in this article focused on participants' decisions and thoughts about reconstructive surgery. Passages referring to mastectomy and reconstructive surgery were marked and identified by descriptive code words. Two of the authors (U.B. and R.L.) developed the codes based on independent coding of a subset of five interviews with women with breast cancer and a subset of five interviews with support persons. Once agreement on the codes and reliability of coding was ensured, all remaining interviews were coded.

RESULTS

Of the 15 sexual minority women treated with mastectomy, eight chose reconstructive surgery and seven decided against it. Mastectomy was the recommended surgical treatment for some, others chose mastectomy to avoid radiation, and approximately half initially underwent breast-conserving surgery or even repeated lumpectomies before mastectomy.

The women who did and who did not undergo reconstruction were similar with respect to cancer stage, receipt of adjuvant therapy, and time since diagnosis (Table 1). Four of the seven women who decided against reconstruction each identified one support person, whereas three women did not identify such a person. Six of the eight women who chose reconstruction identified one support person, and one identified two. All support persons were female. The four support persons in the no-reconstruction group were these women's partners. The eight support persons in the reconstruction group were five partners and the mother, sister, and friend of women with breast cancer. Comparisons of demographic characteristics indicated that women in the two reconstruction groups were similar, as were the two groups of support persons (Table 2).

Certain themes were considerations for all women regardless of the choice they made. Breast size was one such recurrent theme when women considered reconstruction. Although breast size entered into women's decision-making, their actual breast size was not an absolute determinant of a particular choice. However, women with small breasts felt comfortable rejecting reconstruction, arguing that it would not make much of a visible difference in their appearance. Large-breasted women struggled with the implications of a mas-

Table 1. Clinical Characteristics of Women with Breast Cancer

Characteristic	Without Reconstruction (n = 7) (%)	With Reconstruction (*n* = 8) (%)
Mean No. of months since diagnosis	31.9	17.4
Stage of breast cancer: in situ	2 (28.6)	2 (25.0)
I	1 (14.3)	1 (12.5)
II	3 (42.9)	3 (37.5)
Unknown, "early stage"	1 (14.3)	2 (25.0)
Chemotherapy	4 (57.1)	5 (62.5)
Radiation	2 (28.6)	3 (37.5)
Hormonal treatment	4 (57.1)	6 (75.0)
Breast reconstruction procedure		
TRAM flap	0	3 (37.5)
Saline implants	0	5 (62.5)

TRAM, transverse rectus abdominis musculocutaneous.

Table 2. Demographic Characteristics of Women and Their Support Persons

Characteristic	No Reconstruction (n = 7) (%)	Reconstruction (*n* = 8) (%)	Supports for No Reconstruction (*n* = 4) (%)	Supports for Reconstruction (*n* = 8) (%)
Age, years				
Mean	50.7	47.3	48.5	50.4
Range	43–61	41–53	42–57	40–63
Race/ethnicity				
White	7 (100)	7 (87.5)	3 (75.0)	7 (87.5)
Latina	0	0	0	0
African American	0	1 (12.5)	0	1 (12.5)
Asian	0	0	1 (25.0)	0
Education				
High school	0	0	0	0
College	4 (57.1)	5 (62.5)	1 (25.0)	4 (50.0)
Graduate school	3 (42.9)	3 (37.5)	3 (75.0)	4 (50.0)
Employed	5 (71.4)	7 (87.5)	3 (75.0)	6 (75.0)
Income <$20,000 to >$100,000*				
Mean for one person	$57,857, *n* = 7	$52,000, *n* = 5	$73,333, *n* = 3	$44,167, *n* = 6
Mean for more than one person	0	$41,667, *n* = 3	$55,000, *n* = 1	$100,000, *n* = 2
Health insurance	6 (85.7)	8 (100)	75.0 (3)	100 (8)
Sexual orientation				
Lesbian	6 (85.7)	6 (75.0)	3 (75.0)	5 (62.5)
Bisexual	0	1 (12.5)	1 (25.0)	0
Partner with women	1 (14.3)	1 (12.5)	0	0
Heterosexual	0	0	0	3 (37.5)
Disclosure scale mean†	19.3	17.4	16.6	15.9

*Income was collected as a range (e.g., $30,000 to $39,999). To present the mean income as a dollar amount, we used $20,000 as the minimum, $100,000 as the maximum, and the midpoint of each income range (e.g., $35,000).
†The possible range of disclosure was 0 to 24, whereby 0 indicates no disclosure to any of 12 (e.g., family, friends) person groups and 24 indicates complete disclosure, in that the sexual orientation is talked about with all 12 person groups.

tectomy, fears of being unbalanced, and having to wear a heavy prosthesis, which made some of these women choose reconstruction in combination with breast reduction.

The importance of their breast for their self-image was another issue that was mentioned by women who did and did not elect reconstructive surgery. The notion of not being defined by having breasts was closely aligned with a sexual minority (i.e., lesbian identity). Body strength and physical functioning were more important considerations than aesthetic considerations in those interviewed. In the context of women's communications with breast cancer care providers and plastic surgeons, in particular, they noted their "otherness"—the discrepancy of their values and body image with those of mainstream society, as demonstrated by the following quote by one woman who decided for reconstruction:

> *It wasn't that he [plastic surgeon] was—it had to do with me being a lesbian, it was that lesbians sometimes have different views about body image that are different from—what do you call it?—straight world. They have different—it's a wider—we have a wider range of —I don't know what the—how to describe it.*

Similarly, another woman who decided against reconstruction described her displeasure with the information she received on reconstruction. Although she was concerned with the implications on her physical functioning, reconstruction's cosmetic advantages were emphasized:

> *The whole decision about reconstruction or not was such a weird experience for me, because when I would go in and talk with the physicians—with the surgeon—the plastic surgeons—about reconstruction, they'd be talking about, like, all these things that didn't make any sense to me. They were talking about what it looked like, they were talking about droop, want to make sure that your droop is just right, and if we do this—it's, like, I don't care about droop. Tell me about how it affects my body. Like, am I going to be—What can and can't I do? Will I be able to move—and I was asking about sort of how it would limit my ability to do whatever I wanted to do, whether it was skiing or golf or, you know, running around—and they were talking about the cosmetic side of it. It's like, I don't care about the cosmetic side. It's not important to me. And they couldn't, like, get beyond that. And I realized that I didn't need to talk to them, because I didn't need reconstruction because I didn't care about that.*

Regardless of their surgical choice, both groups of women perceived their values with respect to body image as different and rooted in their sexual minority identity. Women who de-

cided against reconstruction emphasized strength, long-term health, and survival, rather than fitting a stereotypical beauty ideal. At times, this group of women portrayed reconstructive surgery as a "straight woman's" choice, emphasizing that the procedure is unnecessary in that it does not provide for any curative benefit and requires additional operations and, in the case of the transverse rectus abdominis musculocutaneous flap procedure, compromising "healthy" body parts.

It is of note that among the women who decided for reconstruction, a similar value system and body image prevailed. One way in which women from the reconstructive surgery group expressed this was in their choice of reconstructive procedure. Avoidance of interfering with healthy body parts was generally the reason for choosing the expander and implant procedure over the transverse rectus abdominis musculocutaneous flap procedure:

> *. . .there were either two kinds of reconstruction surgery that can be done; one is put in the expander the other one is transflap when they use some tissue from the stomach or your back. And that I knew I definitely did not want to do. I didn't want to mess up other parts of my body that were basically okay. So I decided to do the tissue expander and have an implant put in.*

The decision on nipple reconstruction also shows these women's body image values. Several women declined to complete their reconstructive surgery process by foregoing nipple reconstruction, reasoning that it had merely cosmetic value.

> *I decided not to do the—You know, you can do a same way, they can put the nipple in and then they do a—I was going to say branding. That's not what I mean. Anyway, it doesn't really matter. And I just figured, "If they're not going to act like nipples, then I don't—What do I need them for?" So I didn't need them.*

Some themes were linked to a particular reconstructive choice. One such theme, linked to the rejection of reconstruction, was having seen another woman's breast reconstruction and disliking the appearance. Other themes, linked to choosing reconstruction, were avoiding depression, regaining a "normal" appearance, and covering up the physical effects of cancer:

> *I was very, very 50-50 for a long time, even after talking with him [plastic surgeon] about whether I would go one-breasted, or have reconstruction. And I decided to go with the reconstruction, because sort of like for the same reasons I went ahead and got a wig before I knew whether I was going to lose my hair or not, because I wanted the freedom of being able to pass if I needed to; of being able to be in public, and not be a cancer person—you know, a cancer patient.*

Partners to women without reconstruction were in agreement with the body image and value system that motivated this choice. Women who decided against reconstruction declared with great confidence that a female partner is different from a husband, in that her partner agrees with her, or that the decision was made jointly, and that she is as attractive to and desired by her partner as before. Their partners confirmed these statements, as in the following example:

> *. . .she didn't want to do reconstruction. We sort of agonized through that one, too, but the options weren't really realistic for her. She didn't want to do any kind of artificial implant. And they really didn't recommend that anyway for someone her size. And it would have been where they take the flesh from the stomach or the abdomen or the back, and either one probably would have been incapacitating to her in the future for her work.. . . And she wasn't so tied up in her looks that that mattered. And that didn't matter to me either.*

This type of joint decision-making and the concordance in values and body image in couples did not exist to the same extent in couples that chose reconstruction. Partners to women with reconstruction were more passive in the decision-making process and couples displayed at times discordance in values and body image, as in the following example:

> *I wanted her to do exactly what she wanted to do. But deep inside, I was really wanting her to get a bilateral mastectomy without reconstruction.. . . Well, she pretty much right from the beginning was very on the track of she wanted complete reconstruction, because I did at one point try to say, "You know, don't do it right off. See how you feel about it. Leave it the way it is. It's one surgery. And then you don't have to go back for the second surgery." But she really wanted the reconstruction.*

Women who were without a partner at the time of decision-making on breast reconstruction included in their considerations how their choice would affect their dating and a potential partner. Although being single does not automatically determine reconstruction as the choice, single women who chose it often reasoned that their current relationship status influenced their choice. One woman explained, "I'm glad that I

decided to have reconstructive surgery. I think that's a choice that every woman has to make, but since I'm a single woman, I felt more complete with two breasts."

The support persons who were not in a partner relationship reported more reluctance to express their reconstructive preferences. One support person explained to the interviewer that she personally was opposed to reconstruction yet was uncertain whether she ever shared her opinion with her friend.

Women who underwent reconstructive surgery reported considerable discomfort with the reconstruction procedure. Physical problems included numbness and limited range of motion. Because of complications, one woman had the expander removed and another woman decided to reverse her reconstruction permanently. Other women expressed doubts or regrets about their choice and lack of information about potential complications:

> *This reconstruction, the transflap and that flap, it's hard to know. I mean, I'm all—I've got scars here, I've got sort of—I have scar tissue here or something, is numb in places. This is more problematic for me, this one, where they took the latissimus dorsi and lifted around. I've got tissue that's stuck to my ribs and so it's uncomfortable. You know. And I don't know—I think sometimes—like I think if somebody said to me. . .. You know, if I had to do it over again, I don't know what I would choose.*

The cosmetic outcome was satisfying to some women and unsatisfying to others, yet their body image was changed:

> *The cosmetic result I am pleased with. When they first did the surgery, I looked and I cried. That was after the mastectomy. There was like a little breast bud there because they had put in the beginning expander and I wasn't sure how it was going to turn out. When they finally did the reconstruction, it didn't look as I expected it was going to. And I had heard from people that it usually takes some time to settle in and looks better. But I cried a couple of times just because of the change and . . . it didn't look like the other breast that remained. And now it looks pretty close. But I think that the change in the body image, the way I talk about it is it's kind of like when you get a new car. And everything is fine before it has any little dents. Then it gets a small dent. And now I have a big dent.*

Partners to women with reconstruction generally echoed the sentiments, including the doubts and misgivings about reconstruction. Non-partner support persons were not informed about the minute details of how women experienced their reconstructed breast. The partner to a woman who was dissatisfied with the outcome, completely shared the dissatisfaction:

> *. . .she's having difficulty with because she doesn't like the outcome of the surgery and she doesn't like how they look or how they feel and they're very hard and they feel like a wall and that's because the implants are underneath the pecs, so when she contracts her pecs, that's all you see is movement of pecs. It's not like breast tissue. And she has a lot of sensory loss, and they're not even, and they're lumpy, and they feel, you know, lumpy, and they feel kind of like a very thick like bag inside of skin. You can feel like this ripple. So it's just like there is nothing sensual about them. There's nothing that makes me want to touch them, you know. And plus her reactions to it are so like it's almost like she freezes.*

Partners confirmed that outcomes of reconstruction or what to expect had not been thoroughly explained by the plastic surgeon to the couple; rather, they were sold on positives they did not experience. Women without reconstruction place less emphasis on physical problems. Physical problems they mentioned consisted of limited range of motion and having loss of sensation caused by nerve damage. Adjusting to their changed body image required some time, and some commented on mastectomy being a barbaric procedure. Their changed body shape caused some to make adjustments either initially or permanently. Adjustments consisted of wearing a bra by women who were not used to wearing one before and making decisions about when to use prostheses. Depending on where women were in their adjustment process, women expressed different levels of comfort with their changed body:

> *. . .originally I thought I'd have to go find another job, and I went and got a—you know, spent $400 on this prosthetic, which when I wear it, I look fine. And I always hated wearing bras; I always—you know, I was never in love with my breast. I love other women's breasts, but I started—early on, I would wear the prosthetic when I was going out in public. I never wear it now at all. And it's like last May I was going to visit my sister, who I hadn't seen for five years, and she lives up in [U.S. state], and has a kid, and a new husband, and I debated about, well, should I take it and wear it? And I said—and my friends said. . . do what makes you feel comfortable—again, all this incredible support. And I realized that I don't need to wear this for someone else.*

Women who decided against reconstruction commented on how they transformed their changed body shape into something positive for them. Women commented on loving their mastectomy scar and on experiencing the sensitivity they used to experience in their breast now in their scar:

> *I don't even think of it now that I don't have breasts. It took me about a year, though, to kind of get used to that different body image, but I literally was so excited to not have them. I mean it sounds crazy. My chest looks like the nine-year-old boy I always wished I could have been, and so from that aspect it's absolutely satisfying to me that I don't. I just hardly even think about it anymore.*

Contrary to women who chose reconstruction, women voiced no doubts or regrets about their choice. Partners of women without reconstruction echoed the level of satisfaction that the women themselves achieved with respect to their body image, just as the partners in the reconstruction group echoed the level of dissatisfaction. No-reconstruction partners commented on adjustments such as wearing bras and prostheses. One partner of a woman who adjusted well to her changed body, mirrored her satisfaction:

> *I love her. The breast not being there is just as beautiful as it was when the breast was there. . .. No, it hasn't changed anything as far as us making love together, being loving. No. I think she's as beautiful as she's ever been and in fact maybe more beautiful. She just gets better looking with age.*

DISCUSSION

This is the first study we are aware of that focuses on sexual minority women's preferences with respect to reconstructive surgery. Our study points to sexual minority women's value systems and body images that prioritize a sense of well-being including body strength, survival, and physical functioning over outward appearance or normative beauty standards. Body image is an inconsistently defined term,[16] yet studies that measure cancer-related body image prioritize cosmetic outcomes and physical appearance,[16,28,29] with few exceptions,[18,30] and essentially neglect the functional values sexual minority women described as most pertinent.

Sexual minority women who decided against reconstruction did not feel it was necessary for their well-being, whereas those who chose reconstruction hoped it would prevent depression triggered by mastectomy. Both those with and those without reconstructive surgery applied their values and a body image that prioritizes well-being over external appearance. This confirms some research on sexual orientation and body image that suggests sexual minority women place less emphasis on external appearance compared with heterosexual women, sometimes described as sexual minority women's resistance to the dominant societal beauty ideals.[20,21]

Women in our study who underwent reconstruction reported discomfort, regrets, doubts, and discontentment with reconstruction, whereas women without reconstruction seemed more satisfied with their choice and outcome. One possible interpretation is that those choosing reconstruction had expectations that reconstruction would conceal having had breast cancer and that they did not achieve this expectation. This finding and interpretation concur with the results of presumed heterosexual women without reconstruction who rated their satisfaction higher than women with reconstruction,[31] possibly indicating that women may use different criteria on which to base their satisfaction rating. It has been suggested that women with mastectomy alone may be satisfied because they achieved their goal of removing the cancer, whereas women with reconstruction may rate their satisfaction lower because they had to undergo multiple operations.[31] In our earlier assessment of a larger cohort of sexual minority women, we did not find differences in coping and adjustment by reconstructive choice,[32] yet we did not assess body image or satisfaction with the decision in these two reconstructive groups.

Our study contributes new understandings about the perspective and influences of support persons on sexual minority women's decision-making. Partners shared in decision-making of women with breast cancer and played an important role in supporting the decision made by their partners. Women with breast cancer had great confidence that their partner would love them regardless of their physical appearance after mastectomy.

There are several limitations of this study. We relied on a small convenience sample of sexual minority women and their support persons recruited from the community. The sampling procedure and the sample size limit us in several ways. First, we do not suggest making inferences from this study about the prevalence of reconstruction in sexual minority women. Second, we are presenting themes and concerns that entered into the decision-making process on reconstruction and that are connected to the perception and satisfaction with reconstruction

choices, yet we do not propose assumptions about the prevalence of the themes in this particular population. Third, we are aware that our retrospective design is subject to recall bias, which has been a critique of other studies that evaluated satisfaction with reconstruction.[3,31]

Nevertheless, these findings have implications for the discussion about reconstruction options with patients. Consistent with the literature indicating that positive patient–provider communication improves outcomes,[29,33–35] we propose that plastic surgeons inquire about patients' values and important issues with respect to body image. Reconstructive procedures should be presented comprehensively and clearly to indicate to what degree different procedures will match a patient's expectations. Our earlier research indicated that breast cancer providers should inquire about the sexual orientation of their patients rather than assume all patients are heterosexual.[36] Including partners or other support persons in the decision-making discussion may greatly benefit the patient. Also, the concerns, processes, and experiences described by the participants of this study can inform sexual minority patients who are facing this surgical decision in the future.

Ulrike Boehmer, Ph.D.
Center for Health Quality, Outcomes, and Economic Research
200 Springs Road, Building 70
Bedford, Mass. 01730
boehmer@bu.edu

ACKNOWLEDGMENTS

Support for this research was provided by the Massachusetts Department of Public Health Cancer Research Program, FY02, and the American Cancer Society, ROG-03-105-01 (U.B.). Institutional support and work space for the research was provided by the Department of Veterans Affairs, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, Massachusetts. The authors are grateful to the participants who shared their thoughts and experiences.

DISCLOSURE

None of the authors has any conflict of interest resulting from any financial interest related to the procedures mentioned in this article.

REFERENCES

1. Polednak, A. P. How frequent is postmastectomy breast reconstructive surgery? A study linking two statewide databases. *Plast. Reconstr. Surg.* 108: 73, 2001.
2. Alderman, A. K., McMahon, L., Jr., and Wilkins, E. G. The national utilization of immediate and early delayed breast reconstruction and the effect of sociodemographic factors. *Plast. Reconstr. Surg.* 111: 695, 2003.
3. Harcourt, D., and Rumsey, N. Psychological aspects of breast reconstruction: A review of the literature. *J. Adv. Nurs.* 35: 477, 2001.
4. Nissen, M. J., Swenson, K. K., Ritz, L. J., et al. Quality of life after breast carcinoma surgery: A comparison of three surgical procedures. *Cancer* 91: 1238, 2001.
5. Rowland, J. H., Desmond, K. A., Meyerowitz, B. E., et al. Role of breast reconstructive surgery in physical and emotional outcomes among breast cancer survivors. *J. Natl. Cancer Inst.* 92: 1422, 2000.
6. Al-Ghazal, S. K., Fallowfield, L., and Blamey, R. W. Comparison of psychological aspects and patient satisfaction following breast conserving surgery, simple mastectomy and breast reconstruction. *Eur. J. Cancer* 36: 1938, 2000.
7. Pruzinsky, T. Body image adaptation to reconstructive surgery for acquired disfigurement. In T. F. Cash and T. Pruzinsky (Eds.), *Body Image: A Handbook of Theory, Research, and Clinical Practice.* New York: The Guilford Press, 2002. Pp. 440–449.
8. Reaby, L. L., Hort, L. K., and Vandervord, J. Body image, self-concept, and self-esteem in women who had a mastectomy and either wore an external breast prosthesis or had breast reconstruction and women who had not experienced mastectomy. *Health Care Women Int.* 15: 361, 1994.
9. Reaby, L. L., and Hort, L. K. Postmastectomy attitudes in women who wear external breast prostheses compared to those who have undergone breast reconstructions. *J. Behav. Med.* 18: 55, 1995.
10. Reaby, L. L. Breast restoration decision making. *Plast. Surg. Nurs.* 19: 22, 1999.
11. Reaby, L. L. Breast restoration decision making: Enhancing the process. *Cancer Nurs.* 21: 196, 1998.
12. Reaby, L. L. Reasons why women who have mastectomy decide to have or not to have breast reconstruction. *Plast. Reconstr. Surg.* 101: 1810, 1998.
13. Reaby, L. L. The quality and coping patterns of women's decision-making regarding breast cancer surgery. *Psychooncology* 7: 252, 1998.
14. Schain, W. S., Jacobs, E., and Wellisch, D. K. Psychosocial issues in breast reconstruction: Intrapsychic, interpersonal, and practical concerns. *Clin. Plast. Surg.* 11: 237, 1984.
15. Keith, D. J. W., Walker, M. B., Walker, L. G., et al. Women who wish breast reconstruction: Characteristics, fears, and hopes. *Plast. Reconstr. Surg.* 111: 1051, 2003.
16. White, C. A. Body image dimensions and cancer: A heuristic cognitive behavioural model. *Psychooncology* 9: 183, 2000.
17. Tseng, J. F., Kronowitz, S. J., Sun, C. C., et al. The effect of ethnicity on immediate reconstruction rates after mastectomy for breast cancer. *Cancer* 101: 1514, 2004.
18. Fobair, P., O'Hanlan, K., Koopman, C., et al. Comparison of lesbian and heterosexual women's response to newly diagnosed breast cancer. *Psychooncology* 10: 40, 2001.
19. Beren, S. E., Hayden, H. A., Wilfley, D. E., et al. The influence of sexual orientation on body dissatisfaction in adult men and women. *Int. J. Eat. Disord.* 20: 135, 1996.
20. Rothblum, E. Gay and lesbian body images. In T. F. Cash and T. Pruzinsky (Eds.), *Body Image: A Handbook of Theory, Research, and Clinical Practice.* New York: The Guilford Press, 2002. Pp. 257–265.
21. Conner, M., Johnson, C., and Grogan, S. Gender, sexuality, body image and eating behaviours. *J. Health Psychol.* 9: 505, 2004.
22. Sarwer, D. B., Pertschuk, M. J., Wadden, T. A., et al. Psychological investigations in cosmetic surgery: A look back and a look ahead. *Plast. Reconstr. Surg.* 101: 1136, 1998.

23. Watters, J. K., and Biernacki, P. Targeted sampling: Options for the study of hidden populations. *Soc. Probl.* 36: 416, 1989.
24. Biernacki, P., and Waldorf, D. Snowball sampling: Problems and techniques of chain referral sampling. *Sociol. Methods Res.* 10: 141, 1981.
25. Boehmer, U., Freund, K. M., and Linde, R. Support providers of sexual minority women with breast cancer: Who they are and how they impact the breast cancer experience. *J. Psychosom. Res.* 59: 307, 2005.
26. Solarz, A. L. Lesbian Health: Current Assessment and Directions for the Future. Washington, D.C.: Institute of Medicine, Committee on Lesbian Health Research Priorities, Neuroscience and Behavioral Health Program, Health Sciences Policy Program, Health Sciences Section, 1999.
27. Strauss, A., and Corbin, J. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* Newbury Park, Calif.: Sage, 1990.
28. Hopwood, P., Fletcher, I., Lee, A., et al. A body image scale for use with cancer patients. *Eur. J. Cancer* 37: 189, 2001.
29. Figueiredo, M. I., Cullen, J., Hwang, Y.-T., et al. Breast cancer treatment in older women: Does getting what you want improve your long-term body image and mental health? *J. Clin. Oncol.* 22: 4002, 2004.
30. Carver, C. S., Pozo-Kaderman, C., Price, A. A., et al. Concern about aspects of body image and adjustment to early stage breast cancer. *Psychosom. Med.* 60: 168, 1998.
31. Harcourt, D. M., Rumsey, N. J., Ambler, N. R., et al. The psychological effect of mastectomy with or without breast reconstruction: A prospective, multicenter study. *Plast. Reconstr. Surg.* 111: 1060, 2003.
32. Boehmer, U., Linde, R., and Freund, K. M. Sexual minority women's coping and psychological adjustment after a diagnosis of breast cancer. *J. Womens Health* 14: 214, 2005.
33. Liang, W., Burnett, C. B., Rowland, J. H., et al. Communication between physicians and older women with localized breast cancer: Implications for treatment and patient satisfaction. *J. Clin. Oncol.* 20: 1008, 2002.
34. Mandelblatt, J., Figueiredo, M., and Cullen, J. Outcomes and quality of life following breast cancer treatment in older women: When, why, how much, and what do women want? *Health Qual. Life Outcomes* 1: 45, 2003.
35. Ong, L. M., de Haes, J. C., Hoos, A. M., et al. Doctor-patient communication: A review of the literature. *Soc. Sci. Med.* 40: 903, 1995.
36. Boehmer, U., and Case, P. Physicians don't ask, some patients tell: Disclosure of sexual orientation among women with breast cancer. *Cancer* 101: 1882, 2004.

The American Journal of Medicine (2007) 120, 272-279

ELSEVIER

THE AMERICAN JOURNAL of MEDICINE®

CLINICAL RESEARCH STUDY

AJM Theme Issue: Gastroenterology

Cost-effectiveness of Treatment for Hepatitis C in an Urban Cohort Co-infected with HIV

Nicole G. Campos, MSc,[a] Joshua A. Salomon, PhD,[b] Julie C. Servoss, MD, MPH,[c] David P. Nunes, MD,[d] Jeffrey H. Samet, MD, MA, MPH,[e] Kenneth A. Freedberg, MD, MSc,[f,g] Sue J. Goldie, MD, MPH[g]

[a]Program in Health Policy, Harvard University, Cambridge, Mass; [b]Department of Population and International Health, Harvard School of Public Health, Harvard University Initiative for Global Health, Cambridge, Mass; [c]Gastrointestinal Unit, Massachusetts General Hospital, Harvard Medical School, Boston; [d]Section of Gastroenterology, and [e]Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, Mass; [f]Divisions of General Medicine and Infectious Diseases and the Harvard Center for AIDS Research (CFAR), Massachusetts General Hospital, Harvard Medical School, Boston; and [g]Program in Health Decision Science, Department of Health Policy and Management, Harvard School of Public Health, Boston, Mass.

ABSTRACT

PURPOSE: Recent clinical trials have evaluated treatment strategies for chronic infection with hepatitis C virus (HCV) in patients co-infected with human immunodeficiency virus (HIV). Our objective was to use these data to examine the cost-effectiveness of treating HCV in an urban cohort of co-infected patients.

METHODS: A computer-based model, together with available published data, was used to estimate lifetime costs (2004 US dollars), life expectancy, and incremental cost per year of life saved (YLS) associated with 3 treatment strategies: (1) interferon-alfa and ribavirin; (2) pegylated interferon-alfa; and (3) pegylated interferon-alfa and ribavirin. The target population included treatment-eligible patients, based on an actual urban cohort of HIV-HCV co-infected subjects, with a mean age of 44 years, of whom 66% had genotype 1 HCV, 16% had cirrhosis, and 98% had CD4 cell counts >200 cells/mm^3.

RESULTS: Pegylated interferon-alfa and ribavirin was consistently more effective and cost-effective than other treatment strategies, particularly in patients with non-genotype 1 HCV. For patients with CD4 counts between 200 and 500 cells/mm^3, survival benefits ranged from 5 to 11 months, and incremental cost-effectiveness ratios were consistently less than $75,000 per YLS for men and women of both genotypes. Due to better treatment efficacy in non-genotype 1 HCV patients, this group experienced greater life expectancy gains and lower incremental cost-effectiveness ratios.

CONCLUSIONS: Combination therapy with pegylated interferon-alfa and ribavirin for HCV in eligible co-infected patients with stable HIV disease provides substantial life-expectancy benefits and appears to be cost-effective. Overcoming barriers to HCV treatment eligibility among urban co-infected patients remains a critical priority.

KEYWORDS: Hepatitis C virus (HCV); Human immunodeficiency virus (HIV); Cost-effectiveness; Peginterferon-alfa and ribavirin; Clinical guidelines; Treatment eligibility

Among the estimated 950,000 persons infected with human immunodeficiency virus (HIV) in the United States, approximately 30% are co-infected with the hepatitis C virus (HCV).[1,2] While highly active antiretroviral therapy (HAART) has essentially transformed HIV to a chronic disease, co-infected patients are increasingly vulnerable to complications of chronic liver disease, including cirrhosis and liver failure. Compared with HCV mono-infected patients, they tend to have higher levels of HCV RNA and

Supported by the National Institute on Alcohol Abuse and Alcoholism R01-AA13216 and the National Institute of Allergy and Infectious Disease K24 AI062476.

Requests for reprints should be addressed to Nicole G. Campos, MSc, Department of Health Policy and Management, Harvard School of Public Health, 718 Huntington Avenue, 2nd Floor, Boston, MA 02115.

E-mail address: gastin@post.harvard.edu

0002-9343/$ -see front matter
doi:10.1016/j.amjmed.2006.06.036

to progress more rapidly to cirrhosis and end-stage liver disease.[3] Mortality attributable to end-stage liver disease has steadily increased since 1996, and in some HIV patient populations it is now the leading cause of death.[4] The impact of HCV on HIV progression is more controversial.[5-7]

In clinical trials among patients with HCV mono-infection, combination therapy with pegylated interferon-alfa and ribavirin has produced sustained virologic response rates ranging from 54% to 63%.[8-10] Recently, 4 randomized controlled trials evaluated combination therapy with pegylated interferon-alfa and ribavirin compared with interferon-alfa and ribavirin in patients with HIV-HCV co-infection.[11-14] The largest of these trials, the AIDS Pegasys Ribavirin International Co-infection Trial (APRICOT), was conducted at 95 centers in 19 countries with 868 subjects and yielded a sustained virologic response rate of 40%.[11] Based on the APRICOT findings, in February 2005 the United States (US) Food and Drug Administration approved pegylated interferon-alfa-2a and ribavirin for the treatment of HCV in patients with HIV.[15]

CLINICAL SIGNIFICANCE

- In patients infected with both HIV and HCV, therapy with pegylated interferon-alfa and ribavirin for HCV increases life-expectancy and appears to be cost-effective.

To the best of our knowledge, only one cost-effectiveness analysis evaluating the treatment of HCV in HIV-HCV co-infected patients has been published. Kuehne et al demonstrated that combination therapy for histologically moderate HCV in co-infected patients resulted in an increase in quality-adjusted life expectancy while incurring costs comparable with other well-accepted clinical interventions.[16] However, this analysis was performed before randomized controlled trials had established approximate treatment efficacy rates in HIV-HCV co-infected patients, and the APRICOT trial rates of sustained virologic response were generally lower than the lower bounds of the sensitivity analysis performed by Kuehne and colleagues. Since this prior cost-effectiveness analysis was conducted, considerable progress has been made in discerning treatment efficacy rates and relative risk estimates for progression of liver disease in HIV-HCV co-infected patients.

Our objective was to use recent prospective data regarding eligibility for interferon-based treatment, the impact of HIV on the progression of HCV-related liver disease, and demonstrated treatment efficacy from clinical trials to consider the potential health benefits, economic costs, and cost-effectiveness of treatment for HCV among an urban cohort of co-infected patients with stable HIV disease.

METHODS

Overview

We modified an existing Markov model of HCV[17] to reflect co-infection with HIV and examined the cost-effectiveness of the following strategies for HCV treatment in the treatment-eligible segment of an urban co-infected cohort[18]: combination therapy with interferon-alfa-2a and ribavirin; monotherapy with pegylated interferon-alfa-2a; and combination therapy with pegylated interferon-alfa-2a and ribavirin.

Population characteristics (mean age, Metavir score distribution, and mean CD4 cell count) for the modeled cohort were derived from a subgroup of the Hepatitis and AIDS Liver Outcomes (HALO) Study cohort that was co-infected with HIV and HCV, and eligible for treatment (Table 1).[18]

We followed the recommendations of the US Panel on Cost-Effectiveness in Health and Medicine,[19] adopting a societal perspective (although we excluded patient time costs) and discounting all costs and clinical consequences at a rate of 3% per year. The comparative efficiencies of alternative treatment strategies were measured by the incremental cost-effectiveness ratio, defined as the additional cost of a specific treatment strategy divided by its additional health benefit, expressed here as

Table 1 Demographic Characteristics of the Treatment-eligible HIV-HCV Co-infected Subgroup of the HALO Cohort[18]

Variable	Treatment-eligible* n = 44 (%)
Age (years ± SD)	44.8 ± 7.3
Sex	
Male	34 (77)
Female	10 (23)
CD4 cell count >200 cells/mm^3	43 (98)
HCV genotype	
1	29 (66)
2, 3, or 4	15 (34)
Liver biopsy results (Metavir score)†	
F1	8 (25)
F2	14 (44)
F3	5 (16)
F4	5 (16)

*Exclusion criteria for treatment eligibility in this cohort included the following: nonadherence (missing >3 clinic appointments), ongoing alcohol or drug use (other than marijuana) in the preceding 6 months, active psychiatric illness (defined as symptomatic psychosis or depression or a suicide attempt within the previous year), active medical illness (defined as ongoing illness that is a contraindication to interferon therapy or is associated with a life expectancy of <3 years), decompensated liver disease (defined as a Child Pugh score of >7), advanced HIV disease (defined as a CD4 cell count of <100 cells/mm^3 regardless of HIV viral load, or a count of 100-200 cells/mm^3 with a viral load of >10,000 copies per milliliter), neutrophil count of $<1.5 \times 10^9$ cells per liter, and platelet count of $<75 \times 10^9$ cells per liter. Analyses are stratified on HCV genotype (ie, 1 versus non-1), which affects response to therapy, and sex, which affects rate of fibrosis progression.

†Metavir scores were available for 32 subjects. Percentages add to more than 100 due to rounding.

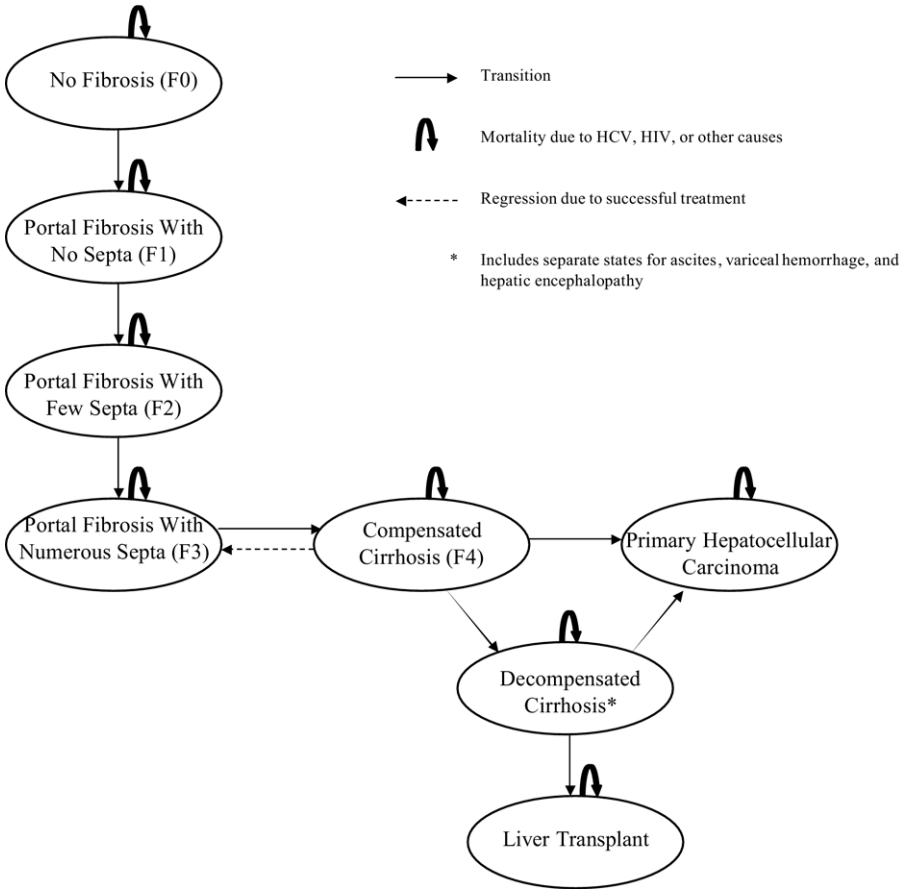


Figure 1 Overview of the Markov model.

years of life saved (YLS). The incremental ratio for a strategy was computed in comparison with the next most effective option after eliminating strategies that were dominated (more costly and less effective than other options) or ruled out by extended (weak) dominance (strategies with higher incremental cost-effectiveness ratios than more effective options). We conducted sensitivity analyses to assess the influence of varying uncertain parameters and adopting alternative assumptions on our results.

Model

A deterministic state-transition Markov model (DATA 4.0; TreeAge Software Inc., Williamstown, Mass) was used to simulate the natural history of HCV infection in patients co-infected with HIV. Early stages of liver disease were classified using the Metavir scoring system, which characterizes the extent of fibrosis that results as damaged liver cells are repaired, including no fibrosis (F0), portal fibrosis without septa (F1), portal fibrosis with few septa (F2), numerous septa without cirrhosis (F3), and cirrhosis (F4). Advanced stages of liver disease were defined clinically as compensated cirrhosis, decompensated cirrhosis (including separate states for ascites, variceal hemorrhage and hepatic encephalopathy), and primary hepatocellular carcinoma (Figure 1). Monthly transition probabilities were derived from the literature and allowed individuals to move through different health states over time.

We made the following assumptions:

- HCV infection may resolve through successful treatment, implying clearance of HCV RNA[20];
- Patients without a sustained viral response to HCV treatment (as defined by an HCV RNA level of <50 IU per milliliter 24 weeks after completion of therapy) received no clinical benefit and were subject to pretreatment rates of HCV-related liver disease progression[11,21];
- Patients in early stages of liver disease who experienced a sustained viral response were no longer at risk for HCV-related liver disease[22];
- Patients with cirrhosis received treatment, and if it was effective they regressed to Metavir stage F3; if treatment was ineffective they were still at risk for decompensated cirrhosis[23];
- Response to treatment was conditional on genotype[11];
- The rate of fibrosis progression in the absence of effective treatment was conditional on age and sex, and remained the same for transition to the next higher Metavir stage;
- All patients were assumed to have a stable CD4 cell count between 200 and 500 cells/mm^3 and to be receiving HAART for HIV; and

- Patients with decompensated cirrhosis were eligible for liver transplantation.[24,25]

Data

Selected parameters used in the model are shown in Table 2. We used baseline age- and sex-specific rates of progression from chronic HCV infection to cirrhosis based on an empirically calibrated model of chronic HCV in mono-infected patients.[17,26] We modified these based on data from studies comparing the relative progression in HIV co-infected patients versus mono-infected patients.[27-30] We assumed that progression from cirrhosis to decompensated cirrhosis was similar in co-infected patients and mono-infected patients, and used rates derived from a cost-effectiveness study in mono-infected patients.[31] Excess mortality due to HIV was based on data from the Multicenter AIDS Cohort Study, from which CD4-specific rates were derived to parameterize a natural history model of HIV/AIDS published by Freedberg et al[32]; an additive relationship to age- and sex-standardized mortality rates was assumed. We then compared the model's predictions of cirrhosis prevalence with a published study by Di Martino et al that was not used for natural history parameter estimation.[33] Among an HIV-HCV co-infected cohort of former intravenous drug users with an average 10.6-year duration of HCV infection and a mean CD4 count of 482 cells/mm^3, 59% of whom had received interferon monotherapy (overall efficacy of 6.4%) and 73% of whom were male, 8.75% had cirrhosis at baseline and 17.5% had cirrhosis at follow-up 4.7 years later.[33] Our model predicted that in a cohort of patients with this sex and treatment profile and duration of HCV infection, 16% had cirrhosis over the same follow-up period.

Annual costs of care related to chronic HCV infection and liver disease included detailed estimates of resource utilization, including hospitalizations, outpatient visits, laboratory tests, medications, and interventions.[31] Treatment costs were based on average wholesale drug prices[34] combined with previously published cost estimates for clinic visits, laboratory tests, and the treatment of adverse events.[35] The annual costs of HIV care for patients on HAART with CD4 count between 200 and 500 cells/mm^3 were obtained from a recently published model for HIV screening.[36,37]

In the base case we assumed treatment for 48 weeks,[11-13] and dosages resembled those in APRICOT: for interferon and ribavirin, 3 million IU interferon alfa-2a subcutaneously 3 times/week plus 800 mg ribavirin/day; for pegylated interferon, 180 μg of pegylated interferon alfa-2a subcutaneously weekly; and for pegylated interferon and ribavirin, pegylated interferon as described above plus 800 mg ribavirin/day. We made the conservative assumption that all patients completed the full course of medication. In sensitivity analyses, we explored a second treatment protocol and assumed that patients without an early virologic response at week 12 discontinued treatment. The percentage of such patients was drawn from APRICOT and varied by treatment arm. We assessed early treatment withdrawal due to adverse events or abnormal laboratory values in a sensitivity analysis.

RESULTS

Table 3 shows the discounted lifetime costs, life expectancy, and incremental cost per YLS for each treatment strategy, stratified by sex and genotype. For the men in our modeled cohort, the average discounted life expectancy without treatment was 11.6 years and lifetime costs were $240,300. In men with genotype 1, treatment for HCV provided incremental gains ranging from 1.0 to 5.2 months compared with no therapy. Combination therapy with pegylated interferon and ribavirin dominated all other strategies because it was both more effective and had a lower (more attractive) cost-effectiveness ratio. Compared with no therapy, its incremental cost-effectiveness ratio was $73,000 per YLS.

For men with non-genotype 1 HCV, treatment provided incremental gains ranging from 3.0 to 10.7 months compared with no therapy. Again, combination therapy with pegylated interferon and ribavirin was the dominant strategy. Compared with no therapy, its incremental cost-effectiveness ratio was $39,700 per YLS. Results in women were very similar.

In an exploratory analysis assuming cessation of treatment when no early virologic response occurs, the rank ordering of strategies remained the same. The incremental cost-effectiveness ratio for combination therapy with pegylated interferon and ribavirin became more attractive—lower by 19% in men with genotype 1 ($59,300 per YLS) and by 17% in men with non-genotype 1 ($33,100 per YLS). Again, results in women were similar.

Results were relatively insensitive to varying parameters across plausible ranges for treatment efficacies among patients with cirrhosis and treatment risks. Results were most sensitive to variation in the annual excess death rate due to HIV, fibrosis progression rates and treatment efficacies in noncirrhotic patients. Results were moderately sensitive to drug costs. For men with genotype 1 infection, when the excess mortality due to HIV was reduced by 97%, the incremental cost-effectiveness ratio decreased to $41,400 (base case: $73,000 per YLS). When excess mortality was increased 11-fold to represent death rates in patients with a history of severe opportunistic infections, the incremental cost-effectiveness ratio increased so greatly that treatment was dominated by nontreatment.

We conducted a 2-way sensitivity analysis in which we varied the effectiveness of combination therapy with pegylated interferon and ribavirin, and the relative risk of fibrosis progression due to co-infection with HIV (Figure 2). When treatment efficacy exceeded 50%, cost-effectiveness ratios were consistently less than $50,000 per YLS, regardless of the relative risk of fibrosis progression. When efficacy was >25%, ratios were consistently <$100,000 per YLS across the entire plausible range of relative risk assumptions. However, when efficacy was <25%, the relative risk of progres-

Table 2 Base Case Values for Model Parameters

Variable	Base Case
HCV natural history parameters*[17,26-29,31]	
Fibrosis progression in men, age (years)†	
0	0.149
50	0.169
60	0.298
≥70	0.406
Fibrosis progression in women, age (years)†	
0	0.108
50	0.135
60	0.208
≥70	0.284
Cirrhosis to ascites	0.025
Cirrhosis to hepatic encephalopathy	0.004
Cirrhosis to hepatocellular carcinoma	0.015
Cirrhosis to variceal hemorrhage	0.011
Liver transplant probability	0.031
Mortality rate per person	
Ascites	0.110
Hepatic encephalopathy (first year)	0.680
Hepatic encephalopathy (subsequent years)	0.400
Hepatocellular carcinoma	0.433
Transplant (first year)	0.210
Transplant (subsequent years)	0.057
Variceal hemorrhage (first year)	0.400
Variceal hemorrhage (subsequent years)	0.130
Treatment parameters[11]	
Treatment response probability	
Interferon-alfa + ribavirin	
Genotype 1	0.07
Non-genotype 1	0.18
Pegylated interferon-alfa	
Genotype 1	0.14
Non-genotype 1	0.31
Pegylated interferon-alfa + ribavirin	
Genotype 1	0.29
Non-genotype 1	0.58
Probability of early virologic response (EVR)	
Interferon-alfa + ribavirin	0.38
Pegylated interferon-alfa	0.55
Pegylated interferon-alfa + ribavirin	0.71
Treatment mortality probability	0.002
Costs (2004 US $)[31,34-37]	
Treatment Protocol 1‡	
Interferon-alfa + ribavirin	$15,568
Pegylated interferon-alfa	$19,305
Pegylated interferon-alfa + ribavirin	$27,880
Treatment Protocol 2§	
Interferon-alfa + ribavirin	$8,769
Pegylated interferon-alfa	$13,109
Pegylated interferon-alfa + ribavirin	$22,022
Costs of annual HCV care	
Chronic HCV	$140
Compensated cirrhosis	$1,017
Ascites	$4,280
Variceal hemorrhage, first year	$23,669
Variceal hemorrhage, subsequent years	$4,632
Hepatic encephalopathy, first year	$15,192
Hepatic encephalopathy, subsequent years	$3,519
Hepatocellular carcinoma	$40,828
Liver transplant, first year	$134,458
Liver transplant, subsequent years	$23,481
Costs of annual HIV care	
CD4 count 200-500/mm^3	$5,096
Three-drug antiretroviral therapy	$13,752

*Annual rates per person are presented here except where specified, but these were converted to monthly probabilities in the model.

†Fibrosis progression rates are assumed to be linearly interpolated in these ranges. Progression rates were the same for Metavir stages F0 to F4.

‡All patients receive the full 48-week course of therapy.

§Patients without an early virologic response receive only 12 weeks of therapy; protocol assumes nondrug costs by week 12 are two-thirds of nondrug costs for 48 weeks of therapy.

Table 3 Cost-effectiveness of 4 Strategies by Treatment Protocol, Sex, and Genotype for Patients with CD4 Count 200-500 Cells/mm^3

Strategy*	Cost, 2004 US$	Life Expectancy, Years	Incremental Cost per YLS
Men			
Genotype 1			
No treatment	$240,300	11.63	–
Interferon + ribavirin	$256,400	11.71	†
Pegylated interferon	$261,100	11.83	†
Pegylated interferon + ribavirin	$271,700	12.06	$73,000
Non-genotype 1			
No treatment	$240,300	11.63	–
Interferon + ribavirin	$257,900	11.88	†
Pegylated interferon	$263,400	12.09	†
Pegylated interferon + ribavirin	$275,600	12.52	$39,700
Women			
Genotype 1			
No treatment	$252,200	12.28	–
Interferon + ribavirin	$268,400	12.37	†
Pegylated interferon	$273,200	12.48	†
Pegylated interferon + ribavirin	$284,000	12.73	$70,700
Non-genotype 1			
No treatment	$252,200	12.28	–
Interferon + ribavirin	$270,000	12.55	†
Pegylated interferon	$275,700	12.76	†
Pegylated interferon + ribavirin	$288,400	13.20	$39,300

*Assumes 48 weeks of HCV therapy for all patients.
†Dominated strategy.

sion had slightly more influence on the cost-effectiveness of treatment.

Results were sensitive to the discount rate used. With no discounting, the incremental cost-effectiveness ratio was approximately 60% lower than in the base case, while a discount rate of 5% resulted in a ratio that was 140% higher than the base case. Varying the cost of pegylated interferon and ribavirin between 50% and 150% of the base-case resulted in cost-effectiveness ratios ranging from $56,300 to $88,500 per YLS.

DISCUSSION

We found that treatment with pegylated interferon and ribavirin for chronic HCV infection in an urban cohort co-infected with HIV, with characteristics similar to the treatment-eligible subgroup in the HALO Study, will provide substantial life expectancy gains. These gains are greatest in patients with non-genotype 1 infection. Combination therapy with pegylated interferon and ribavirin was the most effective and cost-effective treatment strategy regardless of genotype or sex.

There were substantial differences in the cost-effectiveness ratios between patients with genotype 1 and non-genotype 1 HCV, mainly attributable to differences in treatment response rates. There were only small differences in the cost-effectiveness of treatment for men and women.

Four trials to date have evaluated the efficacy of HCV treatment in HIV-HCV co-infected patients; we used efficacy results from the largest multicenter trial, APRICOT, for the base case analysis.[11-14] Each trial found treatment to be more efficacious in non-genotype 1 patients, but the magnitude of the differences varied by trial, as shown in Figure 2. Because cost-effectiveness ratios are sensitive to treatment efficacy, a wide range of cost-effectiveness ratios is possible. Despite our conservative assumptions, our cost-effectiveness results may be particularly favorable because APRICOT efficacies are high relative to other trials' results. However, the APRICOT study's treatment protocol requiring 800 mg of ribavirin daily has become a relatively standard clinical practice. The AIDS Clinical Trials Group (ACTG) trial, on the other hand, administered ribavirin according to a dose-escalation schedule.[13] The Agence Nationale de Recherches sur le Sida (ANRS) HCO2-RIBAVIC trial's exclusion criteria for subjects was less stringent than the other trials, allowing patients with alcohol intake up to 40 grams per day for women or 50 grams per day for men to participate; 21% of this study population had psychiatric disorders, and 40% had bridging fibrosis or cirrhosis.[12] These factors may explain the higher efficacies demonstrated in the APRICOT study, but may also indicate that the RIBAVIC trial results might be more applicable to the urban cohort modeled here if treatment eligibility criteria are relaxed.

In contrast to results from a recent cost-effectiveness analysis in HCV mono-infected patients, the present study suggests that monotherapy with pegylated interferon and combination therapy with interferon and ribavirin are dominated strategies in co-infected populations.[17] Our results also differ from those of Kuehne et al,[16] because this earlier cost-effectiveness analysis of HIV-HCV co-infected pa-

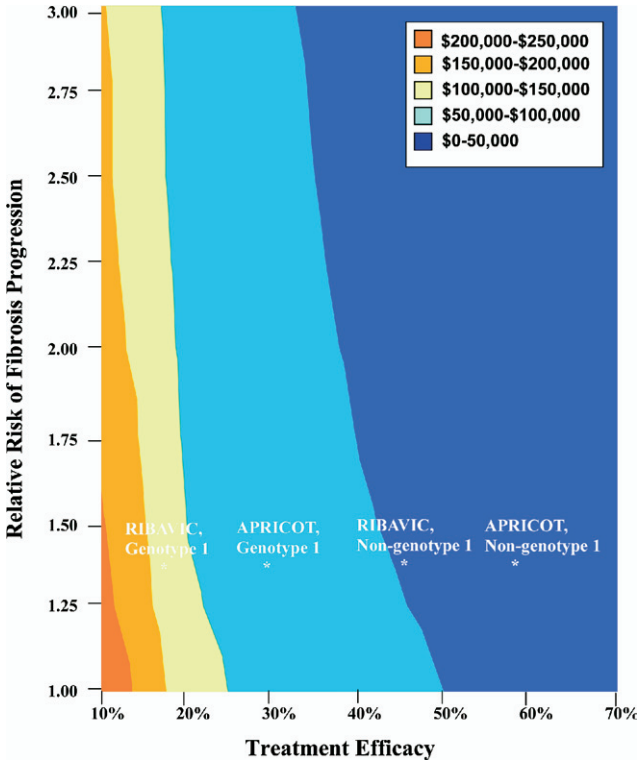


Figure 2 Two-way sensitivity analysis for treatment efficacy and the relative risk of fibrosis progression due to co-infection with HIV. * Indicates percentage of patients with a sustained viral response. The base case value for relative risk of fibrosis progression was 1.35. Efficacy points are merely for reference and do not necessarily represent cost-effectiveness ratios for the particular trial or study they mark due to different drug regimens, protocols, and subject characteristics.

tients assumed treatment efficacy and protocols were the same as in mono-infected patients. The APRICOT study not only demonstrated that treatment efficacy is substantially lower in co-infected patients, but that combination therapy with interferon and ribavirin rarely results in a sustained viral response, even in non-genotype 1 patients. While Kuehne et al found this strategy to be cost-effective ($18,500 per YLS relative to no treatment in genotype 1 patients and $63,500 per YLS relative to 24 weeks of combination therapy with interferon and ribavirin in non-genotype 1 patients), our analysis suggests that combination therapy with pegylated interferon and ribavirin dominates other treatment strategies.

Due to limitations in data that would ideally inform a model of co-infection, our goals for this analysis were modest and exploratory in nature, aiming to update a previous analysis with treatment efficacy data and identify influential parameters. Implications of our analysis are restricted to a specific target population of treatment-eligible patients with stable HIV disease and stable CD4 cell counts between 200 and 500 cells/mm^3. A more sophisticated model of co-infection that fully represents the natural history of HIV disease will be necessary to explore important questions regarding the optimal timing of treatment for chronic HCV relative to antiretroviral therapy for HIV disease, and potential interactions or additive toxicities between treatments. Other limitations of our study include the uncertainty surrounding many of the model's parameters. Also, we did not consider health-related quality of life in co-infected patients, and all of our costs were literature-based.

We chose to use population characteristics and treatment eligibility criteria from an urban cohort that has been previously described.[18] While the treatment-eligible portion of the HALO Study cohort was in many respects similar to the population studied by APRICOT, efficacy as demonstrated in a randomized controlled trial is not the same as effectiveness in a typical clinical setting. Prospective studies currently underway with the HALO Study and other cohorts will provide insight to treatment effectiveness in a particular population of HIV-HCV patients. It is important to note that only 30% of the co-infected subgroup of the HALO Study cohort was eligible for HCV treatment; the remainder were not eligible for a variety of reasons, including unstable social circumstances, concern about potential adverse effects, and concern about ability to work during the treatment course.[38]

As treatment-eligible co-infected patients are not currently the norm, further studies are needed to establish the effectiveness of combination HCV therapy in populations with low eligibility for treatment. Overcoming barriers to HCV treatment eligibility and initiation in HIV-HCV co-infected patients remains a priority, now that combination therapy has been demonstrated efficacious in certain populations. For co-infected patients with stable HIV disease, treatment appears to be not only life-prolonging but cost-effective as well.

References

1. U.S. Department of Health and Human Services. *HIV/AIDS Statistics, NIAID Facts and Figures*; July 2004.
2. Sulkowski MS, Mast EE, Seeff LB, Thomas DL. Hepatitis C virus infection as an opportunistic disease in persons infected with human immunodeficiency virus. *Clin Infect Dis*. 2000;30(Suppl 1):S77-S84.
3. Graham CS, Baden LR, Yu E, et al. Influence of human immunodeficiency virus infection on the course of hepatitis C virus infection: a meta-analysis. *Clin Infect Dis*. 2001;33:562-569.
4. Bica I, McGovern B, Dhar R, et al. Increasing mortality due to end-stage liver disease in patients with human immunodeficiency virus infection. *Clin Infect Dis*. 2001;32:492-497.
5. Greub G, Ledergerber B, Battegay M, et al. Clinical progression, survival, and immune recovery during antiretroviral therapy in patients with HIV-1 and hepatitis C virus coinfection: the Swiss HIV Cohort Study. *Lancet*. 2000;356:1800-1805.
6. Anderson KB, Guest JL, Rimland D. Hepatitis C virus coinfection increases mortality in HIV-infected patients in the highly active antiretroviral therapy era: data from the HIV Atlanta VA Cohort Study. *Clin Infect Dis*. 2004;39:1507-1513.
7. Sulkowski MS, Moore RD, Mehta SH, Chaisson RE, Thomas DL. Hepatitis C and progression of HIV disease. *JAMA*. 2002;288:199-206.
8. Hadziyannis SJ, Sette H Jr, Morgan TR, et al. Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med*. 2004; 140:346-355.

9. Fried MW, Shiffman ML, Reddy KR, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med*. 2002; 347:975-982.
10. Manns MP, McHutchison JG, Gordon SC, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet*. 2001;358: 958-965.
11. Torriani FJ, Rodriguez-Torres M, Rockstroh JK, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection in HIV-infected patients. *N Engl J Med*. 2004;351:438-450.
12. Carrat F, Bani-Sadr F, Pol S, et al. Pegylated interferon alfa-2b vs standard interferon alfa-2b, plus ribavirin, for chronic hepatitis C in HIV-infected patients: a randomized controlled trial. *JAMA*. 2004;292: 2839-2848.
13. Chung RT, Andersen J, Volberding P, et al. Peginterferon alfa-2a plus ribavirin versus interferon alfa-2a plus ribavirin for chronic hepatitis C in HIV-coinfected persons. *N Engl J Med*. 2004;351:451-459.
14. Laguno M, Murillas J, Blanco JL, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for treatment of HIV/HCV co-infected patients. *AIDS*. 2004;18:F27-F36.
15. U.S. Food and Drug Administration Center for Drug Evaluation and Research. CDER Fast Track Products Approved Since 1998 Through September 30, 2005. Available at: http://www.fda.gov/cder/rdmt/internetftap.htm. Accessed January 19, 2006.
16. Kuehne FC, Bethe U, Freedberg K, Goldie SJ. Treatment for hepatitis C virus in human immunodeficiency virus-infected patients: clinical benefits and cost-effectiveness. *Arch Intern Med*. 2002;162:2545-2556.
17. Salomon JA, Weinstein MC, Hammitt JK, Goldie SJ. Cost-effectiveness of treatment for chronic hepatitis C infection in an evolving patient population. *JAMA*. 2003;290:228-237.
18. Fleming CA, Craven DE, Thornton D, Tumilty S, Nunes D. Hepatitis C virus and human immunodeficiency virus coinfection in an urban population: low eligibility for interferon treatment. *Clin Infect Dis*. 2003;36:97-100.
19. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in Health and Medicine*. New York, NY: Oxford University Press; 1996.
20. Lau DT, Kleiner DE, Ghany MG, Park Y, Schmid P, Hoofnagle JH. 10-Year follow-up after interferon-alpha therapy for chronic hepatitis C. *Hepatology*. 1998;28:1121-1127.
21. Feld JJ, Hoofnagle JH. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature*. 2005;436:967-972.
22. Marcellin P, Boyer N, Gervais A, et al. Long-term histologic improvement and loss of detectable intrahepatic HCV RNA in patients with chronic hepatitis C and sustained response to interferon-alpha therapy. *Ann Intern Med*. 1997;127:875-881.
23. Arima M, Terao H, Kashima K, Arita T, Nasu M, Nishizono A. Regression of liver fibrosis in cases of chronic liver disease type C: quantitative evaluation by using computed image analysis. *Intern Med*. 2004;43:902-910.
24. Maida I, Nunez M, Gonzalez-Lahoz J, Soriano V. Liver transplantation in HIV-HCV coinfected candidates: what is the most appropriate time for evaluation? *AIDS Res Hum Retroviruses*. 2005;21:599-601.
25. Moreno S, Fortun J, Quereda C, et al. Liver transplantation in HIV-infected recipients. *Liver Transpl*. 2005;11:76-81.
26. Salomon JA, Weinstein MC, Hammitt JK, Goldie SJ. Empirically calibrated model of hepatitis C virus infection in the United States. *Am J Epidemiol*. 2002;156:761-773.
27. Martinez-Sierra C, Arizcorreta A, Diaz F, et al. Progression of chronic hepatitis C to liver fibrosis and cirrhosis in patients coinfected with hepatitis C virus and human immunodeficiency virus. *Clin Infect Dis*. 2003;36:491-498.
28. Mohsen AH, Easterbrook PJ, Taylor C, et al. Impact of human immunodeficiency virus (HIV) infection on the progression of liver fibrosis in hepatitis C virus infected patients. *Gut*. 2003;52:1035-1040.
29. Benhamou Y, Bochet M, Di Martino V, et al. Liver fibrosis progression in human immunodeficiency virus and hepatitis C virus coinfected patients. *Hepatology*. 1999;30:1054-1058.
30. Marine-Barjoan E, Saint-Paul MC, Pradier C, et al. Impact of antiretroviral treatment on progression of hepatic fibrosis in HIV/hepatitis C virus co-infected patients. *AIDS*. 2004;18:2163-2170.
31. Bennett WG, Inoue Y, Beck JR, Wong JB, Pauker SG, Davis GL. Estimates of the cost-effectiveness of a single course of interferon-alpha 2b in patients with histologically mild chronic hepatitis C. *Ann Intern Med*. 1997;127:855-865.
32. Freedberg KA, Scharfstein JA, Seage GR 3rd, et al. The cost-effectiveness of preventing AIDS-related opportunistic infections. *JAMA*. 1998;279:130-136.
33. Di Martino V, Rufat P, Boyer N, et al. The influence of human immunodeficiency virus coinfection on chronic hepatitis C in injection drug users: a long-term retrospective cohort study. *Hepatology*. 2001; 34:1193-1199.
34. *Red Book 2004: Pharmacy's Fundamental Reference*. Stamford, CT: Thomson Healthcare; 2004.
35. Wong JB, Poynard T, Ling MH, Albrecht JK, Pauker SG. Cost-effectiveness of 24 or 48 weeks of interferon alpha-2b alone or with ribavirin as initial treatment of chronic hepatitis C. International Hepatitis Interventional Therapy Group. *Am J Gastroenterol*. 2000;95: 1524-1530.
36. 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;352:570-585.
37. Bozzette SA, Joyce G, McCaffrey DF, et al. Expenditures for the care of HIV-infected patients in the era of highly active antiretroviral therapy. *N Engl J Med*. 2001;344:817-823.
38. Fleming CA, Tumilty S, Murray JE, Nunes D. Challenges in the treatment of patients coinfected with HIV and hepatitis C virus: need for team care. *Clin Infect Dis*. 2005;40(Suppl 5):S349-S354.

A Cross-sectional Measurement of Medical Student Empathy

Daniel Chen, MD[1], Robert Lew, PhD[2,3], Warren Hershman, MD, MPH[1], and Jay Orlander, MD, MPH[1,4]

[1]Section of General Internal Medicine, Evans Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [2]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; [3]MAVERIC, VA Boston Healthcare System, Boston, MA, USA; [4]Medical Service, VA Boston Healthcare System, Boston, MA, USA.

BACKGROUND: Empathy is important in the physician–patient relationship. Prior studies have suggested that physician empathy may decline with clinical training.

OBJECTIVE: To measure and examine student empathy across medical school years.

DESIGN AND PARTICIPANTS: A cross-sectional stud of students at Boston University School of Medicine in 2006. Incoming students plus each class near the end of the academic year were surveyed.

MEASUREMENTS: The Jefferson Scale of Physician Empathy–Student Version (JSPE-S), a validated 20-item self-administered questionnaire with a total score ranging from 20 to 140. JSPE-S scores were controlled for potential confounders such as gender, age, anticipated financial debt upon graduation, and future career interest.

RESULTS: 658 students participated in the study (81.4% of the school population). The first-year medical student class had the highest empathy scores (118.5), whereas the fourth-year class had the lowest empathy scores (106.6). Measured empathy differed between second- and third-year classes (118.2 vs 112.7, $P<.001$), corresponding to the first year of clinical training. Empathy appears to increase from the incoming to the first-year class (115.5 vs 118.5, $P=.02$). Students preferring *people-oriented* specialties had higher empathy scores than students preferring *technology-oriented* specialties (114.6 vs 111.4, $P=.002$). Female students were more likely than male students to choose people-oriented specialties (51.5 vs 26.9%, $P<.001$). Females had higher JSPE-S scores than males (116.5 vs 112.1, $P<.001$). Age and debt did not affect empathy scores.

CONCLUSIONS: Empathy scores of students in the preclinical years were higher than in the clinical years. Efforts are needed to determine whether differences in empathy scores among the classes are cohort effects or represent changes occurring in the course of medical education. Future research is needed to confirm whether clinical training impacts empathy negatively, and, if so, whether interventions can be designed to mitigate this impact.

Received February 19, 2007
Revised June 12, 2007
Accepted July 3, 2007
Published online July 26, 2007

KEY WORDS: empathy; medical student education; physician attitudes.
J Gen Intern Med 22(10):1434–8
DOI 10.1007/s11606-007-0298-x

INTRODUCTION

Empathy is the cornerstone of the physician–patient relationship. It is the physician's ability to cognitively recognize a patient's perspectives and experiences, and convey such an understanding back to the patient.[1,2] This understanding allows the patient to feel respected and validated.[3,4] Empathy promotes patient and physician satisfaction, contributes to patient enabling and participation, and may improve patient outcomes.[1,5–10] Furthermore, empathy improves the quality of data obtained from the patient, improves the physician's diagnostic ability, and decreases the rate of miscommunication and lawsuits.[1,3,4,11]

There is concern among educators that clinical training may have an adverse effect on medical resident and student empathy. In one cohort of internal medical residents, empathy was measured to be highest at the beginning but decreasing by the end of internship, and remained low through to the end of residency.[12,13] The work-related challenges, including long work hours and sleep deprivation, are reasons believed to contribute to this decline.[14] Studies among students have shown that empathy measured over the third year of one cohort of medical students declined,[15] and that a single medical school class had higher measured empathy at the start compared to the end of medical school.[16]

This study investigates empathy more closely across the entirety of medical school education while controlling for the potential confounding effects of gender, age, anticipated financial debt upon graduation, and future career interests.

METHODS

This is a cross-sectional study of all medical students enrolled at Boston University School of Medicine (BUSM) during 2006. This study was approved by the Boston University Medical Center Institutional Review Board.

STUDY PARTICIPANTS

All incoming medical students and those completing first- through fourth-year medical school in 2006 were eligible to participate in the study.

The BUSM curriculum is a traditional 4-year medical school with 2 years of preclinical study, with limited patient contact, followed by 2 years of clinical clerkships and electives.

STUDY DESIGN

One author (DC) distributed the self-administered, anonymous surveys to the medical students between March and September 2006. Incoming medical students were surveyed during Orientation Week, before the beginning of first-year medical school classes. First- through fourth-year medical students were surveyed during classes or class events, where attendance was recommended but not mandatory, at the end of their academic year. In total, 5 medical school classes were studied.

The primary measure of empathy, the *Jefferson Scale of Physician Empathy–Student Version* (JSPE-S), is a 20-item psychometrically validated instrument measuring components of empathy among health professionals in patient care situations.[2,17,18] Respondents indicate their level of agreement to each item on a 7-point Likert scale. The JSPE-S total score ranges from 20 to 140, with higher values indicating a higher degree of empathy.[2,17,18] In past studies, total scores among medical students ranged from 115 to 123.1 and standard deviations ranged from 9.9 to 14.1.[2,15,19,20]

Students also specified gender, age, anticipated financial debt, and likelihood of choosing various specialties. Gender was included because practicing female physicians and medical students have been found to have higher empathy than their male counterparts.[2,19] As empathy involves aspects of perception and concern, which may be gained with more maturity, we included age as a confounder.[21] The anticipated level of financial indebtedness at graduation was assessed to the nearest $25,000. Financial indebtedness may potentially influence the selection of career choice and cause high debt students to prefer more lucrative (often technical) specialties.

Students indicated their career specialty intentions, in terms of likelihood of entering each of the specialties listed in Table 1, on a 4-point Likert scale (very unlikely=1,...very likely=4). The *people-oriented* and *technology-oriented* specialty categorizations were based on categories determined in prior studies.[2,17] Each student was assigned to one of these two categories after comparing his or her average Likert score for each group of specialties. For example, if the average score for all people-oriented specialties was 2.0 and the average score for technology-oriented specialties was 2.8, the student was considered preferring technology-oriented specialties. Students with no difference in their scores were not included in analyses of specialty preference. We believed that students with higher measured empathy might associate with the people-oriented careers. As such, student career preferences could potentially confound our results and, thus, needed to be controlled. This construct does not imply that career preference calibrates empathy but instead that students who feel that empathy enhances their skills would gravitate toward higher levels of patient contact.

Table 1. Career Preference Categories [17,18]

People-oriented specialties	Technology-oriented specialties
Internal medicine	Pathology
Family medicine	Surgery and surgical subspecialties
Pediatrics	Radiology
Neurology	Radiation oncology
Rehabilitation medicine	Anesthesiology
Psychiatry	
Emergency medicine	
Obstetrics and gynecology	
Ophthalmology	
Dermatology	

A nonresponder was defined as a student who failed to return an administered survey. An adequate response to the survey was defined as having 16 or more of the 20 JSPE-S questions answered. Surveys with fewer than 16 JSPE-S questions answered were discarded. In cases where surveys were incomplete but had more than 16 JSPE-S responses, we prorated the total scores to give a score with a denominator of 140.

Missing values were rare for most demographic factors and were simply imputed: age (overall mean) and debt category (mode). Missing values could not be imputed in a simple way for gender as imputation of gender affected the analysis. Thus, several approaches were taken. First, data were stratified into three groups, *male*, *female*, and *gender unspecified*. Second, using cases where gender was known as the end point, we constructed a discriminate function from the 20 JSPE-S questions to discriminate male from female. Next, we applied this rule to the gender unspecified subgroup and imputed their gender.

Descriptive statistics and analyses of variance (ANOVAs) were run to compare the different JSPE-S scores among the different classes and categorized groups, whereas controlling for the effects of gender, age, anticipated financial indebtedness, and career preference. Post hoc ANOVA pairwise comparisons were made using Tukey's HSD test. All computations were done with SAS statistical software version 9.

RESULTS

Of the 723 surveys distributed, 658 surveys were returned for an overall response rate of 91.0%. These 658 respondents represent 81.4% of the total students at BUSM in 2006. No differences are seen in the demographic features (age and gender) between responders and nonresponders in the medical school (data not shown). Third-year students have the lowest response rate and the fourth year students have the lowest percentage surveyed (see Table 2).

Table 2 shows the number of students by class among the 658 responding medical students. The number of surveys used in the analysis was 648 because 10 surveys had fewer than 16 out of 20 responses.

The primary multivariate ANOVA considers 4 factors: class, gender, anticipated financial debt, and career preference as well as age. The initial ANOVA model contains all interactions, but highly nonsignificant interaction terms are discarded (data not shown). Hence, the ANOVA factors of interest are class ($P<.001$), gender ($P<.001$), age ($P=.04$), debt ($P=.71$), career preference ($P=.003$), and the gender–class interaction term ($P=.11$).

The 15 subjects with unspecified gender have the lowest mean total scores, indicating that the gender values are not missing at random. A discriminant function based on the 20-item JSPE-S, applied to the gender unspecified surveys, classifies all the unspecified surveys as males. While ex-

Table 2. Demographics and Characteristics of the Medical School Classes

	Incoming	First-year	Second-year	Third-year	Fourth-year	Totals all classes
Number of students	179	160	148	167	154	808
Number of surveys distributed	179	146	147	149	102	723
Number of responders (response rate %)	172 (96.1%)	138 (94.5%)	142 (96.6%)	115 (77.2%)	91 (89.2%)	658 (91.0%)
Percentage of class surveyed (%)	96.1%	86.3%	96.0%	68.9%	59.1%	81.4%

treme, the resultant proportions become more concordant with the proportions of male and female in the medical school (data not shown).

Table 3 shows the JSPE-S scores by class. The first-year medical school class has the highest empathy scores (118.5), whereas the fourth-year class has the lowest empathy scores (106.6). No difference is seen between first- and second-year classes (118.5 vs 118.2, P=.77), or between third- and fourth-year classes (112.7 vs 106.6, P=.10). There is a difference between second- and third-year classes (118.2 vs 112.7, P<.001), which corresponds to the first clinical year in medical school. There is also a difference in JSPE-S scores between incoming and first-year classes (115.5 vs 118.5, P=.02), and between incoming and second-year classes (115.5 vs118.2, P= .04). The incoming class has suggestive differences in empathy scores when compared to the third-year class (115.5 vs 112.7, P=.05), and the incoming class differs from the fourth-year class (115.5 vs 106.6, P=.02).

When looking at the differences in JSPE-S scores by gender, female medical students have higher empathy than male medical students (116.5 vs 112.1, P<.001). Students preferring people-oriented specialties as a career have higher empathy than students preferring technology-oriented specialties (114.6 vs 111.4, P=.002). Age, while significant, has a small effect on empathy scores (scores rise 0.6 with age), but has no effect on other associations. Female students prefer people-oriented specialties more than men (61.9 vs 36.1%, P<.001).

In our analysis, no association is noted between career preference and anticipated financial debt among women (P=.33) or men (P=.96). There is no relationship seen between gender and anticipated financial indebtedness (P=.29) or between different medical school classes and anticipated financial indebtedness (P=.59). In addition, we find that 72.1% of medical students anticipate having more than $200,000 debt after graduation, whereas 14.8% of students anticipate having less than $25,000 debt.

Table 3. JSPE-S Scores by Medical School Class

Class	JSPE-S score	Standard error
Incoming	115.5[A]	1.8
First-year	118.5[B]	1.8
Second-year	118.2[B]	1.8
Third-year	112.7[A,C]	1.9
Fourth-year	106.6[C]	2.3

The class was adjusted for gender, age, anticipated financial indebtedness, career preference, and gender–class interaction.
Groups that share the same superscript are not significantly different from one another. All other differences in JSPE-S scores are significant at the P<.05 level.

DISCUSSION

In our cross-sectional study, empathy appears to increase during the first year of medical school, but falls after the third year (first clinical year) and remains down through the final year of medical school. JSPE-S scores differ by as great as 11.9 points between the first- and fourth-year classes after adjusting for gender, age, financial indebtedness, and career preferences.

Our results, although cross-sectional, are consistent with previous studies, suggesting that empathy decreases after clinical training in medical school. Using the JSPE-S, one cohort of medical students had a decline in empathy during the third year of medical school.[15] This group of 125 third-year medical students exhibited a postclerkship decline in empathy of 2.5 points (123.1 to 120.6). The authors found no significant associations between changes in empathy scores and gender, age, or academic performance on step 1 of the USMLE.[15] Another group measured empathy in a cohort of medical students at the beginning of medical school and just before graduation and found lower empathy scores in the graduating class.[16] Among another group of health care professionals, dental students, empathy scores also decreased after patient care responsibilities began.[20] The only other cross-sectional study of multiple medical school classes that we could find did not demonstrate differences in empathy across classes, but this study used an outcome measure, which was not specifically designed for health professionals.[22]

Studies of medical resident empathy have noted similar declines. A cross-sectional study in an internal medicine program observed that first-year residents scored 4 points higher on the JSPE–Physician Version than third-year residents (117.5 vs 113.5, P=.31).[11]

Whereas these studies lack an assessment of behavior, one recent report showed a positive association between the individuals' scores on the JSPE-S in medical school and ratings of their empathic behavior made by their residency program director 3 years later.[23] This study suggests a long-term predictive validity of the self-report empathy scale.

Various stressful aspects of medical education and training, such as long work-hours and sleep deprivation, as well as dependence on technology for diagnoses, shorter patient hospitalizations, and limited bedside interactions may contribute to decreases in empathy.[14,24–27] In response, some programs now include clinical narrative or critical incident writing; classes on medically themed creative writing, literature and art; journal writing; and use of standardized patients in the medical education curriculum to maintain or increase empathy.[28–34] Studies offer conflicting results with respect to their impact on empathy. One group preliminarily measured an increase in empathy in students who participated in role-playing and simulated patient scenarios.[35] In contrast, an entire

medical school class taking a 4-month patient-interviewing course designed to teach communication and emphasize empathy did not show an improvement in the latter.[16] A recent review suggests that empathy may be amenable to a range of interventional strategies.[36] Qualitative data from independent observations and unvalidated surveys note that these interventions improve student communication skills and empathy. Lastly, student course evaluations and feedback suggest that students respond positively to these educational interventions and perceive themselves to be more sensitive and empathic toward their patients from such activities despite a lack of more objective outcomes.[26,28,31–34,36,37]

Another possible explanation for the observed decrease in empathy may be an acculturation phenomenon. Student doctors experience a wide range of emotions and stresses and may struggle to maintain their empathy.[14,38] This would suggest that to remain effective for patients, students and trainees become less empathic as they face emotionally challenging and draining situations. Outcome measures to assess such a hypothesis should be included in future research.

We found that medical students expressing a preference for people-oriented specialties had higher empathy scores than those expressing a preference for technology-oriented specialties. These data are consistent with another study, which found that students likely choosing family medicine, internal medicine, psychiatry, pediatrics, and obstetrics and gynecology had higher empathy scores than all other specialties, when controlled for gender effects.[22]

Previous studies have demonstrated a difference in empathy among practicing physicians of different specialties. Physicians in people-oriented specialties, such as primary care specialties (family medicine, internal medicine, and pediatrics), obstetrics and gynecology, emergency medicine, psychiatry, and medical subspecialties, had higher average empathy scores than physicians in technology-oriented specialties—anesthesiology, radiology, pathology, surgery, and surgical subspecialties (see Table 1).[2,17] Psychiatrists had the highest mean JSPE–Physician Version score (127.0), primary care specialists scored from 120–122, and the lowest values were noted in orthopedic surgeons and anesthesiologists (approx. 116).[18]

Students may possibly be prestratified in career preferences before coming to medical school, with those students who are naturally endowed with more empathy attracted to people-oriented specialties. Although we categorized medical students as preferring either people-oriented or technology-oriented specialties, the vast majority of incoming and first- through third-year medical students had small mean differences ranging from 0.46 to 0.62 when comparing their average Likert score for the people-oriented and technology-oriented specialty groups. This suggests that they may not be definitive in their career preference early in medical school and raises the possibility of changing career preferences with different experiences in medical school. Future studies should determine whether fostering empathy skills impacts student career preferences. With fewer graduating medical students selecting careers in primary care (people oriented) specialties,[39–41] if enhancement in empathy can be achieved and be shown to modify career preference, potential policy implications regarding medical curricula and resource allocation could be possibly driven by societal or regional needs for primary care clinicians.

There are several limitations of our study. First, our measurement of empathy, the JSPE-S, is self-reported. It measures medical students' orientation to empathy and is not correlated with behavior. A recently demonstrated correlation between individuals' empathy scores at the beginning of third year of medical school and ratings of their empathic behavior at the end of their first year of postgraduate training does suggest predictive validity of the JSPE-S.[23] Studies of practicing physicians have noted that JSPE score difference is as great as 11 between practicing psychiatrists and anesthesiologists,[18] a range difference seen in some of our comparisons.

A second limitation of our study is the possibility of cohort effects. We recognize this as a limitation of all cross-sectional studies. However, except for our new finding of differences seen in the preclinical years, our data are consistent with other studies of medical student empathy, which suggest a decline during medical school.[15,16] Unexpectedly, we found that there is an increase in empathy scores from beginning to end of first-year among the medical students. As the JSPE-S questions were designed to assess the empathy of health care providers in patient-provider situations, it is possible that a complete lack of clinical exposure impacts how the JSPE-S is completed by incoming students and hence the instrument may be invalid in this group. Whereas there is limited patient contact in the first 2 years of our medical school, students do interact with patients when shadowing practicing physicians and participating in their clinical skills training courses. Patient contact in the context of the first 2 years of medical school possibly alters the perception of students so that the subjective anchors on JSPE-S questions and Likert-scale anchors are interpreted differently. Alternatively, the limited clinical exposure during the first 2 years of medical school may positively influence medical students' empathy by reinforcing their desire to help people through medicine. A third hypothesis is that this represents a cohort effect. The robustness of our observation could be tested by sequentially tracking these medical student cohorts.

Lastly, we acknowledge that attendance, survey participation, and possibly response (e.g., level of empathy) may be influenced by the situations or events during which we obtained the data.

Although our study is limited to one medical school, we feel that our results can be generalized to medical schools that have a traditional structure similar to ours. In particular, we are struck by the consistency of the decline in score after a full year of clinical exposure.

CONCLUSION

Empathy is important in the physician–patient relationship and has clear benefits for the patient and the physician. We found that there are differences in the empathy among the different classes and that empathy declines with increased clinical training in medical school. Whether the decline is reflective of the prevalent teaching methods and modifiable with better methods or is an unavoidable psychological effect of the acculturation process into the medical profession is not yet known.

The association of measured empathy and career preference among medical students is interesting. Although medical students indicate a career preference, the vast majority of

them are not strongly committed in their choice in the first 3 years of medical school. This association suggests the possibility that career preferences can possibly be changed with changes in empathy. Current available data on the impact of interventions does not provide conclusive evidence that empathic behavior can be effectively and permanently improved. Future interventions should examine relationships between empathy, career preferences, and links with clinician behavior, as such finding would have the largest impact on educational policy and practice.

Acknowledgment: *Permission to use the JSPE-S was obtained from the Jefferson Medical College Center for Research in Medical Education and Health Care. We thank Phyllis Carr, MD, BUSM, for her role in reviewing the manuscript.*

Funding sources: *None of the authors received any funding support for the study.*

Conflict of interest statement: *None disclosed.*

Corresponding Author: *Daniel Chen, MD; Section of General Internal Medicine, Evans Department of Medicine, Boston University School of Medicine, 91 East Concord Street, MAT 2, Boston, MA 02118, USA (e-mail: daniel.chen@bmc.org).*

REFERENCES

1. **Coulehan JL, Platt FW, Egener B, et al.** "Let me see if I have this right...": words that help build empathy. Ann Intern Med. 2001;135:221–7.
2. **Hojat M, Mangione S, Nasca TJ, et al.** The Jefferson Scale of Physician Empathy: development and preliminary psychometric data. Educ Psychol Meas. 2001;61:349–65.
3. **Beckman HB, Markakis KM, Suchman AL, Frankel RM.** The doctor–patient relationship and malpractice: lessons from plaintiff depositions. Arch Intern Med. 1994;154:1365–70.
4. **Levinson W, Roter DL, Mullooly JP, Dull VT, Frankel RM.** The relationship with malpractice claims among primary care physicians and surgeons. JAMA. 1997;277:553–9.
5. **Bikker AP, Mercer SW, Reilly D.** A pilot prospective study on the consultation and relational empathy, patient enablement, and health changes over 12 months in patients going to the Glasgow Homoeopathic Hospital. J Altern Complement Med. 2005;11:591–600.
6. **Kim SS, Kaplowitz S, Johnston MV.** The effects of physician empathy on patient satisfaction and compliance. Eval Health Prof. 2004;27:237–51.
7. **MacPherson H, Mercer SW, Scullion T, Thomas KJ.** Empathy, enablement, and outcome: an exploratory study on acupuncture patients' perceptions. J Altern Complement Med. 2003;9:869–76.
8. **Mercer SW, Reynolds WJ.** Empathy and quality of care. Br J Gen Pract. 2002;52(suppl):S9–12.
9. **Vermeire E, Hearnshaw H, Van Royen P, Denekens J.** Patient adherence to treatment: three decades of research. A comprehensive review. J Clin Pharm Ther. 2001;26:331–42.
10. **Suchman AL, Roter D, Green M, Lipkin M.** Physician satisfaction with primary care office visits. Collaborative Study Group of the American Academy on Physician and Patient. Med Care. 1993;31:1083–92.
11. **Mangione S, Kane GC, Caruso JW, Gonnella JS, Nasca TJ, Hojat M.** Assessment of empathy in different years of internal medicine training. Med Teach. 2002;24:370–3.
12. **Bellini LM, Baime M, Shea JA.** Variation of mood and empathy during internship. JAMA 2002;287:3143–6.
13. **Bellini LM, Shea JA.** Mood change and empathy decline persist during three years of internal medicine training. Acad Med. 2005;80:164–7.
14. **Rosen IM, Gimotty PA, Shea JA, Bellini LM.** Evolution of sleep quantity, sleep deprivation, mood disturbances, empathy and burnout among interns. Acad Med. 2006;81:82–5.
15. **Hojat M, Mangione S, Nasca TJ, et al.** An empirical study of decline in empathy in medical school. Med Educ. 2004;38:934–41.
16. **Diseker RA, Michielutte R.** An analysis of empathy in medical students before and following clinical experience. J Med Educ. 1981;56:1004–10.
17. **Hojat M, Gonnella JS, Nasca TJ, Mangione S, Veloksi JJ, Magee M.** The Jefferson Scale of Physician Empathy: further psychometric data and differences by gender and specialty at item level. Acad Med. 2002;77 (suppl):S58–60.
18. **Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M.** Physician empathy: definition, components, measurement, and relationship to gender and specialty. Am J Psychiatry. 2002;159:1563–9.
19. **Hojat M, Gonnella JS, Mangione S, et al.** Empathy in medical students as related to academic performance, clinical competence and gender. Med Educ. 2002;36:522–7.
20. **Sherman JJ, Cramer A.** Measurement of changes in empathy during dental school. J Dent Educ. 2005;69:338–45.
21. **Davis MH.** Measuring individual differences in empathy; evidence for a multidimensional approach. J Pers Soc Psychol. 1983;44:113–26.
22. **Newton BW, Savidge MA, Barber L, et al.** Differences in medical students' empathy. Acad Med. 2000;75:1215.
23. **Hojat M, Mangione S, Nasca TJ, Gonnella JS, Magee M.** Empathy scores in medical school and ratings of empathic behavior in residency training 3 years later. J Soc Psychol. 2005;145:663–72.
24. **Benbassat J, Baumal R.** What is empathy, and how can it be promoted during clinical clerkships? Acad Med. 2004;79:832–9.
25. **Kay J.** Traumatic deidealization and the future of medicine. JAMA. 1990;263:572–3.
26. **Kramer D, Ber R, Moore M.** Increasing empathy among medical students. Med Educ. 1989;23:168–73.
27. **Marcus ER.** Empathy, humanism, and the professionalization process of medical education. Acad Med. 1999;74:1211–5.
28. **Anderson R, Schiedermayer D.** The art of medicine through the humanities: an overview of a one-month humanities elective for fourth year students. Med Educ. 2003;37:560–2.
29. **Charon R.** Narrative medicine: form, function, and ethics. Ann Intern Med. 2001;134:83–7.
30. **Charon R.** Narrative and medicine. N Engl J Med. 2004;350:862–4.
31. **DasGupta S, Charon R.** Personal illness narratives: using reflective writing to teach empathy. Acad Med. 2004;79:351–6.
32. **Hatem D, Ferrara E.** Becoming a doctor: fostering humane caregivers through creative writing. Patient Educ Couns. 2001;45:13–22.
33. **Lichstein PR, Young G.** "My most meaningful patient": reflective learning on a general medicine service. J Gen Intern Med. 1996;11: 406–9.
34. **Shapiro J, Lie D.** Using literature to help physician-learners understand and manage "difficult" patients. Acad Med. 2000;75:765–8.
35. **Feighny KM, Monaco M, Arnold L.** Empathy training to improve physician–patient communication skills. Acad Med. 1995;70:435–6.
36. **Stepien KA, Baernstein A.** Educating for empathy: a review. J Gen Intern Med. 2006;21:524–30.
37. **Winefield HR, Chur-Hansen A.** Evaluating the outcome of communication skill teaching for entry-level medical students: does knowledge of empathy increase? Med Educ. 2000;34:90–4.
38. **Branch W, Pels RJ, Lawrence RS, Arky R.** Becoming a doctor. critical-incident reports from third-year medical students. N Engl J Med. 1993;329:1130–2.
39. **Bodenheimer T.** Primary care—will it survive? N Engl J Med. 2006;355:861–4.
40. **Joyce CM, McNeil JJ.** Fewer medical graduates are choosing general practice: a comparison of four cohorts, 1980–1995. Med J Aust. 2006;185:102–4.
41. American College of Physicians: The impending collapse of primary care medicine and its implications for the state of the nation's health care. http://www.acponline.org/hpp/statehc06_1.pdf. Accessed September 19, 2006.

Impact of Hepatitis C on HIV Progression in Adults With Alcohol Problems

Debbie M. Cheng, David Nunes, Howard Libman, John Vidaver, Julie K. Alperen, Richard Saitz, and Jeffrey H. Samet

Background: Coinfection of human immunodeficiency virus (HIV) and hepatitis C virus (HCV) is a substantial medical and public health concern due to its increasing prevalence and complex patient management. Alcohol use may worsen HCV-related liver disease and interfere with adherence to antiretroviral therapy (ART) and medical care. We therefore studied the association between HCV infection and markers of HIV disease progression in adults with alcohol problems.

Methods: This is a longitudinal study of 396 HIV-infected persons with alcohol problems, 199 (50%) of whom were coinfected with HCV (positive HCV RNA test). CD4 cell counts and HIV RNA levels were assessed at baseline and then every 6 months for up to 42 months. Hepatitis C virus RNA status was determined at study enrollment. We examined the relationship between HCV infection and laboratory markers of HIV progression (CD4 cell count and $\log_{10}$ HIV RNA) by fitting multivariable longitudinal regression models for each outcome.

Results: Among subjects who were adherent to ART, the presence of HCV infection was associated with a lower CD4 cell count (adjusted mean difference -46.0 cells/μL, $p = 0.03$). There was no association observed between HCV infection and CD4 cell count among those not adherent to ART or those not taking ART. No significant association was observed between HCV infection and HIV RNA regardless of ART status.

Conclusions: Hepatitis C virus infection has an adverse effect on CD4 cell count in patients with alcohol problems who are adherent to ART. Addressing HCV coinfection among these patients may confer additional immunologic benefit for this patient population.

Key Words: Hepatitis C Virus, HIV, Coinfection, Disease Progression, Alcohol Abuse.

COINFECTION WITH HUMAN immunodeficiency virus (HIV) and hepatitis C virus (HCV) is a substantial medical and public health concern because of its increasing prevalence and problematic patient management. With improved treatment of HIV infection, coinfection with HCV is responsible for substantial morbidity and mortality in this patient population (Sulkowski et al., 2000). An estimated 15 to 30% of HIV-infected persons in the United States and Europe are coinfected with HCV (Sulkowski et al., 2002). In certain subgroups of HIV-infected patients, the prevalence of HCV coinfection is even higher. For example, 60 to 90% of HIV-infected persons with a history of injection drug use are coinfected with HCV (Bonacini and Puoti, 2000; Sulkowski and Thomas, 2003).

The impact of HCV on HIV disease progression has been an area of active investigation with conflicting conclusions (Braitstein et al., 2006; De Luca et al., 2002; Dorrucci et al., 1995; Greub et al., 2000; Klein et al., 2003; Macias et al., 1998; Piroth et al., 1998; Rancinan et al., 2002; Staples et al., 1999; Sulkowski et al., 2000; Sulkowski and Thomas, 2003; Weis et al., 2006). For example, the results of the 2 largest studies to date addressing this issue are inconsistent. Greub et al. (2000) reported that HCV infection was associated with decreased survival and more rapid progression to acquired immunodeficiency syndrome (AIDS) in 3,111 HIV-infected persons enrolled in the Swiss HIV Cohort Study. However, Sulkowski et al. (2002) found no association between HCV infection and

From the Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts (DMC); the Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, Massachusetts (DMC, JKA, RS, JHS); the Section of Gastroenterology, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts (DN); the Department of Medicine, Beth Israel Deaconess Medical Center, Divisions of General Medicine and Primary Care and Infectious Diseases, Harvard Medical School, Boston, Massachusetts (HL); DM-STAT, Inc., Malden, Massachusetts (JV); the Department of Epidemiology and Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, Massachusetts (RS); and the Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts (JHS).

Received for publication June 27, 2006; accepted January 29, 2007.

This work was supported by the National Institute on 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).

Reprint requests: Debbie M. Cheng, ScD, Department of Biostatistics, Boston University School of Public Health, 715 Albany Street, T4E, Boston, MA 02118; Fax: 617-414-4676; E-mail: dmcheng@bu.edu

DOI: 10.1111/j.1530-0277.2007.00381.x

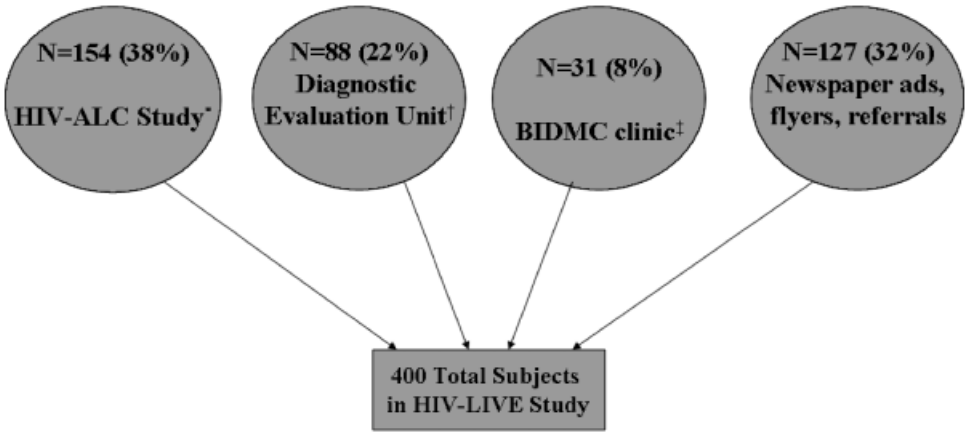


Fig. 1. Distribution of subject recruitment.

survival, progression to AIDS, or response to highly active antiretroviral therapy (ART) in a prospective cohort study of 1,955 subjects from an urban HIV clinic in the United States.

Recommendations for management of HCV infection in HIV-infected patients continue to evolve (Alberti et al., 2005; Soriano et al., 2005), although HCV treatment is often the exception rather than the rule (Fleming et al., 2003). Patient management is complex because of the limited effectiveness of and numerous contraindications to HCV therapy, its duration and adverse effects, potential hepatotoxicity of ART, and lack of clarity concerning the effect of HCV infection on HIV disease progression (Chung et al., 2004; Sulkowski and Thomas, 2003; Torriani et al., 2004).

The objective of this study was to test the hypothesis that HCV infection has an adverse effect on markers of HIV disease progression in adults with alcohol problems. Previous studies of patients starting ART have suggested that HCV infection is associated with a blunted CD4 cell recovery (Greub et al., 2000). The biological mechanism of such an effect is unclear, but HCV is known to replicate in mononuclear cells and recent studies have suggested that both HCV infection and HCV proteins may enhance lymphocyte apoptosis (Nunez et al., 2006; Thoren et al., 2004).

It is important to study whether HCV impacts HIV disease progression in subjects with alcohol problems as alcohol use can be hepatotoxic and its use is prevalent among HIV-infected persons (Cook et al., 1997; Krupitsky et al., 2005; Samet et al., 2004b). Alcohol use has also been shown to worsen HCV-related liver disease (Schiff, 1999) and interfere with patient adherence to ART and medical care (Kresina et al., 2002; Samet et al., 2004a). The impact of alcohol on both HCV and HIV-related outcomes makes it imperative to understand the effect of HCV on HIV disease progression among subjects with alcohol problems to inform the medical management of this patient population.

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 alcohol problems. The study sample included a spectrum of alcohol problems that was not limited to a diagnosis of current alcohol abuse or dependence. Data were collected at baseline and every 6 months thereafter for up to 42 months.

Of the 400 subjects, 38% ($n = 154$) were recruited from the HIV-ALC (HIV-Alcohol Longitudinal Cohort) study, a previous cohort study at Boston Medical Center (BMC). Inclusion and exclusion criteria for HIV-ALC and HIV-LIVE were identical. Subjects also were recruited from the Diagnostic Evaluation Unit ($n = 88$, 22%) (Samet et al., 1995), an intake clinic for HIV-infected patients at BMC, and the HIV Primary Care and Specialty Clinics at Beth Israel Deaconess Medical Center (BIDMC) ($n = 31$, 8%). Remaining subjects ($n = 127$, 32%) were recruited through flyers distributed in health care centers, homeless shelters, and drug treatment programs, advertisements in newspapers and referrals from other HIV-LIVE subjects. Figure 1 illustrates the distribution of subject recruitment across various sources.

Eligibility criteria were (1) documented HIV antibody test by ELISA 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 1 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. Eligible subjects who wished to participate provided written informed consent before enrollment. Enrollment began August 2001 and ended July 2003. Most interviews took place at the clinical

research units. The Institutional Review Boards of Boston University Medical Center 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, including questions on demographics, HIV risk behaviors, alcohol consumption, and ART use in the past 30 days. Past month alcohol consumption was assessed using a validated calendar method (Sobell and Sobell, 1992). The Composite International Diagnostic Interview (CIDI) 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. Recent drug use and current (past 12 months) diagnosis of drug dependence was assessed at enrollment using the CIDI. Drug use and drug dependence diagnosis (past 6 months) were assessed at each follow-up using the CIDI Short Form (Kessler et al., 1991). Interviews were conducted in English or Spanish.

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. Human immunodeficiency virus RNA testing was performed using either a branched-chain DNA assay or polymerase chain reaction (PCR) (Pachl et al., 1995). The lower threshold of detection was between 50 and 75 copies/mL over the course of the study. All subjects were tested for HCV infection by measurement of HCV antibody. Antibody-positive subjects were tested for HCV RNA if this was unavailable from medical records. Hepatitis C virus RNA was measured using commercially available assays, either by branched chain DNA or PCR-based assays. The lower level of detection of the assays was 615 IU/mL. HCV antibody-negative subjects were assumed to be HCV RNA negative (Thio et al., 2000).

Outcomes

The primary outcomes were CD4 cell count per microliter and $\log_{10}$ HIV RNA copies per milliliter, 2 laboratory markers of HIV disease progression. For analysis purposes, undetectable HIV RNA levels were assigned half the value of the lower limit of detection. CD4 count/total lymphocyte count percentage was analyzed as a secondary endpoint. This value may be more reflective of immune status than absolute CD4 count in the setting of asplenia or advanced liver disease (Zurlo et al., 1995).

Primary Independent Variable

The main independent variable was HCV infection status, defined as HCV RNA positive versus HCV RNA negative.

Potential Confounding Factors

Adherence to antiretroviral medications was determined using the AIDS Clinical Trials Group Questionnaire for Adherence to Anti-Retroviral Medications (Chesney et al., 2000). Patients reported their current antiretroviral medications, number of daily doses, and pills. Patients who reported being <100% adherent during the previous 3 days were considered not adherent; the 100% cut-off was used due to the brief timeframe of assessment. We used information concerning current receipt of ART, defined as receiving at least one antiretroviral medication, and adherence to create a 3-category variable representing ART status: receiving ART and adherent, receiving ART and not adherent, or not receiving ART.

Other potential covariates were gender, age, race (black, white, other), whether the subject was recruited from a previous cohort study of HIV-infected subjects with alcohol problems, homelessness, alcohol consumption, drug dependence diagnosis, injection drug use, time since starting HIV ART, and time since study enrollment. Homelessness was defined as at least 1 night in a shelter or on the street in the past 6 months (Kertesz et al., 2005). Alcohol use was categorized as heavy, moderate, or abstinent. Heavy alcohol use was defined as >14 drinks/wk or $\geq$5 drinks on 1 occasion for men <66 years old and >7 drinks/wk or $\geq$4 drinks on 1 occasion for men $\geq$66 years old and all women (NIAAA, 1995). Moderate alcohol use was defined as any drinking less than heavy amounts but not abstinent. Self-reported information was available on whether subjects injected drugs in the 6 months before enrollment and lifetime injection behaviors. The latter variable was highly correlated with HCV infection, whereas the former was not. Thus, current injection drug use was selected as the covariate to avoid collinearity and because current injection drug use would be more likely to affect HIV disease progression.

Statistical Analysis

Descriptive statistics were used to characterize the study sample at enrollment overall and by HCV RNA status. Continuous variables were compared between HCV RNA groups using *t*-tests or the Wilcoxon rank sum test. Chi-square and Fisher's exact tests were used as appropriate to compare categorical variables. We examined the relationship between HCV infection and HIV progression by fitting separate multivariable longitudinal regression models for each outcome. Generalized linear mixed effects models (Laird and Ware, 1982) were used to account for the correlation from using repeated observations on the same subject. Because the effect of HCV infection on HIV progression was expected to differ based on ART status, interactions between HCV RNA status and ART status were included in all models. The estimated effect of HCV infection is reported separately by ART status. Potential confounding factors were included in regression models. Alcohol use, homelessness, and ART status were included in regression models as time-varying covariates. Baseline CD4 cell count was included as a covariate in analyses of CD4 cell count. Analyses of HIV RNA did not adjust for baseline values as it was expected that subjects had achieved their viral "set points," and thus HIV RNA would remain somewhat constant over time. We evaluated whether alcohol use was a confounder of the relationship between HCV infection and HIV progression by excluding alcohol use and refitting the multivariable regression models. The regression coefficients for HCV infection for the models with and without alcohol use were then compared to assess whether there was confounding by alcohol use. A change in estimate of >10% was used to identify confounding. Secondary analyses assessed whether time since starting HIV ART confounded the relationship between HCV infection and markers of HIV disease progression by including it as an additional covariate in 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., >0.40). All analyses were conducted using 2-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. Of the 396 subjects with known HCV RNA status, 185 were adherent to their HIV medications (99 HCV RNA positive; 86 HCV RNA negative), 61 were not adherent to their HIV medications (29 HCV RNA positive; 32 HCV RNA negative), and 150 were not on HIV medications (71 HCV RNA positive; 79 HCV RNA negative) at study enrollment. These sample sizes allow our study 80% power to detect minimum differences in CD4 cell count of 134, 233, and 149 between HCV RNA positive versus negative among subjects who were adherent, not adherent, and not on HIV medications, respectively. For the outcome log HIV RNA, our study has 80% power to detect differences as small as

0.54, 0.93, and 0.60 between HCV RNA positive versus negative subjects who were adherent, not adherent, and not on HIV medications, respectively. Power calculations were based on 2-sided 2-sample *t*-tests using a significance level of 0.05. Standard deviations of 325 and 1.3 were assumed for CD4 cell count and log HIV RNA, respectively. Power calculations were based on a simple setting that utilizes a single time point, however the data were analyzed using longitudinal regression methods that allow for repeated observations from subjects, therefore we expected to be able to detect smaller differences with equal power.

RESULTS

Study Subjects

Of the 400 HIV-LIVE subjects, 235 (59%) were HCV antibody-positive. Of the 231 HCV antibody-positive subjects who had an HCV RNA test, 199 (86%) were HCV RNA positive. One HCV-infected subject received interferon therapy during the study. Analyses were conducted on 396 subjects with known HCV RNA status.

Baseline characteristics of the study subjects are displayed in Table 1. The majority were male (75%), nonwhite race (67%), and housed (75%). The average age of the subjects was 42.5 years. At enrollment, 14% reported injection drug use (past 6 months), 43% had a current diagnosis (past 12 months) of drug dependence, 70% had a lifetime diagnosis of alcohol dependence and 18% alcohol abuse; 10% had a current diagnosis (past 6 months) of alcohol dependence. Thirty-one percent reported heavy alcohol consumption, 11% moderate consumption, and 58% were abstinent in the past 30 days. Median CD4 cell count was 403 cells/μL. Median $\log_{10}$ HIV RNA was 2.95 and 31% of subjects had undetectable HIV RNA. Sixty-two percent were on ART (93% of these subjects were taking at least 3 HIV medications).

Hepatitis C virus RNA-positive subjects were older (mean age 44 vs 41 years), more likely to have used injection drugs in the past 6 months (23 vs 5%), and had a lower median CD4 cell count (356 vs 424 cells/μL) compared with HCV RNA-negative subjects.

Multivariable Regression Results

The primary analyses used data from 396 subjects contributing 1,171 observations to examine CD4 cell count and 1,542 observations for log HIV RNA. Human immunodeficiency virus RNA analyses included a larger number of observations as log HIV RNA at enrollment was modeled as part of the longitudinal outcome whereas baseline CD4 cell count was a covariate in CD4 cell count analyses. The number of subjects included in CD4 cell count analyses ($n = 344$) was less than that in HIV RNA analyses as only those with a follow-up interview (87%) were included in analyses of CD4 cell count. The median number of observations per subject was 4 (interquartile range: 3–5) for CD4 cell count and HIV RNA. Table 2 presents results of the longitudinal regression analyses. Among subjects adherent to ART, those with detectable

Table 1. Baseline Characteristics of the 396 HIV-Infected Subjects in the Study Cohort, Overall and by HCV RNA Status

	Total sample ($N = 396$)	HCV RNA+ ($N = 199$)	HCV RNA − ($N = 197$)	*p*-Value
Age (y), mean (SD)	42.5 (7.4)	43.9 (7.0)	40.9 (7.5)	<0.0001
Gender				
Female	25%	27%	23%	0.42
Male	75%	73%	77%	
Race				
Black	41%	38%	44%	0.46
White	33%	34%	32%	
Other	26%	28%	24%	
Median CD4 count[a] (IQR)	403 (238, 624)	356 (231, 542)	424 (257, 655)	0.02
Median $\log_{10}$ HIV RNA[b] (IQR)	2.95 (1.57, 4.14)	2.92 (1.57, 4.14)	3.03 (1.40, 4.18)	0.94
Undetectable HIV RNA				
Yes	31%	29%	33%	0.49
No	69%	71%	67%	
HIV ART status[c]				
Adherent	47%	50%	44%	0.48
Not adherent	15%	15%	16%	
Not on meds	38%	36%	40%	
Injection drug use ever[d]				
Yes	55%	89%	20%	<.0001
No	45%	11%	80%	
Injection drug use, past 6 mo[d]				
Yes	14%	23%	5%	<.0001
No	86%	77%	95%	
Drug dependence, past 12 mo				
Yes	43%	44%	40%	0.42
No	57%	56%	60%	
Homeless				
Yes	25%	29%	20%	0.06
No	75%	71%	80%	
Lifetime alcohol dependence				
Dependence	70%	74%	66%	0.14
Abuse	18%	17%	19%	
No diagnosis	12%	9%	15%	
Alcohol dependence, past 6 mo				
Dependence	10%	11%	9%	0.70
Abuse	2%	2%	2%	
No diagnosis	88%	87%	89%	
Alcohol consumption, past mo[d]				
None	58%	63%	53%	0.09
Moderate	11%	8%	13%	
Heavy	31%	29%	34%	
Participated in previous HIV study[e]				
Yes	38%	43%	33%	0.04
No	62%	57%	67%	
Time since diagnosis of HIV (y), mean (SD)	9.2 (4.9)	9.8 (4.9)	8.6 (4.8)	0.02
Time since starting HIV ART (y), mean (SD)	6.9 (4.1)	7.1 (4.2)	6.7 (4.0)	0.30

[a] $N = 375$, 187, and 188, respectively.
[b] $N = 361$, 180, and 181, respectively.
[c] Adherent: $N = 185$, 99, and 86, respectively; Not adherent: $N = 61$, 29, and 32, respectively; Not on meds: $N = 150$, 71, and 79, respectively.
[d] $N = 396$, 200, and 196, respectively.
[e] "Yes" indicates subjects were previously enrolled in an observational cohort study conducted by the study investigators.
HIV, human immunodeficiency virus; ART, antiretroviral therapy.

Table 2. Adjusted Mean Differences in CD4 Cell Count, Log_{10} HIV RNA, and Percent CD4 Cells from Longitudinal Regression Models Comparing HCV RNA-Positive Subjects Versus HCV RNA-Negative Subjects According to ART Status

		CD4 cell count[a]		Log_{10} HIV RNA[b]		Percent CD4 cells[c]	
ART status	HCV status	Adjusted mean difference (SE)[d]	*p*-Value	Adjusted mean difference (SE)[d]	*p*-Value	Adjusted mean difference (SE)[d]	*p*-Value
Adherent to ART	HCV RNA+	**−46.0 (21)**	**0.03**	0.086 (.12)	0.46	**−2.3 (.76)**	**<0.01**
	HCV RNA−	—		—		—	
Not adherent to ART	HCV RNA+	−4.5 (28)	0.88	−0.010 (0.16)	0.96	−1.3 (1.0)	0.18
	HCV RNA−	—		—		—	
Not on ART	HCV RNA+	−16.3 (23)	0.48	0.23 (0.13)	0.08	−0.48 (0.84)	0.57
	HCV RNA−	—		—		—	

Bold font indicates significant differences ($p<0.05$).
[a]Analyses based on 344 subjects and 1,171 observations.
[b]Analyses based on 395 subjects and 1,542 observations.
[c]Analyses based on 341 subjects and 1,160 observations.
[d]All analyses adjusted for: ART status, interaction between HCV status and ART status, gender, age, race, homelessness, participation in previous HIV cohort study, alcohol consumption, drug dependence diagnosis, injection drug use past 6 mo, time, and baseline value of the outcome (CD4 cell count analyses only).
HIV, human immunodeficiency virus; ART, antiretroviral therapy.

HCV RNA had lower CD4 cell counts (adjusted mean difference −46.0 in CD4 cells/μL, *p*-value = 0.03). No association was observed between HCV infection and CD4 cell count for subjects not adherent to ART or not taking ART. No significant association was observed between HCV RNA status and log HIV RNA regardless of ART status (Table 2).

For the secondary endpoint percent CD4 cells, HCV RNA positivity was associated with a lower percentage of CD4 cells (adjusted mean difference −2.3%, *p*-value <0.01) among those adherent to ART. No association was observed between HCV infection and percentages of CD4 cells among those not adherent to ART and those not on ART (Table 2).

We also evaluated whether alcohol use was a confounder of the relationship between HCV infection and HIV disease progression. In the multivariable regression models that did not adjust for alcohol use, the estimated effects of HCV infection were similar to the primary models that adjusted for alcohol use, suggesting that alcohol use was not a confounder. Among subjects adherent to ART, those with detectable HCV RNA had lower CD4 cell counts (adjusted mean difference −44.9 in CD4 cells/μL, *p*-value = 0.03) and lower percentage of CD4 cells (adjusted mean difference −2.2%, *p*-value <0.01).

The results were similar after adjusting for time since starting HIV ART in the regression models. Among subjects adherent to ART, subjects with detectable HCV RNA had significantly lower CD4 cell counts (adjusted mean difference 44.5 in CD4 cells/μL, *p*-value = 0.03) and percentage of CD4 cells (adjusted mean difference 2.2%, *p*-value <0.01).

DISCUSSION

Coinfection with HCV is common in HIV-infected patients, especially among those with injection drug use. Determining whether HCV coinfection affects HIV disease progression is important for understanding the optimal timing and sequencing of therapies for these infections. This is a particularly key issue for coinfected persons with alcohol problems due to several complicating factors: alcohol exposure can result in more rapid progression of liver disease (Schiff, 1999); heavy alcohol use is a contraindication for interferon therapy (Soriano et al., 2005); and alcohol use is associated with worse adherence to ART (Kresina et al., 2002). Analysis of this prospective cohort of HIV-infected persons with alcohol problems demonstrated a significant association between coinfection with HCV and lower CD4 cell counts among patients adherent to ART. Reductions were present in both CD4 cell count and percentage, suggesting these effects were probably not related to hypersplenism secondary to the underlying liver disease. Analyses controlled for multiple characteristics associated with HIV disease progression.

The observation of a lower CD4 cell count in patients with HCV infection that were receiving and adherent to ART is of particular interest. These findings are consistent with those described by Greub et al. (2000) in the Swiss HIV Cohort Study in which initiation of ART in HCV coinfected persons was associated with an attenuated recovery in CD4 cell count compared with those without HCV. Similar to subjects adherent to ART in the current study, Greub and colleagues found a decreased response of CD4 cell count to ART among coinfected persons despite no observed differences in HIV RNA. Interestingly, a follow-up report from the Swiss HIV Cohort Study investigators showed that differences in CD4 cell count between these groups did not persist after 4 years of follow-up (Kaufmann et al., 2003). This "muting" effect of HCV infection on CD4 cell count response to ART has been shown in other studies (De Luca et al., 2002; Law et al., 2004), but differs from the recently published EUROSIDA

study and a smaller prospective U.S. study which showed no effect of HCV on CD4 cell recovery (Chung et al., 2002; Rockstroh et al., 2005).

The reason for discrepancies across study findings is unclear. They may be related in part to the timepoints at which CD4 cell recovery was measured, particularly if HCV infection delays but does not prevent immunologic reconstitution. Why HCV should affect CD4 cell count recovery in the face of similar HIV viral suppression is unknown, though it may indicate that HCV infection has a direct effect on CD4 cells rather than being mediated through HIV viral replication. This study's finding that coinfection of HCV and HIV is associated with lower CD4 cell counts among those adherent to ART is consistent with other observations suggesting a direct effect of HCV on the immune system (Bare et al., 2003, 2005; El-Serag et al., 2002; Pawlotsky et al., 1995). Recent data have shown that HCV proteins and HCV infection of mononuclear cells are associated with enhanced lymphocyte apoptosis (Nunez et al., 2006; Thoren et al., 2004). It is unclear why we observed an effect of HCV only in those individuals receiving ART. However it is possible that exposure to antiretroviral medications enhances the oxidative or apoptotic "stress" within these cells. It is also possible that HCV induces alterations in CD4 cells through an alternative but poorly defined mechanism as HCV infection is associated with a number of lymphoproliferative and autoimmune conditions including cryoglobulinemia, lymphoma, and autoimmune phenomena.

Whether HCV infection has an effect on mortality in HIV-infected patients is controversial. Several studies have shown increased mortality in those with HCV coinfection (Anderson et al., 2004; Backus et al., 2005; Braitstein et al., 2005; Weis et al., 2006), while others have observed similar rates (El-Serag et al., 2005). Although we did not examine the effect of HCV on this clinical endpoint, a major strength of this study was the ability to control for factors such as presence of detectable HCV RNA, active drug and alcohol use, medication adherence and social factors such as homelessness. These factors were not addressed in several larger cohort studies (Backus et al., 2005; De Luca et al., 2002; El-Serag et al., 2005; Jaggy et al., 2003; Keiser et al., 2004; Law et al., 2004; Quintana et al., 2003; Rockstroh et al., 2005).

Although alcohol use may worsen HCV-related liver disease and interfere with medical care, we did not find evidence in our data to suggest that alcohol use is a confounder of the association between HCV status and HIV disease progression. Our findings that HCV infection has an adverse effect on CD4 cell count and percent CD4 cells in patients adherent to ART persisted regardless of whether alcohol use was accounted for in the analyses.

This study's findings have important implications for understanding the optimal management approach for coinfected patients. Recent work has shown that effective treatment of HIV infection can significantly reduce the risk of HCV-related liver disease progression (Qurishi et al., 2003). However, our results suggest that HCV infection may adversely impact HIV-related immune reconstitution. If this is true, additional improvement in HIV outcomes may be achieved by effective treatment of HCV infection in patients who are adherent to ART. Corroboration of these findings would add effective HCV treatment to the multidimensional approach to address HIV infection and strengthen the argument to initiate HCV therapy early in the course of HIV disease. These findings may also support the earlier introduction of ART in HCV coinfected individuals to both offset the negative impact of HIV on HCV disease progression and the potentially negative impact of HCV on CD4 cell recovery.

The limitations of this study merit explanation. As HCV coinfection is more common in injection drug users, this group's delayed presentation for medical care, compared with noninjection drug users, may account for lower CD4 cell counts among coinfected individuals (Samet et al., 2001). However, adjustment for CD4 cell count at study enrollment in the multivariable analysis should mitigate this argument. Also, HCV-negative subjects were not diagnosed with HIV earlier and had not been on ART longer than HCV-positive subjects, suggesting our results are not explained by the groups being in care for different lengths of time. It is possible that we did not adequately control for all potential contributing factors to HIV disease progression. Nonetheless, multiple characteristics known to be associated with HIV disease progression were included as covariates in the multivariable analyses. This study may have been underpowered to detect a statistically significant effect of HCV infection. The observed differences in CD4 cell count and log HIV RNA between the HCV RNA-positive and negative subjects were smaller than anticipated, thus it is likely that the study was not adequately powered to detect associations of the observed magnitude.

In summary, in this cohort of subjects with alcohol problems, HCV infection is associated with lower CD4 cell counts in patients adherent to ART. Addressing HCV coinfection in HIV-infected patients who have maximized ART may confer additional immunologic benefit for this patient population.

ACKNOWLEDGMENTS

We appreciate the contributions of the research associates and data managers. Research was conducted in part in the General Clinical Research Centers at Boston University School of Medicine, USPHS Grant MO1 RR00533 and Beth Israel Deaconess Medical Center, USPHS Grant M01 RR01032.

REFERENCES

Alberti A, Clumeck N, Collins S, Gerlich W, Lundgren J, Palu G, Reiss P, Thiebaut R, Weiland O, Yazdanpanah Y, Zeuzem S (2005) Short statement of the first European Consensus Conference on the

treatment of chronic hepatitis B and C in HIV co-infected patients. J Hepatol 42:615–624.

Anderson KB, Guest JL, Rimland D (2004) Hepatitis C virus coinfection increases mortality in HIV-infected patients in the highly active antiretroviral therapy era: data from the HIV Atlanta VA Cohort Study. Clin Infect Dis 39:1507–1513.

Backus LI, Phillips BR, Boothroyd DB, Mole LA, Burgess J, Rigsby MO, Chang SW (2005) Effects of hepatitis C virus coinfection on survival in veterans with HIV treated with highly active antiretroviral therapy. J Acquir Immune Defic Syndr 39:613–619.

Bare P, Massud I, Belmonte L, Corti M, Villafane M, Perez Bianco R, de Tezanos-pinto M, de Bracco MM, Ruibal-Ares B (2003) HCV recovery from peripheral blood mononuclear cell culture supernatants derived from HCV-HIV co-infected haemophilic patients with undetectable HCV viraemia. Haemophilia 9:598–604.

Bare P, Massud I, Parodi C, Belmonte L, Garcia G, Nebel MC, Corti M, Pinto MT, Bianco RP, Bracco MM, Campos R, Ares BR (2005) Continuous release of hepatitis C virus (HCV) by peripheral blood mononuclear cells and B-lymphoblastoid cell-line cultures derived from HCV-infected patients. J Gen Virol 86:1717–1727.

Bonacini M, Puoti M (2000) Hepatitis C in patients with human immunodeficiency virus infection: diagnosis, natural history, meta-analysis of sexual and vertical transmission, and therapeutic issues. Arch Intern Med 160:3365–3373.

Braitstein P, Yip B, Montessori V, Moore D, Montaner JS, Hogg RS (2005) Effect of serostatus for hepatitis C virus on mortality among antiretrovirally naive HIV-positive patients. Can Med Assoc J 173: 160–164.

Braitstein P, Zala C, Yip B, Brinkhof MW, Moore D, Hogg RS, Montaner JS (2006) Immunologic response to antiretroviral therapy in hepatitis C Virus-coinfected adults in a population-based HIV/AIDS treatment program. J Infect Dis 193:259–268.

Buchsbaum DG, Buchanan RG, Centor RM, Schnoll SH, Lawton MJ (1991) Screening for alcohol abuse using CAGE scores and likelihood ratios. Ann Intern Med 115:774–777.

Chesney MA, Ickovics JR, Chambers DB, Gifford AL, Neidig J, Zwickl B, Wu AW (2000) 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 12:255–266.

Chung RT, Andersen J, Volberding P, Robbins GK, Liu T, Sherman KE, Peters MG, Koziel MJ, Bhan AK, Alston B, Colquhoun D, Nevin T, Harb G, van der Horst C (2004) Peginterferon Alfa-2a plus ribavirin versus interferon alfa-2a plus ribavirin for chronic hepatitis C in HIV-coinfected persons. N Engl J Med 351:451–459.

Chung RT, Evans SR, Yang Y, Theodore D, Valdez H, Clark R, Shikuma C, Nevin T, Sherman KE (2002) Immune recovery is associated with persistent rise in hepatitis C virus RNA, infrequent liver test flares, and is not impaired by hepatitis C virus in co-infected subjects. AIDS 16:1915–1923.

Cook RT, Stapleton JT, Ballas ZK, Klinzman D (1997) Effect of a single ethanol exposure on HIV replication in human lymphocytes. J Invest Med 45:265–271.

De Luca A, Bugarini R, Lepri AC, Puoti M, Girardi E, Antinori A, Poggio A, Pagano G, Tositti G, Cadeo G, Macor A, Toti M, D'Arminio Monforte A (2002) Coinfection with hepatitis viruses and outcome of initial antiretroviral regimens in previously naive HIV-infected subjects. Arch Intern Med 162:2125–2132.

Dorrucci M, Pezzotti P, Phillips AN, Lepri AC, Rezza G (1995) Coinfection of hepatitis C virus with human immunodeficiency virus and progression to AIDS. Italian Seroconversion Study. J Infect Dis 172:1503–1508.

El-Serag HB, Giordano TP, Kramer J, Richardson P, Souchek J (2005) Survival in hepatitis C and HIV co-infection: a cohort study of hospitalized veterans. Clin Gastroenterol Hepatol 3: 175–183.

El-Serag HB, Hampel H, Yeh C, Rabeneck L (2002) Extrahepatic manifestations of hepatitis C among United States male veterans. Hepatology 36:1439–1445.

Fleming CA, Craven DE, Thornton D, Tumilty S, Nunes D (2003) Hepatitis C virus and human immunodeficiency virus coinfection in an urban population: low eligibility for interferon treatment. Clin Infect Dis 36:97–100.

Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state" A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12:189–198.

Greub G, Ledergerber B, Battegay M, Grob P, Perrin L, Furrer H, Burgisser P, Erb P, Boggian K, Piffaretti JC, Hirschel B, Janin P, Francioli P, Flepp M, Telenti A (2000) Clinical progression, survival, and immune recovery during antiretroviral therapy in patients with HIV-1 and hepatitis C virus coinfection: the Swiss HIV cohort study. Lancet 356:1800–1805.

Jaggy C, von Overbeck J, Ledergerber B, Schwarz C, Egger M, Rickenbach M, Furrer HJ, Telenti A, Battegay M, Flepp M, Vernazza P, Bernasconi E, Hirschel B (2003) Mortality in the Swiss HIV cohort study (SHCS) and the Swiss general population. Lancet 362:877–878.

Kaufmann GR, Perrin L, Pantaleo G, Opravil M, Furrer H, Telenti A, Hirschel B, Ledergerber B, Vernazza P, Bernasconi E, Rickenbach M, Egger M, Battegay M (2003) CD4 T-lymphocyte recovery in individuals with advanced HIV-1 infection receiving potent antiretroviral therapy for 4 years: the Swiss HIV Cohort Study. Arch Intern Med 163:2187–2195.

Keiser O, Taffe P, Zwahlen M, Battegay M, Bernasconi E, Weber R, Rickenbach M (2004) All cause mortality in the Swiss HIV Cohort Study from 1990 to 2001 in comparison with the Swiss population. AIDS 18:1835–1843.

Kertesz SG, Larson MJ, Horton NJ, Winter M, Saitz R, Samet JH (2005) Homeless chronicity and health-related quality of life trajectories among adults with addictions. Med Care 43:574–585.

Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen H (1991) The world health organization composite international diagnostic interview short form (CIDI-SF). Int J Methods Psychiatr Res 7:171–185.

Klein MB, Lalonde RG, Suissa S (2003) The impact of hepatitis C virus coinfection on HIV progression before and after highly active antiretroviral therapy. J Acquir Immune Defic Syndr 33:365–372.

Kresina TF, Flexner CW, Sinclair J, Correia MA, Stapleton JT, Adeniyi-Jones S, Cargill V, Cheever LW (2002) Alcohol use and HIV pharmacotherapy. AIDS Res Hum Retrov 18:757–770.

Krupitsky EM, Horton NJ, Williams EC, Lioznov D, Kuznetsova M, Zvartau E, Samet JH (2005) Alcohol use and HIV risk behaviors among HIV-infected hospitalized patients in St. Petersburg, Russia. Drug Alcohol Depend 79:251–256.

Laird NM, Ware JH (1982) Random-effects models for longitudinal data. Biometrics 38:963–974.

Law WP, Duncombe CJ, Mahanontharit A, Boyd MA, Ruxrungtham K, Lange JM, Phanuphak P, Cooper DA, Dore GJ (2004) Impact of viral hepatitis co-infection on response to antiretroviral therapy and HIV disease progression in the HIV-NAT cohort. AIDS 18: 1169–1177.

Macias J, Pineda JA, Leal M, Abad MA, Garcia-Pesquera F, Delgado J, Gallardo JA, Sanchez-Quijano A, Lissen E (1998) Influence of hepatitis C virus infection on the mortality of antiretroviral-treated patients with HIV disease. Eur J Clin Microbiol Infect Dis 17:167–170.

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

National Institute on Alcohol Abuse and Alcoholism (1995) *The Physician's Guide to Helping Patients With Alcohol Problems*. National Institute on Alcohol Abuse and Alcoholism, Washington, DC.

Nunez M, Soriano V, Lopez M, Ballesteros C, Cascajero A, Gonzalez-Lahoz J, Benito JM (2006) Coinfection with hepatitis C virus increases lymphocyte apoptosis in HIV-infected patients. Clin Infect Dis 43:1209–1212.

Pachl C, Todd JA, Kern DG, Sheridan PJ, Fong SJ, Stempien M, Hoo B, Besemer D, Yeghiazarian T, Irvine B, Kolberg J, Kokka R, Neuwald P, Urdea MS (1995) Rapid and precise quantification of HIV-1 RNA in plasma using a branched DNA signal amplification assay. J Acq Immun Def Synd Hum Retrov 8:446–454.

Pawlotsky JM, Bouvier M, Fromont P, Deforges L, Duval J, Dhumeaux D, Bierling P (1995) Hepatitis C virus infection and autoimmune *Thrombocytopenic purpura*. J Hepatol 23:635–639.

Piroth L, Duong M, Quantin C, Abrahamowicz M, Michardiere R, Aho LS, Grappin M, Buisson M, Waldner A, Portier H, Chavanet P (1998) Does hepatitis C virus co-infection accelerate clinical and immunological evolution of HIV-infected patients? AIDS 12:381–388.

Quintana M, del Amo J, Barrasa A, Perez-Hoyos S, Ferreros I, Hernandez F, Villar A, Jimenez V, Bolumar F (2003) Progression of HIV infection and mortality by hepatitis C infection in patients with haemophilia over 20 years. Haemophilia 9:605–612.

Qurishi N, Kreuzberg C, Luchters G, Effenberger W, Kupfer B, Sauerbruch T, Rockstroh JK, Spengler U (2003) Effect of antiretroviral therapy on liver-related mortality in patients with HIV and hepatitis C virus coinfection. Lancet 362:1708–1713.

Rancinan C, Neau D, Saves M, Lawson-Ayayi S, Bonnet F, Mercie P, Dupon M, Couzigou P, Dabis F, Chene G (2002) Is hepatitis C virus co-infection associated with survival in HIV-infected patients treated by combination antiretroviral therapy? AIDS 16:1357–1362.

Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N, Towle LH (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.

Rockstroh JK, Mocroft A, Soriano V, Tural C, Losso MH, Horban A, Kirk O, Phillips A, Ledergerber B, Lundgren J (2005) Influence of hepatitis C virus infection on HIV-1 disease progression and response to highly active antiretroviral therapy. J Infect Dis 192:992–1002.

Samet JH, Freedberg KA, Savetsky JB, Sullivan LM, Stein MD (2001) Understanding delay to medical care for HIV infection: the long-term non-presenter. AIDS 15:77–85.

Samet JH, Horton NJ, Meli S, Freedberg KA, Palepu A (2004a) Alcohol consumption and antiretroviral adherence among HIV-infected persons with alcohol problems. Alcohol Clin Exp Res 28:572–577.

Samet JH, Libman H, LaBelle C, Steger K, Lewis R, Craven DE, Freedberg KA (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.

Samet JH, Phillips SJ, Horton NJ, Traphagen ET, Freedberg KA (2004b) Detecting alcohol problems in HIV-infected patients: use of the CAGE questionnaire. AIDS Res Hum Retroviruses 20:151–155.

Schiff ER (1999) The alcoholic patient with hepatitis C virus infection. Am J Med 107:95S–99S.

Smith KL, Horton NJ, Saitz R, Samet JH (2006) The use of the mini-mental state examination in recruitment for substance abuse research studies. Drug Alcohol Depend 82:231–237.

Sobell LC, Sobell MB (1992) Timeline follow-back: a technique for assessing self-reported alcohol consumption, in *Measuring Alcohol Consumption: Psychosocial and Biochemical Methods* (Litten RZ, Allen JP eds), pp 41–69. Humana Press, Totowa, NJ.

Soriano V, Puoti M, Bonacini M, Brook G, Cargnel A, Rockstroh J, Thio C, Benhamou Y (2005) Care of patients with chronic hepatitis B and HIV co-infection: recommendations from an HIV-HBV International Panel. AIDS 19:221–240.

Staples CT Jr., Rimland D, Dudas D (1999) Hepatitis C in the HIV (human immunodeficiency virus) Atlanta V.A. (Veterans Affairs Medical Center) Cohort Study (HAVACS): the effect of coinfection on survival. Clin Infect Dis 29:150–154.

Sulkowski MS, Moore RD, Mehta SH, Chaisson RE, Thomas DL (2002) Hepatitis C and progression of HIV disease. JAMA 288: 199–206.

Sulkowski MS, Thomas DL (2003) Hepatitis C in the HIV-infected person. Ann Intern Med 138:197–207.

Sulkowski MS, Thomas DL, Chaisson RE, Moore RD (2000) Hepatotoxicity associated with antiretroviral therapy in adults infected with human immunodeficiency virus and the role of hepatitis C or B virus infection. JAMA 283:74–80.

Thio CL, Nolt KR, Astemborski J, Vlahov D, Nelson KE, Thomas DL (2000) Screening for hepatitis C virus in human immunodeficiency virus-infected individuals. J Clin Microbiol 38:575–577.

Thoren F, Romero A, Lindh M, Dahlgren C, Hellstrand K (2004) A hepatitis C virus-encoded, nonstructural protein (NS3) triggers dysfunction and apoptosis in lymphocytes: role of NADPH oxidase-derived oxygen radicals. J Leukoc Biol 76:1180–1186.

Torriani FJ, Rodriguez-Torres M, Rockstroh JK, Lissen E, Gonzalez-Garcia J, Lazzarin A, Carosi G, Sasadeusz J, Katlama C, Montaner J, Sette H Jr., Passe S, De Pamphilis J, Duff F, Schrenk UM, Dieterich DT (2004) Peginterferon Alfa-2a plus ribavirin for chronic hepatitis C virus infection in HIV-infected patients. N Engl J Med 351:438–450.

Weis N, Lindhardt BO, Kronborg G, Hansen AB, Laursen AL, Christensen PB, Nielsen H, Moller A, Sorensen HT, Obel N (2006) Impact of hepatitis C virus coinfection on response to highly active antiretroviral therapy and outcome in HIV-infected individuals: a nationwide cohort study. Clin Infect Dis 42:1481–1487.

Zurlo JJ, Wood L, Gaglione MM, Polis MA (1995) Effect of splenectomy on T lymphocyte subsets in patients infected with the human immunodeficiency virus. Clin Infect Dis 20:768–771.

Violence and Victims, Volume 22, Number 2, 2007

Substance Use During Sexual and Physical Assault in HIV-Infected Persons

Cynthia H. Chuang, MD, MSc
Penn State College of Medicine and Milton S. Hershey Medical Center
Hershey, PA

Jane M. Liebschutz, MD, MPH
Boston Medical Center, MA

Debbie M. Cheng, ScD
Boston Medical Center and Boston University, MA

Anita Raj, PhD, MS
Boston University, MA

Jeffrey H. Samet, MD, MA, MPH
Boston Medical Center and Boston University, MA

Data from the HIV-Alcohol Longitudinal Cohort was used to determine the prevalence of substance use by victims and assailants during physical and sexual assault against HIV-infected persons and whether these findings differed by gender. Of the sexually assaulted participants, 31% of victims and 70% of assailant(s) had used drugs/alcohol during sexual assault. Compared with men, women had higher odds of substance use during sexual assault (adjusted odds ratio [OR] 3.8, 95% confidence interval [CI] 1.6 to 8.7) and of substance use by their assailant(s) during sexual assault (adjusted OR 5.9, 95% CI 1.7 to 20.6) in adjusted analysis. Of the physically assaulted participants, 66% of victims and 85% of assailants used drugs/alcohol during physical assault; these results did not differ by gender.

Keywords: alcohol use; drug use; violence; victims; abuse

The association of substance use with violence has been well established. High rates of chronic drug and alcohol abuse have been described among both assailants and victims of violent crimes, sexual assault, and intimate partner violence (Abbey, Zawacki, Buck, Clinton, & McAuslan, 2001; Chermack & Blow, 2002; Cunradi, Caetano, Clark, & Schafer, 1999; Cunradi, Caetano, & Schafer, 2002; Kilpatrick, Acierno, Resnick, Saunders, & Best, 1997; Lown & Vega, 2001; Rivara et al., 1997; Sharps, Campbell, Campbell, Gary, & Webster, 2001; Zavala & French, 2003). Other work has also shown that assailants and victims of sexual and physical violence are often substance using at the time of the event (Caetano, Schafer, & Cunradi, 2001; Chermack & Blow, 2002; Kantor

& Straus, 1989; Mohler-Kuo, Dowdall, Koss, & Wechsler, 2004; Thompson & Kingree, 2004). In fact, one review estimated that 45% of intimate partner violence perpetrated by men involved alcohol (Roizen, 1993). However, the role of substance use during violence experienced by HIV-infected groups and how it is affected by gender of the victim has not been well described.

Violence among those with HIV and at risk for HIV may be even greater than among the general population (Bedimo, Kissinger, & Bessinger, 1997; Cohen et al., 2000; McDonnell, Gielen, & O'Campo, 2003; Vlahov et al., 1998; Zierler et al., 2000). Sixty-six percent of HIV-infected women in one cohort reported a history of assault, with 21% having been victimized in the past year and 31% reporting a history of sexual assault before age 18 (Cohen et al., 2000). Another study of HIV-infected women found that 32% had a history of sexual assault during their lifetime (Bedimo et al., 1997). Although less is known about the prevalence of abuse among HIV-infected men, it appears that they are commonly victims of violence as well. Data from the HIV Costs and Service Utilization Study revealed that 20.5% of women, 11.5% of men having sex with men, and 7.5% of heterosexual men reported physical harm since HIV diagnosis, nearly half of whom reported HIV as the cause of violence (Zierler et al., 2000). This is highly concerning given the association that victimization has with several negative health measures in cross-sectional studies of HIV-infected persons. For example, victimization has been associated with increased medical disease, health care utilization, and substance use characteristics (Eisenman, Cunningham, Zierler, Nakazono, & Shapiro, 2003; Liebschutz, Feinman, Sullivan, Stein, & Samet, 2000). Additionally, sexual victimization has been associated with high-risk sexual behavior in HIV-infected persons (Bogart et al., 2005; Chuang, Liebschutz, Horton, & Samet, 2006). These findings further emphasize the need to better characterize violence against HIV-infected persons.

The pathway through which substance use contributes to violence is likely multifold and perhaps different from one situation to another. For example, some assailants may have the desire to commit sexual assault, and that desire also leads them to drink alcohol. Alcohol use could then be a consequence rather than a cause for the assault. Conversely, alcohol use may actually be the facilitator to violence, which is supported by the hypothesized proximal role of substance use in violent acts (Fals-Stewart, Golden, & Schumacher, 2003). In these situations, alcohol may affect cognitive ability and reduce inhibitions that ultimately lead to assault. Another contributing factor may be the social situations in which violence occurs; for example, certain situations may encourage both drinking and sexual assault, such as bars or parties (Abbey et al., 2001).

Since HIV-infected groups may be more at risk for both chronic substance use (Bing et al., 2001; Cook et al., 2001; Galvan et al., 2002; Samet, Phillips, Horton, Traphagen, & Freedberg, 2004) and violence victimization (Bedimo et al., 1997; Cohen et al., 2000; McDonnell et al., 2003), we hypothesize that they may be particularly susceptible to substance-involved acts of violence during their lifetimes. Little is known about the role of acute intoxication during violence experienced in the lives of HIV-infected individuals. One study found that victims of violence who were HIV-infected women were more likely to report using alcohol before or during a violent episode than HIV-negative women (McDonnell et al., 2003). Research in this area has focused on female victims of substance-involved assault, but men are likely at risk as well. Although women are more likely to be victims of assault in the context of intimate relationships, men are overall more often the victims of physical violence (Craven, 1997). Interestingly one study found that incident-specific alcohol use by perpetrators was more common during intimate partner

violence against women than against men (Thompson & Kingree, 2004). Conversely, another study found that men were more likely than women to have been drinking during the violent episode, independent of whether the man was the perpetrator or the victim (Caetano et al., 2001). The present study uses the baseline cross-sectional data from the HIV-Alcohol Longitudinal Cohort (HIV-ALC) study, a cohort of HIV-infected persons with a lifetime history of alcohol problems. In this cohort, 81% of the participants had experienced physical or sexual violence in their lifetime, and 40% of the participants experienced physical or sexual violence at some time during the 3-year follow-up period (Liebschutz, Geier, Horton, Chuang, & Samet, 2005). Our objective was to describe the prevalence of substance use by both assailants and victims during lifetime violence using data collected at the baseline time point and to observe whether the results differed by gender of the victim.

MATERIALS AND METHODS

Study Design and Participant Recruitment

The HIV-ALC study recruited HIV-infected individuals with a history of alcohol problems with the primary aim of evaluating the effect of alcohol use on HIV progression (Samet, Horton, Traphagen, Lyon, & Freedberg, 2003). The current study is a cross-sectional analysis using the baseline data of subsamples of this cohort who reported experiencing either sexual or physical assault. The Institutional Review Boards of Boston Medical Center and Beth Israel Deaconess Medical Center approved this study. Patients were recruited principally from the Boston Medical Center HIV Diagnostic Evaluation Unit (Samet et al., 1995), a weekly clinic for engaging HIV-infected persons into medical care. Participants were also recruited from other sites: the Beth Israel Deaconess Medical Center, a respite facility for homeless persons, a methadone clinic, Boston Medical Center's primary care practices, referrals by friends, and through posted flyers at homeless shelters and HIV/AIDS social service agencies in the Boston area.

Eligibility criteria were confirmed HIV infection, a lifetime history of alcohol problems (defined as two or more positive responses to the CAGE questionnaire [Ewing, 1984]), and age ≥18 years. Those patients recruited from the Boston Medical Center HIV Diagnostic Evaluation Unit who did not meet CAGE criteria were eligible if one of two attending physicians made a specific diagnosis of alcohol abuse or dependence. Other inclusion criteria were fluency in English or Spanish, a Mini-Mental State Examination score ≥21 (Folstein, Folstein, & McHugh, 1975), and no plans to move residence from the Boston area for the next 2 years. Recruitment ran from June 1997 to July 2001. Informed consent was obtained from subjects who met the eligibility criteria and agreed to participate in the study.

A trained research associate administered structured in-person confidential interviews in English or Spanish. For the measures used in this analysis, the Spanish items were created by translating from English, then back translating to check for accuracy, and corrected. Most interviews were performed at the General Clinical Research Center of the Boston University School of Medicine.

Definition of Variables

Data collected included demographics (age, sex, race/ethnicity, education, marital status, homelessness), interpersonal violence exposure, alcohol and drug use, and HIV

transmission category. Homelessness was defined as at least one night in a shelter or on the street in the past 6 months. Alcohol use in the past 30 days was categorized as hazardous, moderate, or abstinent. These categories were derived from the Department of Health and Human Services (2000) definition for hazardous use (> 14 drinks per week for men < 66 years old and > 7 drinks per week for men ≥ 66 years old and all women or ≥ 5 drinks on one occasion for men < 66 years old and ≥ 4 drinks on one occasion for men ≥ 66 years old and all women). Moderate alcohol use was defined as any drinking less than hazardous.

The questions used to assess violence exposure were adapted from a previous study designed to describe interpersonal violence among persons with a history of substance abuse (Liebschutz et al., 2002). To assess for sexual violence, the subjects were asked, "Have you ever been sexually assaulted (for example: unwanted sexual touching anywhere on your body, touching of genitals and/or breasts, or made to have oral sex or vaginal or anal intercourse against your will by force or the threat of force)?" If an assault history was reported, the participants were asked whether they were using alcohol or drugs when they were assaulted. Response categories were "never," "some cases," "most cases," or "all cases." The outcome variable created to represent substance use during assault was dichotomous (never vs. at least some cases). Participants were then asked whether the person who assaulted them was using alcohol or drugs when they were assaulted, using the same response categories and outcome variable. To assess for a history of physical assault, the participants were asked, "Have you ever been physically abused or assaulted (for example: kicked, hit, choked, shot, stabbed, burned, or held at gunpoint)?" If a history of physical assault was reported, subjects were asked whether they or their assailants were using drugs or alcohol at the time of physical assault, in the same manner as for sexual assault. For each type of assault reported (sexual or physical), participants were asked their age at first victimization and whether the assailant(s) were stranger(s), nonstranger(s), or both.

Statistical Analysis

Statistical analysis was performed using SAS/STAT v.8.2 (SAS Institute, 2001). The data analysis for this study was restricted to the subgroups of participants who reported a history of sexual or physical assault. Descriptive data for demographic and substance use characteristics are presented by assault history and gender. The proportion of assault victims and assailants with a history of substance use during sexual and physical assault is reported. The odds of substance use by the victim and the assailant during both sexual and physical assault was then compared by gender using logistic regression modeling, adjusting for age (continuous), race (White vs. non-White), and the victim–assailant relationship (stranger, nonstranger, or both). For all analyses, two-tailed tests were performed using $p < .05$ as criterion for statistical significance.

Childhood sexual assault and childhood physical assault (occurring before age 13) were considered as potential confounders in the analyses of substance abuse during assault but were excluded from the final logistic regression models because of potential collinearity with victim–assailant relationship (i.e., participants reporting childhood assault were significantly more likely to report nonstranger assailants than those not reporting childhood assault). In subsequent analysis, the childhood assault variables were assessed as potential confounders in regression models that did not include the victim–assailant relationship variable. In this analysis the primary relationship of gender with substance use during assault was not confounded by a history of childhood assault.

RESULTS

The HIV-ALC cohort ($N = 349$) was recruited from the following locations: 56% from the Boston Medical Center HIV Diagnostic Evaluation Unit, 16% from posted flyers, 13% from Boston Medical Center's primary care practices, 5% from a respite facility for homeless persons, 4% from a methadone clinic, 4% from friend referrals, and 2% from the Beth Israel Deaconess Medical Center. Most study subjects (313/349 [90%]) met the eligibility criteria of two or more positive responses to the CAGE questionnaire (Ewing, 1984); the remainder qualified on the basis of clinical assessment (36/349 [10%]). The majority of the participants completed the interviews in English, while 26 (3%) completed the interview in Spanish.

Participant Characteristics

Eighty-one percent of the cohort (281/349) had experienced either sexual and/or physical assault in their lifetime, with 139 (40%) reporting a history of sexual assault and 266 (76%) reporting a history of physical assault. The sexual and physical assault groups were not mutually exclusive since more than one-third of the participants experienced both sexual and physical assault (36%). All subsequent analysis is restricted to subjects reporting some type of assault. Characteristics of the participants who experienced either sexual assault or physical assault with comparisons by gender are shown in Table 1. Among those reporting a lifetime history of sexual assault, the men were more likely to have used alcohol in the past 30 days but less likely to have used heroin in the past 30 days than the women. Men reporting a lifetime history of physical assault were older, more likely to be homeless, more likely to have consumed alcohol in the past 30 days, but less likely to have used heroin in the past 30 days than women. Childhood assault was very common, with 63% of sexual assault victims and 47% of physical assault victims reporting first assault before age 13.

Substance Use During Sexual Assault

Data concerning substance use during sexual assault is shown in Table 2. Of the 139 subjects who had been sexually assaulted, 31% (95% CI 23% to 39%) reported they used drugs or alcohol during at least some cases of sexual assault, and 70% (95% CI 64% to 76%) reported their assailant had used drugs or alcohol during at least some cases of sexual assault. Gender differences were clinically and statistically significant. Female sexual assault victims were significantly more likely than male sexual assault victims to report substance use during sexual assault (51% vs. 19%, adjusted OR 3.8, 95% CI 1.6 to 8.7), adjusting for age, race/ethnicity, and victim–assailant relationship. Substance use by the assailant during sexual assault was also reported significantly more often by female victims (88% vs. 58%, adjusted OR 5.9, 95% CI 1.7 to 20.6) in the adjusted analysis.

Substance Use During Physical Assault

Data concerning substance use during physical assault are also shown in Table 2. Of the 266 subjects who had been physically assaulted, 66% (95% CI 60% to 72%) reported they had used drugs or alcohol during at least some cases of physical assault, and 85% (95% CI 81% to 89%) reported their assailant had used drugs or alcohol during at least some cases of physical assault. The odds of substance use by either the subject or the assailant were similar for men and women.

TABLE 1. Characteristics of the HIV-Alcohol Longitudinal Cohort Participants With Lifetime Histories of Sexual and Physical Victimization

	Victims of Sexual Assault[a] (n = 139)		Victims of Physical Assault[a] (n = 266)	
Characteristic	Men N = 86	Women N = 53	Men N = 206	Women N = 60
Age—mean in years (*SD*)	39 (7)	38 (7)	41 (7)*	38 (7)*
High school education or more—no. (%)	53 (62)	33 (62)	121 (59)	38 (63)
Married—no. (%)	7 (8)	6 (11)	14 (7)	8 (13)
Homeless—no. (%)	22 (26)	10 (19)	65 (32)*	8 (13)*
Race—no. (%)				
Black	33 (38)	26 (49)	80 (39)	31 (52)
White	34 (40)	17 (32)	78 (38)	18 (30)
Hispanic	18 (21)	9 (17)	47 (23)	10 (17)
Other	1 (1)	1 (2)	1 (0.5)	1 (2)
HIV transmission risk category**—no. (%)				
Heterosexual	8 (9)	22 (42)	30 (15)	27 (45)
Injection drug use	41 (48)	31 (58)	127 (62)	33 (55)
Men who have sex with men (MSM)	37 (43)	N/A	49 (24)	N/A
Alcohol use in past 30 days—no. (%)				
Abstinent	50 (58)*	41 (77)*	116 (56)*	44 (73)*
Moderate	11 (13)*	1 (2)*	24 (12)*	1 (2)*
Hazardous	25 (29)*	11 (21)*	66 (32)*	15 (25)*
Any cocaine use in past 30 days—no. (%)	25 (29)	10 (19)	51 (25)	10 (17)
Any heroin use in past 30 days—no. (%)	2 (2)*	8 (15)*	14 (7)*	10 (17)*
First assault occurring prior to age 13	60 (70)	28 (54)	98 (49)	26 (46)

[a]Sexual assault and physical assault groups are not mutually exclusive.
*$p < .05$ comparing men and women.
**HIV transmission risk category could not be compared by gender because MSM are exclusively men.

DISCUSSION

Substance use during assault by both HIV-infected victims and their assailants was high in this cohort of HIV-infected persons with a history of alcohol problems. In fact, the majority of the assault victims reported that their assailants were substance using during both physical and sexual assault. Similarly, a large number of participants reported that they themselves were substance using when they were victimized, particularly during physical assault.

Compared with men, women experiencing sexual assault were significantly more likely to report substance use by both themselves and their assailants. The reasons for this are unclear. We initially hypothesized that men who experienced sexual assault may be more likely to have been victimized as children exclusively, when they would be less likely to be substance using. However, in a separate analysis looking only at the participants who experienced sexual assault after childhood (age ≥13), we found that women victims were still significantly more likely to report substance use during sexual assault

TABLE 2. Frequency of Substance Use by Victims and Assailants During Sexual and Physical Assault Against HIV-Infected Persons and Logistic Regression Analysis Results of the Association Between Substance Use and Gender

	Total	Men	Women	Adjusted OR[a] (95% CI)
Substance use during sexual assault (N = 139):				
By victim	31%	19%	51%	3.79 (1.64–8.72)
By assailant	70%	58%	88%	5.88 (1.68–20.58)
Substance use during physical assault (N = 266):				
By victim	66%	66%	65%	1.07 (0.55–2.07)
By assailant	85%	83%	89%	1.80 (0.59–5.50)

[a]Odds ratio of substance use during assault for women compared with men, controlling for age, race, and victim–assailant relationship.

than men victims (63% vs. 36%, adjusted OR 4.22, 95% CI 1.05 to 16.95). Women were also still more likely than men to report substance use by their assailants among those only experiencing sexual assault after childhood (88% vs. 59%, unadjusted OR 5.25, 95% CI 0.90 to 30.62) compared with the full sample (88% vs. 58%, unadjusted OR 5.39, 95% CI 1.67 to 17.42); this difference was no longer statistically significant, likely because of the smaller sample size. Note that the adjusted OR for the full sample was 5.88 (95% CI 1.68 to 20.58), but we were unable to run an adjusted model for those experience sexual assault after childhood because of a limited number of events. An alternative explanation to the gender findings is the possibility that the social situations around which sexual assault occurs differ for men and women. Women may be more likely fall victim to sexual assault when they are in social situations that encourage both drinking and sexual pursuit, like bars or parties (Abbey et al., 2001), and perhaps the environments for victimization of men differ. Additionally, whether differences exist in the social situations and environments that lead to intimate partner violence against men who have sex with men versus heterosexual men is unclear from our study. Further investigation of these issues is needed to elucidate the relationship of substance use and violence against HIV-infected men.

We were interested in whether a history of childhood assault would bias the study findings. Since childhood sexual abuse victims are more susceptible to both increased substance use and violence victimization in adulthood (Bartholow et al., 1994; DiIorio, Hartwell, & Hansen, 2002; Johnsen & Harlow, 1996; Wingood & DiClemente, 1997), we hypothesized that they would be more susceptible to substance involved violence experienced in adulthood. However, the limited data collected on childhood assault did not allow us to fully address this hypothesis; whether a participant had experienced assault during childhood only or in adulthood as well could not be distinguished. In this sample, childhood sexual assault was found to be negatively associated with substance use by the victim during sexual assault. We suspect that this finding was driven by participants who experienced sexual assault during childhood only, when substance use by the victim would be less likely. As described previously, collinearity between childhood assault and the victim–assailant relationship variable did not allow us to assess the effects of both variables

simultaneously and led to exclusion of the childhood assault variable from the final analysis. However, the primary relationship of gender with substance use during assault was not modified when the childhood assault variable was assessed in regression models that did not include the victim–assailant relationship variable.

This study has certain limitations. There is potential for recall and reporting bias with self-report of sensitive data that reflects a lifetime experience. The topics being recalled, however, concerned very important life events, which are not likely to be forgotten. In addition, the interviews were performed confidentially by staff trained to facilitate patient comfort in order to minimize the potential for inaccurate reporting. Since only the victims were interviewed, we relied on the victims' perceptions of whether their assailants were substance using during the episodes. Another limitation is the limited data collected in the survey instrument on victimization. The terminology of the questions and the limited questions on assault may have resulted in underreporting. However, unless the degree of underreporting differed by gender, it would not bias estimates of the association between gender and substance use during assault. An additional limitation of the data was that detail was not obtained regarding recurrent episodes of violence throughout the lifetime, which could involve different assailants and different relationships. Future exploration about the social contexts where substance involved violence occurs will help clarify these issues. Additionally, since lifetime experience of violence was measured, it is not possible to distinguish whether violence occurred before or after HIV infection.

HIV-infected persons have high prevalence of lifetime violence, which has been shown to negatively impact their physical and mental health status. This study demonstrated that substance use by both the victim and the assailant during cases of physical and sexual assault is very common in this HIV-infected sample. Furthermore, women were at even higher risk than men for substance-involved sexual assault. The reasons for these differences need to be further investigated. Whether these episodes of violence are occurring before or after HIV diagnosis is also deserving of further study. Such information can help guide the incorporation of violence prevention into substance abuse treatment programs aimed at HIV-infected individuals. Education about the risks of violence, techniques to avoid situations that could potentially lead to violence, as well as screening for violence and its sequelae are important issues in the substance abuse management of HIV-infected individuals.

REFERENCES

Abbey, A., Zawacki, T., Buck, P. O., Clinton, A. M., & McAuslan, P. (2001). Alcohol and sexual assault. *Alcohol Research and Health, 25,* 43–51.

Bartholow, B. N., Doll, L. S., Joy, D., Douglas, J. M., Jr., Bolan, G., Harrison, J. S., et al. (1994). Emotional, behavioral, and HIV risks associated with sexual abuse among adult homosexual and bisexual men. *Child Abuse and Neglect, 18,* 747–761.

Bedimo, A. L., Kissinger, P., & Bessinger, R. (1997). History of sexual abuse among HIV-infected women. *International Journal of STD and AIDS, 8,* 332–335.

Bing, E. G., Burnam, M. A., Longshore, D., Fleishman, J. A., Sherbourne, C. D., London, A. S., et al. (2001). Psychiatric disorders and drug use among human immunodeficiency virus-infected adults in the United States. *Archives of General Psychiatry, 58,* 721–728.

Bogart, L. M., Collins, R. L., Cunningham, W., Beckman, R., Golinelli, D., Eisenman, D., et al. (2005). The association of partner abuse with risky sexual behaviors among women and men with HIV/AIDS. *AIDS and Behavior, 9,* 325–333.

Caetano, R., Schafer, J., & Cunradi, C. B. (2001). Alcohol-related intimate partner violence among white, black, and Hispanic couples in the United States. *Alcohol Research and Health, 25,* 58–65.

Chermack, S. T., & Blow, F. C. (2002). Violence among individuals in substance abuse treatment: The role of alcohol and cocaine consumption. *Drug and Alcohol Dependence, 66,* 29–37.

Chuang, C. H., Liebschutz, J. M., Horton, N. J., & Samet, J. H. (2006). Association of violence victimization with inconsistent condom use in HIV-infected persons. *AIDS and Behavior, 10,* 201–207.

Cohen, M., Deamant, C., Barkan, S., Richardson, J., Young, M., Holman, S., et al. (2000). Domestic violence and childhood sexual abuse in HIV-infected women and women at risk for HIV. *American Journal of Public Health, 90,* 560–565.

Cook, R. L., Sereika, S. M., Hunt, S. C., Woodward, W. C., Erlen, J. A., & Conigliaro, J. (2001). Problem drinking and medication adherence among persons with HIV infection. *Journal of General Internal Medicine, 16,* 83–88.

Craven, D. (1997). *Special report: Sex differences in violent victimization* (NCJ-164508). Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.

Cunradi, C. B., Caetano, R., Clark, C. L., & Schafer, J. (1999). Alcohol-related problems and intimate partner violence among white, black, and Hispanic couples in the U.S. *Alcoholism, Clinical and Experimental Research, 23,* 1492–1501.

Cunradi, C. B., Caetano, R., & Schafer, J. (2002). Alcohol-related problems, drug use, and male intimate partner violence severity among US couples. *Alcoholism, Clinical and Experimental Research, 26,* 493–500.

Dilorio, C., Hartwell, T., & Hansen, N. (2002). Childhood sexual abuse and risk behaviors among men at high risk for HIV infection. *American Journal of Public Health, 92,* 214–219.

Eisenman, D. P., Cunningham, W. E., Zierler, S., Nakazono, T. T., & Shapiro, M. F. (2003). Effect of violence on utilization of services and access to care in persons with HIV. *Journal of General Internal Medicine, 18,* 125–127.

Ewing, J. A. (1984). Detecting alcoholism. The CAGE questionnaire. *Journal of the American Medical Association, 252,* 1905–1907.

Fals-Stewart, W., Golden, J., & Schumacher, J. A. (2003). Intimate partner violence and substance use: A longitudinal day-to-day examination. *Addictive Behaviors, 28,* 1555–1574.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Galvan, F. H., Bing, E. G., Fleishman, J. A., London, A. S., Caetano, R., Burnam, M. A., et al. (2002). The prevalence of alcohol consumption and heavy drinking among people with HIV in the United States: Results from the HIV Cost and Services Utilization Study. *Journal of Studies on Alcohol, 63,* 179–186.

Johnsen, L. W., & Harlow, L. L. (1996). Childhood sexual abuse linked with adult substance use, victimization, and AIDS-risk. *AIDS Education and Prevention, 8,* 44–57.

Kantor, G. K., & Straus, M. A. (1989). Substance abuse as a precipitant of wife abuse victimizations. *American Journal of Drug and Alcohol Abuse, 15,* 173–189.

Kilpatrick, D. G., Acierno, R., Resnick, H. S., Saunders, B. E., & Best, C. L. (1997). A 2-year longitudinal analysis of the relationships between violent assault and substance use in women. *Journal of Consulting and Clinical Psychology, 65,* 834–847.

Liebschutz, J., Savetsky, J. B., Saitz, R., Horton, N. J., Lloyd-Travaglini, C., & Samet, J. H. (2002). The relationship between sexual and physical abuse and substance abuse consequences. *Journal of Substance Abuse Treatment, 22,* 121–128.

Liebschutz, J. M., Feinman, G., Sullivan, L., Stein, M., & Samet, J. (2000). Physical and sexual abuse in women infected with the human immunodeficiency virus: Increased illness and health care utilization. *Archives of Internal Medicine, 160,* 1659–1664.

Liebschutz, J. M., Geier, J. L., Horton, N. J., Chuang, C. H., & Samet, J. H. (2005). Physical and sexual violence and health care utilization in HIV-infected persons with alcohol problems. *AIDS Care, 17,* 566–578.

Lown, A. E., & Vega, W. A. (2001). Alcohol abuse or dependence among Mexican American women who report violence. *Alcoholism, Clinical and Experimental Research, 25,* 1479–1486.

McDonnell, K. A., Gielen, A. C., & O'Campo, P. (2003). Does HIV status make a difference in the experience of lifetime abuse? Descriptions of lifetime abuse and its context among low-income urban women. *Journal of Urban Health, 80,* 494–509.

Mohler-Kuo, M., Dowdall, G. W., Koss, M. P., & Wechsler, H. (2004). Correlates of rape while intoxicated in a national sample of college women. *Journal of Studies on Alcohol, 65,* 37–45.

Rivara, F. P., Mueller, B. A., Somes, G., Mendoza, C. T., Rushforth, N. B., & Kellermann, A. L. (1997). Alcohol and illicit drug abuse and the risk of violent death in the home. *Journal of the American Medical Association, 278,* 569–575.

Roizen, J. (1993). Issues in the epidemiology of alcohol and violence. In S. E. Martin (Ed.), *Alcohol and interpersonal violence: Fostering multidisciplinary perspectives* (pp. 3–36). Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Samet, J. H., Horton, N. J., Traphagen, E. T., Lyon, S. M., & Freedberg, K. A. (2003). Alcohol consumption and HIV disease progression: Are they related? *Alcoholism, Clinical and Experimental Research, 5,* 862–867.

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. *Archives of Internal Medicine, 155,* 1629–1633.

Samet, J. H., Phillips, S. J., Horton, N. J., Traphagen, E. T., & Freedberg, K. A. (2004). Detecting alcohol problems in HIV-infected patients: Use of the CAGE questionnaire. *AIDS Research and Human Retroviruses, 20,* 151–155.

SAS Institute Inc. (2001). SAS/STAT v.8.2. Cary, NC: Author.

Sharps, P. W., Campbell, J., Campbell, D., Gary, F., & Webster, D. (2001). The role of alcohol use in intimate partner femicide. *American Journal on Addictions, 10,* 122–135.

Thompson, M. P., & Kingree, J. B. (2004). The role of alcohol use in intimate partner violence and non-intimate partner violence. *Violence and Victims, 19,* 63–74.

U.S. Department of Health and Human Services and U.S. Department of Agriculture (USDA). (2000). *Nutrition and your health: Dietary guidelines for Americans* (5th ed.). Home and Garden Bulletin No. 232. Washington, DC: USDA.

Vlahov, D., Wientge, D., Moore, J., Flynn, C., Schuman, P., Schoenbaum, E., et al. (1998). Violence among women with or at risk for HIV infection. *AIDS and Behavior, 2,* 53–60.

Wingood, G. M., & DiClemente, R. J. (1997). Child sexual abuse, HIV sexual risk, and gender relations of African-American women. *American Journal of Preventive Medicine, 13,* 380–384.

Zavala, S. K., & French, M. T. (2003). Dangerous to your health: The role of chronic drug use in serious injuries and trauma. *Medical Care, 41,* 309–322.

Zierler, S., Cunningham, W. E., Andersen, R., Shapiro, M. F., Nakazono, T., Morton, S., et al. (2000). Violence victimization after HIV infection in a US probability sample of adult patients in primary care. *American Journal of Public Health, 90,* 208–215.

Acknowledgments. Data management was provided by DM-STAT, Inc., Medford, Massachusetts, in particular Nicole Tibbets, Karen Traister, and Fay Robinson. The authors appreciate the contributions of the clinical staff of the HIV Diagnostic Evaluation Unit at Boston Medical Center, including Colleen LaBelle, RN, and Jennifer Doyle. We also acknowledge the invaluable work provided by the research associates and the project manager, Seville Meli, MPH. Finally, the authors appreciate the contributions of Howard Libman, MD, at the Beth Israel Deaconess Medical Center. Support for this study came from the following grants from the National Institute on Alcohol Abuse and Alcoholism (NIAAA): RO1-AA13766 (Clinical Impact of HCV and Alcohol in HIV-Infected Persons); RO1-AA11785 (Medication Adherence in Alcohol Abusing HIV Patients); RO1-AA10870 (Enhanced Linkage of Alcohol Abusers to Primary Care). This research was conducted in part in the General Clinical Research Center at Boston University School of Medicine, USPHS grant M01 RR00533.

Correspondence regarding this article should be directed to Cynthia H. Chuang, MD, MSc, Division of General Internal Medicine, HU15, Penn State Milton S. Hershey Medical Center, 500 University Drive, P.O. Box 850, Hershey, PA 17033-0850. E-mail: cchuang@psu.edu

Cancer Screening Participation: Comparative Willingness of San Juan Puerto Ricans versus New York City Puerto Ricans

Cristina Claudio, PhD; Ralph V. Katz, DMD, MPH, PhD; B. Lee Green, PhD; Nancy R. Kressin, PhD; Min Qi Wang, PhD; and Stefanie L. Russell, DDS, MPH, PhD

Financial support: The Cancer Screening Questionnaire Study was supported by National Institute of Dental and Craniofacial Research/National Institutes of Health grant U54 DE 14257, the New York University Oral Cancer Research on Adolescent and Adult Health Promotion Center, an Oral Health Disparities Research Center.

Objectives: The specific aim of this study was to determine the self-reported likelihood of New York Puerto Ricans (NYPR) and San Juan Puerto Ricans (SJPR) to participate in: 10 site-specific cancer screenings, cancer-screenings conducted by different specific persons/agencies and cancer-screening under specific conditions of what one was asked to do as a part of cancer screening.

Methods: The Cancer Screening Questionnaire (CSQ) was administered via random-digit-dial telephone interviews to 154 adults living in San Juan, PR and 155 in New York, NY.

Results: Although the self-reported willingness to participate across the 10 site-specific cancer screening exams was consistently high in both cities, SJPR had higher rates, as compared to NYPR for all 10 site-specific cancer screening exams in the unadjusted analyses. A similar pattern was observed regarding the influence of both "who conducts the cancer-screening exam" and "what one is asked to do in a cancer-screening exam" as factors in the willingness to participate in such exams. Adjusted multivariate analysis showed that the odds of SJPR participating in skin cancer screening as compared to NYPR, were three-fold higher to participate in skin cancer screening and were two-fold higher to participate in a cancer screening where they have to be interviewed about their alcohol habits. These two observed differences might reflect the effect of acculturation in the NYPR.

Key words: cancer ■ prevention ■ Latinos ■ acculturation

 From Community Dentistry Section, School of Dentistry, University of Puerto Rico, Río Piedras, PR (Claudio); Department of Epidemiology & Health Promotion, NYU College of Dentistry, New York University, New York, NY (Katz, Russell); Office of Institutional Diversity, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL (Green); Department of General Internal Medicine, Boston University Medical School, and Center for Health Quality, Outcomes & Economic Research, Department of Veterans Affairs, Bedford, MA (Kressin); and Department of Statistics, University of Maryland, College Park, MD (Wang). Send correspondence and reprint requests for *J Natl Med Assoc.* 2007;99:542–549 to: Dr. Cristina Claudio, Community Dentistry Section, School of Dentistry, University of Puerto Rico, Río Piedras, PR; phone: (787) 758-2525, ext. 1132; fax: (787) 756-7145; e-mail: cclaudio@rcm.upr.edu

INTRODUCTION

Cancer, the second leading cause of death in the United States, accounts for 22.7% of all deaths. According to the Surveillance Epidemiology and End Results (SEER, 2005) Cancer Statistics, prostate and breast cancers have the highest age-adjusted (to the 2000 U.S. standard population) incidence among all cancers with a rate of 182.5 and 137.3 per 100,000 respectively, followed by lung and colorectal cancers. These latter cancers had, for both genders, an age-adjusted (to the 2000 U.S. standard population) incidence rate of 63.7 per 100,000 and 53.1 per 100,000, respectively.[1]

While the scientific literature states that early detection of cancer is one of the most effective means of assuring timely treatment and survival, too few people take advantage of the available early screening tests for common cancers.[2] For instance, in the United States, only about 40–50% of the population receives screening for colon cancer, 58% of the women ≥40 receives a mammogram and 54% of men ≥50 get a PSA test for prostate cancer.[3] Only about 20% of the population receives an oral cancer exam[4] despite a recent finding that oral cancer screening in high-risk subjects is effective in reducing of oral cancer mortality.[5]

In Puerto Rico, cancer is also the second cause of death, accounting for 17% of all deaths.[6] According to the Puerto Rico Central Cancer Registry (2006), cancers with the highest age-adjusted (to 2000 U.S. standard population) incidence are, as in the United States, breast and prostate. They have an incidence rate of 98.0 and 145.6 per 100,000, respectively, followed by colorectal cancer with 42.8 per 100,000 for both genders.[7] These cancers are also, for both countries, the most common fatal cancers.[8]

Lower screening rates have direct implications for the quality of cancer care since low utilization of early

detection tests increases the risk of a diagnosis of late-stage cancer in populations. In Puerto Rico, for example, a study which characterized head and neck squamous cell carcinoma (HNSCC), which constitutes approximately 90% of all oral cancers, found that 61% of the patients presented stage 3–4 disease at the time of diagnosis.[9] This late-stage diagnosis compromises not only the quality of life but also the five-year survival prognosis of the patient, even in places like the United States, where at the present time the survival rate is 65%.[10]

Disparities in risk for cancer exist among racial and ethnic groups in the United States. It is well documented in the literature that Hispanics and blacks have lower screening rates than whites.[11-13] Black and Hispanic patients, as well as others with lower levels of formal education, are less likely to have such examinations due perhaps to lack of access to medical care.[4] Low income, poor knowledge and attitudes towards the screening process, not having a regular physician, language, cultural beliefs, and competing demands of day-to-day living are, among others, barriers that contribute to the underutilization of cancer screening exams.[14-16] Thus, it is not surprising that the burden of cancer deaths is particularly high among blacks and Hispanics, as compared to Caucasians.

In minority groups, such as Hispanics, the effect of acculturation in health-related behaviors has also been studied. Acculturation is the process by which members of a different cultural or ethnic group in a given society come to identify with or adopt the cultural changes that result from continuous contact between two cultural groups.[17] Several factors contribute to the process of acculturation, e.g., length of residence, language proficiency, ethnic/national attitudes of the individual, social support, discrimination, a reciprocal interaction between the individual and the environment, cultural beliefs and so forth.[18,19]

The effect of acculturation on Hispanics' health in the United States is inconsistent[18,20-23] due perhaps to the multidimensional factors involved in the acculturation process.[24,25] Although several studies have found that acculturation negatively affects the health practices of Hispanics and other minority groups,[26-29] other studies concluded that high levels of acculturation are associated with better self-reported measures of general health, better oral health status index score (OSHI), influenced care-seeking behaviors and increased health preventive behaviors such as clinical breast exams and mammograms.[18,30-32]

Since Puerto Ricans are recognized as minorities on the U.S. mainland but not on the island of Puerto Rico, they face many challenges involved in the process of acculturation, i.e., new language, different customs and norms for social interactions, unfamiliar rules and lifestyle changes. Consequently, it can be hypothesized that since Puerto Ricans in New York City have undergone a process of change and adaptation, they will behave differently than those Puerto Ricans that have lived life long in Puerto Rico, especially regarding their cancer screening participation/behavior. This being the situation, the strategies used to motivate this population to participate in cancer screening exams, for example, tailored-made cancer screening campaigns, should acknowledge these changes.

The specific purpose of this analysis of the Cancer-Screening Questionnaire (CSQ) Study was to determine whether—and if so, the extent to which—Puerto Ricans living in New York, NY (NYPR), as compared to Puerto Ricans living life long in San Juan, Puerto Rico (SJPR), differed in their willingness to participate in cancer screening exams. Specifically, this analysis aimed to determine the comparative likelihood of NYPR and SJPR to participate in: 1) 10 different site-specific type of cancer screening exams, 2) cancer screenings conducted by different specific persons/agencies, and 3) cancer screenings under specific conditions of what one was asked to do as a part of the cancer screening exams. The CSQ Study was approved by the University of Puerto Rico, Medical Sciences Campus institutional review board (IRB) and the IRB of New York University (NYU).

METHODS

The Cancer Screening Questionnaire

The instrument used for this study was the CSQ, which was developed in 2001–2002 within the NYU Oral Cancer Research on Adolescent and Adult Health Promotion Center, an NIDCR/National Institutes of Health Oral Health Disparities Center. The English and Spanish versions of the questionnaire were developed by a multidisciplinary and multiuniversity research team. The Spanish version was validated through pilot studies in San Juan, Puerto Rico. The CSQ addresses a range of issues related to the willingness of minorities to participate in cancer screening examinations. This instrument has 60 questions, eight of which obtained demographic information. It is divided into nine topics, including

Table 1. Demographics of San Juan and New York Puerto Ricans

	SJPR	NYPR
	n=154	n=155
% Female	67 %	72 %
Mean Age (± SD)	50 ± 16.5	44 ± 14.8
Income Level		
<$20,000	42.6%	36.2%
$20,000–34,999	23.4%	28.4%
≥$35,000	34 %	35.5%
Educational Level		
< High school	11.7%	27.7%
High-school graduate	13.0%	31.0%
Some college or higher	74.7%	40.6%

willingness to participate in different site-specific cancer screening exams, the influence of "who conducts the cancer screening exam" and of "what one is asked to do in a cancer-screening exam" as key factors in the subject's willingness to participate in cancer screening exams. These latter topics are discussed in this report.

The Spanish version used in San Juan had additional eligibility questions that asked subjects if they had lived all their life in Puerto Rico. Those who had not and had lived on the U.S. mainland for longer than two years in a row were not eligible for the study. However, no eligibility question regarding length of time living in the United States was used for the NYC subjects because of the possibility that it could be threatening to Hispanics in the United States regardless of their legal or illegal status.

The Spanish version was administered to San Juan subjects, while the NYPR subjects could decide whether they preferred to be interviewed in English or Spanish.

A five-point Likert scale was utilized as response choices for subjects on the CSQ. The scale consists of very likely (VL), somewhat likely (SL), not quite sure (NQS), somewhat unlikely (SU) and very unlikely (VU) for the questions analyzed in this report. To create a dichotomous response for certain analyses, VL + SL were combined to create the positive response of "willingness to participate" used in this report.

To gather the demographic information, the following variables were used: the "date of birth" variable was used to calculate the subjects' ages; the level of education and level of income variables were collected in an ordinal listings of nine ascending categories of educational level and of 10 ascending categories of income level. These variables were then each collapsed into three categories for the demographic table and the multivariate analyses.

Study Population

The sample for this study is composed of 309 Puerto Ricans: 154 SJPR and 155 NYPR. It was drawn from the total noninstitutionalized adult population (≥18 years) residing in telephone-equipped dwelling units in New York City and San Juan. This study provided for a disproportionally allocated, stratified random-digit sample of telephone-equipped residential households in both cities. The percentage of households with telephone service for NYPR was 97.6% and 84% for SJPR. Each city was sampled independently, with a target number of 150

Table 2. Percent difference between San Juan Puerto Ricans and New York Puerto Ricans in willingness to participate in nine site-specific cancer screening exams

Site-specific Cancer Screening	Males				Females		
	SJPR N=51 % Willing	NYPR N=44 % Willing	% Difference (SJPR-NYPR)		SJPR N=103 % Willing	NYPR N=111 % Willing	% Difference (SJPR-NYPR)
Skin	86.3	61.4	24.9* ‡		86.4	73	13.4* ‡
Rectal	88.2	63.6	24.6* ‡		77.5	81.1	-3.6
Blood	90.2	65.9	24.3* ‡		86.4	87.4	-1
Liver	84.3	68.2	16.1		76.7	74.8	1.9
Oral	82.4	68.2	14.2		81.6	75.7	5.9
Colon	88	74.4	13.6		82.4	79.1	3.3
Stomach	86.3	75	11.3		82.5	79.1	3.4
Lung	88.2	77.3	10.9		85.4	74.8	10.6
Prostate	88.2	81.4	6.8		-	-	-
Breast	-	-	-		95.1	93.7	1.4

* Significant difference, p≤0.05, unadjusted analysis; ‡ Significant difference, p≤0.05, multivariate analysis, controlling for age, sex, income and education; - The question was not asked to these subjects

completed interviews within each city. The NYPR population includes both Puerto Ricans who where born in NYC and persons of Puerto-Rican descent.

ORC Macro, an opinion research corporation, administered the CSQ to adults living in these two cities in 2003. The instrument was administered via random-digit dial using a computer-assisted telephone interview) system for data collection. The telephone survey followed a 10-attempt dialing protocol in which up to 10 attempts were made unless a final disposition was obtained. The response rates (number interviewed / number of phone numbers contacted) for San Juan and New York were 57.9% and 45.1%, respectively. The completion rate (number of completed interviews / total number of interviews initiated) was 84% for San Juan and 83% for New York City.

Statistical Analysis

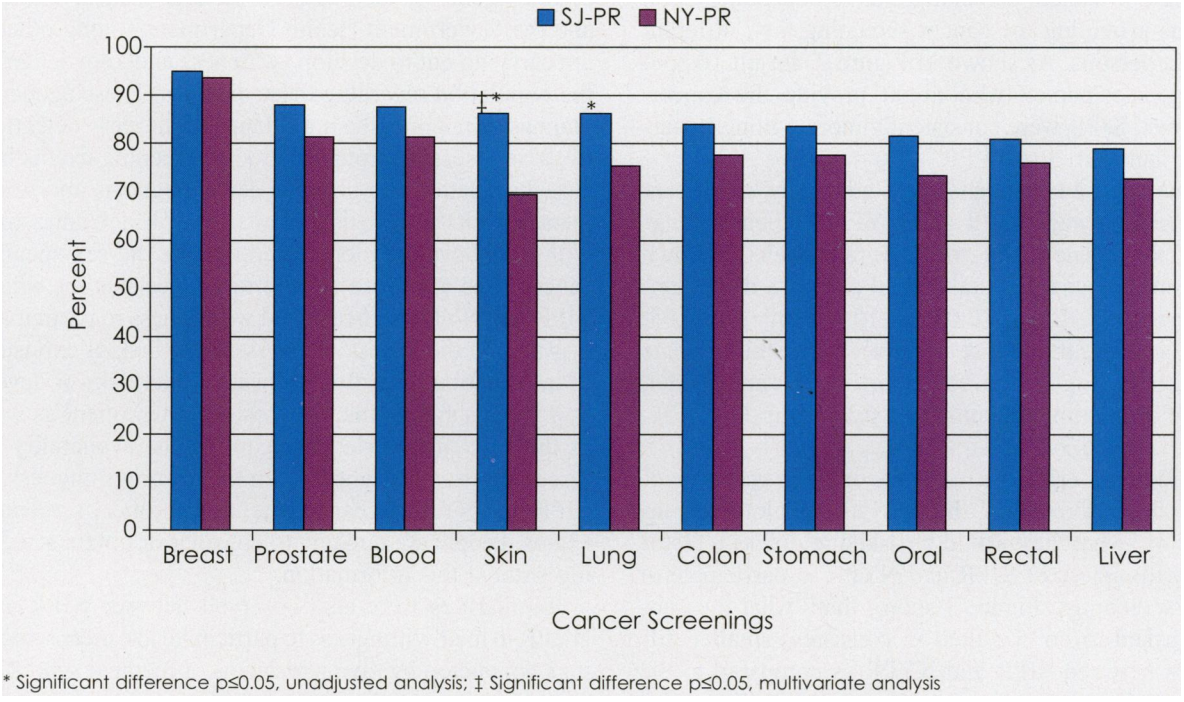
The aim of the statistical analysis performed for this study was to determine if SJPR and NYPR differed in their willingness to participate in cancer screenings. In addition, statistical analysis was performed by gender to determine if there were significant sex differences between SJPR's and NYPR's willingness to participate in 10 site- and sex-specific cancer exams. Unadjusted analysis was used as a pathway leading to adjusted multivariate analyses. Statistical significance in cross-tabulations was evaluated by means of Chi-squared tests with the significance level set at $p \leq 0.05$. Multivariate logistic regression analysis adjusted for gender, education, age and income was also performed at 0.05 level of significance.

RESULTS

Table 1 shows the composition of the sample by gender, age, education and income level. NYPR had slightly higher percentage of females (72% vs. 67%) and had slightly higher income than SJPR; these differences were not statistically significant. SJPR, however, did report significantly higher levels of education than NYPR, i.e., 74.5% with some college or higher educational level versus 40% in NYPR (Pearson $\chi^2=37.0$, df=2, $p \leq 0.001$), and were significantly older than NYPR (50 years vs. 44 years), (t (307)=3.47, p=0.001).

Figure 1 shows the comparative rank order of the 10 site-specific cancer screening exams for SJPR (n=154) and NYPR (n=155) listed in order based upon the descending rank order of SJPR. It also shows that while overall self-reported willingness to participate across the 10 site-specific cancer screening exams was consistently high in both cities (ranging from 70% to a high of 95%) SJPR had higher rates, as compared to NYPR for all 10 site-specific cancer screening exams. As shown in Figure 1, subjects within each city placed the sex-specific cancers screenings (i.e., breast cancer and prostate cancer) and blood cancer screening exams in the top tercile, while placing liver and oral cancer screenings in the lowest tercile. The only significant difference be-

Figure 1. Comparison between San Juan Puerto Ricans' and New York Puerto Ricans' willingness to participate in cancer screenings for the 10 site-specific cancer screenings



* Significant difference p≤0.05, unadjusted analysis; ‡ Significant difference p≤0.05, multivariate analysis

tween SJPR and NYPR in willingness to participate was evidenced for skin cancer screening exams. This significant difference was observed in both the unadjusted and the multivariate analysis adjusted for demographics (χ^2=12.51, df=1, p≤0.05; OR=3.33, 95% CI: 1.40–7.71, respectively).

Unadjusted analysis by gender (Table 2) shows differences in willingness to participate between SJPR and NYPR for the nine site- and sex-specific cancer screenings. For males, SJPR were consistently more willing to participate in all nine site- and sex-specific cancer-screenings (mean of 16% with a range from a high difference of 24.9% to a low of 6.8%). Significant differences were observed in both unadjusted and adjusted multivariate analysis controlling for demographics, for skin (χ^2=7.763, df=1, p≤0.05; OR=3.84, 95% CI: 1.28–11.49), rectal (χ^2=8.032, df=1 p≤0.05; OR=4.10, 95% CI: 1.38–12.17) and blood (χ^2=8.383, df=1, p≤0.05; OR=4.79, 95% CI: 1.46–15.70) screenings. For females, although the same direction of findings was observed in the unadjusted analyses (i.e., generally albeit not consistently higher willingness to participate in SJPR females, as compared to NYPR females), the magnitude of the difference in willingness to participate was much reduced in the females (mean=4%, ranging from -3.6%–13.4%) as compared to the males. Statistically significant differences were only found between SJPR and NYPR females in their willingness to participate in skin cancer-screening exams (χ^2=5.904, df=1, p≤0.05). This difference remained significant when the multivariate analysis controlling for demographic characteristics was performed (OR=2.64, 95% CI: 1.20–5.80).

Subjects were also asked about their likelihood to participate in cancer screening exams depending upon who was providing the cancer screening, i.e., different agencies/persons. As shown in Figure 2, for all 10 specific agencies/person who might provide the cancer-screenings, SJPR were consistently more willing to participate than were the NYPR.

Unadjusted analysis showed statistically significant differences between SJPR and NYPR in their willingness to participate if the exams were provided by: own physician, university dental school and by a drug company (χ^2=3.992, df=1, p≤0.05; χ^2=10.825, df=1, p≤0.05; χ^2=15.74, df=1, p≤0.05, respectively). In a multivariate analysis adjusting for age, education, income and gender, only the drug company remained significant (OR=3.21, 95% CI: 1.63–6.34).

Besides exploring the influence of "who was providing the cancer screening," the study also explored the influence of "what they would be asked to do" as a factor in the willingness of SJPR and NYPR to participate in cancer screenings. Figure 3 shows that "what the subject is asked to do" resulted in consistent smaller differences between SJPR and NYPR as compared to the factor of who was conducting the screening (Figure 2). This figure also shows that SJPR self-reported a greater willingness to participate in cancer screening for 10 of the 11 specific "what one was asked to do" conditions (mean=7.5%, range 2.3–12.7 %). The sole exception in direction of findings was for the circumstance of "having a nurse examine you" for which the NYPR self-reported a higher willingness to participate. Unadjusted analysis showed that statistically significant differences were found between SJPR and NYPR in their willingness to participate in cancer screenings when they were asked to: have a piece of skin removed and a physician examining them (χ^2=5.49, df=1, p≤0.05 and χ^2=9.04, df=1,p≤0.05). The multivariate analysis adjusting for demographics, however, revealed that the odds of SJPR participating in a cancer screening in which they are interviewed about their alcohol drinking habits were two-fold higher as compared to NYPR (OR=2.11, 95% CI: 1.09–4.08) as the only statistically significant finding.

DISCUSSION

Findings of this study show more similarities than differences in the willingness of SJPR and NYPR to participate in 10 site-specific cancer screening exams regardless of the influence of "who was conducting the exam" or "what they are asked to do" factor. SJPR and NYPR, for example, self-reported more willingness to participate in the site-specific cancer-screening exams of breast and prostate than in oral cancer screening exam. Findings reveal that these sex-specific cancer screening exams obtained the highest percentage of willingness to participate within each gender. The high willingness to participate in the sex-specific cancer screening exams likely reflects the considerable efforts that public and private agencies, such as the American Cancer Society and the Government Health Department among others, are carrying out to develop awareness and knowledge in the population regarding these diseases. These agencies emphasize not only the importance of the early detection of these diseases through periodic screening exams but also the relationship between early diagnosis and treatment with its prognosis and quality of life. Contrasting with the above-mentioned findings for the sex-specific cancer sites are those related to oral cancer for which SJPR and NYPR self-reported willingness to participate was among the lowest. Conversely, the lack of exposure of information regarding oral cancer, the risks of developing it, its symptoms, diagnosis and treatment as well as the lack of knowledge regarding the availability of screening exams might explain the lower willingness of SJPR and NYPR to participate in oral cancer screening exams. People are unlikely to consider or obtain screening without this information.[33]

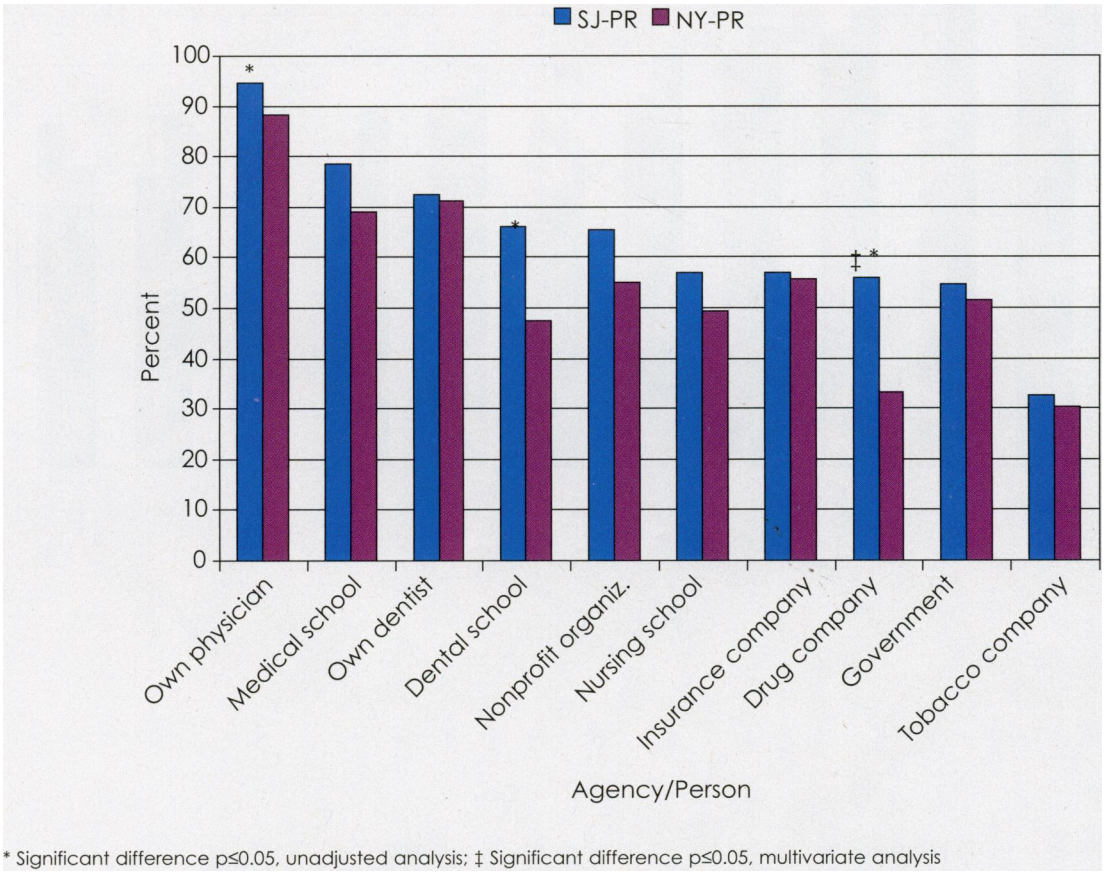
Similarities were also observed between SJPR and NYPR in their willingness to participate in cancer screenings conducted by own physicians, a medical school or own dentists as well as their unwillingness to participate

in cancer screenings conducted by a tobacco company. This willingness to participate in cancer screenings conducted by different healthcare professionals stresses how much the primary care provider can do to increase the participation in cancer screening exams of this Puerto-Rican population. Thus, the incorporation of different healthcare providers in the development of educational programs tailored to increase knowledge and to modify health preventive behaviors such as cancer screening exams should be considered as an adequate strategy for both populations.

On the other hand, a few statistically significant differences emerged as well. SJPR and NYPR differed significantly in their willingness to participate in skin cancer screening, screening exams conducted by drug companies and screening exams where they have to be interviewed about their alcohol drinking habits. As the results showed that the odds of SJPR participating in skin cancer-screening were three-fold higher than for NYPR, this willingness to participate in skin cancer screening exams might be due to their perception of having a lower risk of developing skin cancer. According to Pipitone (2002), Hispanics on the mainland perceive that they have less skin sensitivity and tendency to sunburn than their white counterparts. This belief could have being reinforced by multimedia messages that link skin cancer and early detection strategies with having sun-sensitive skin.[34] On the other hand, SJPR as tropical islanders are constantly exposed to sun, the major environmental risk for skin cancer. This fact makes them more aware of their potential high risk of developing skin cancer.

Similarly, when the 10 types of cancer screenings were analyzed by gender, SJPR and NYPR were generally found to be very similar on willingness to participate for most types of cancer screenings; however, a few statistically significant differences were detected. The odds of SJPR males as compared to NYPR males were 4–5-fold higher for self-reported willingness to participate in cancer screening exams for skin, rectal and blood cancer, while the odds of SJPR females were almost three-fold higher for self-reported willingness to participate in skin cancer screening exams, as compared to NYPR females. This comparative lack of willingness of NYPR males to participate in rectal cancer screening is congruent with previous studies conducted with Puerto-Rican males on the U.S. mainland.[35,36] According to these studies, Puerto-Rican men indicated that it would be too embarrassing for them to get a rectal exam, and

Figure 2. Comparison between San Juan Puerto Ricans and New York Puerto Ricans on willingness to participate in cancer screening based on what agency/person conducts the screening



* Significant difference p≤0.05, unadjusted analysis; ‡ Significant difference p≤0.05, multivariate analysis

they felt that others would disapprove of them getting such an exam. Another study that compared differences in screening exams, such as the digital rectal exam and prostate-specific antigen tests, among men of seven ethnic groups, including Puerto Ricans, found that fear was among the key psychological determinants that negatively influence male screening behavior.[37]

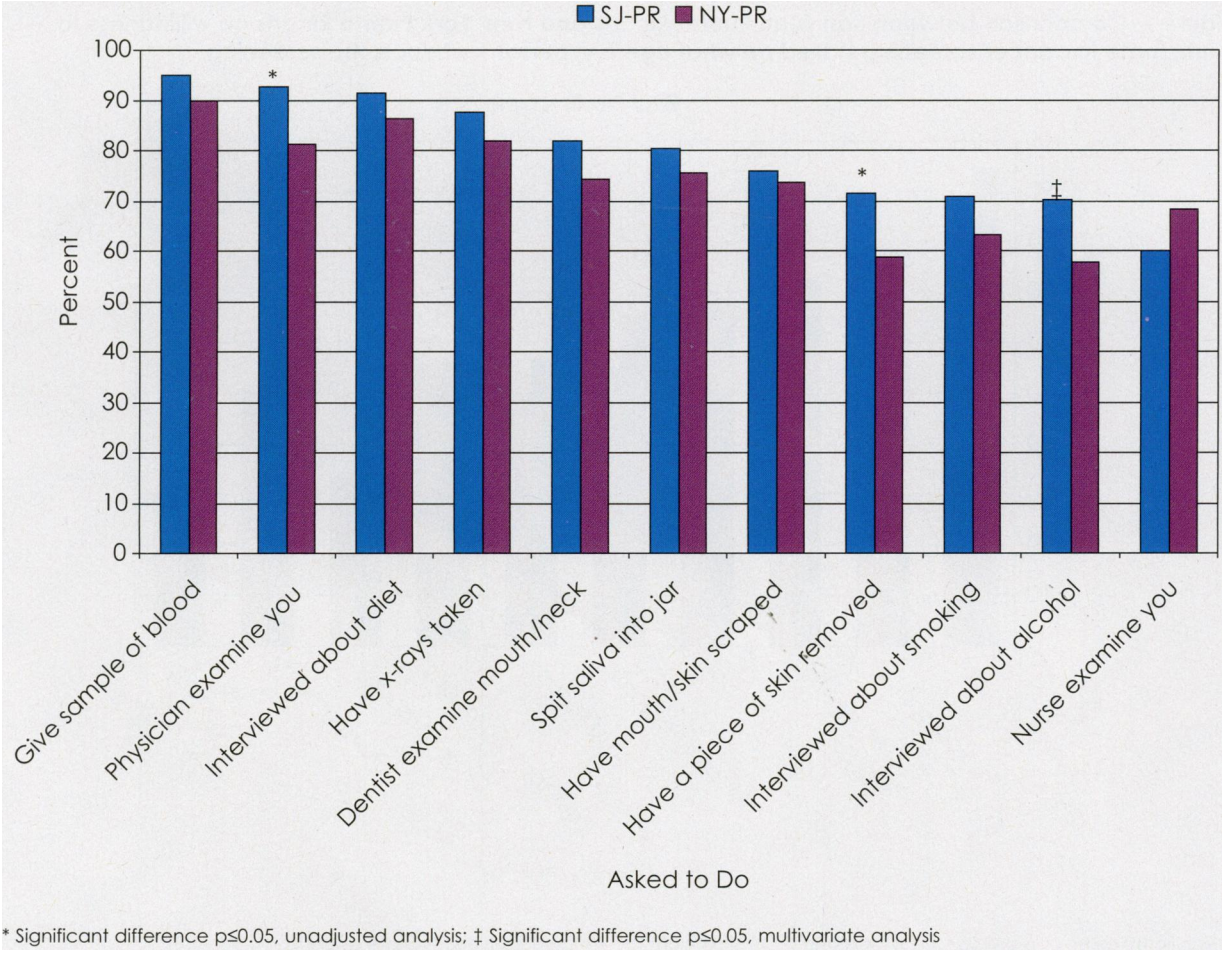
In regards to the influence of the factor "who is conducting the exam," the results showed that the odds of SJPR participating in cancer screening conducted by a drug company were almost three-fold higher than for NYPR. The extent of influence of drug companies is particularly high in Puerto Rico due to the fact that the pharmaceutical industry constitutes one of the leading industries in the country. It provides and promotes a diversity of health information and services, individually or in alliances with other health organizations, thus increasing SJPR exposure to activities sponsored by those companies. Cultural diversity in NYC, language barriers or promotional strategies may contribute to the minimal exposure of NYPR to the outreach services of these organizations.

In a similar fashion, for the "what they are asked to do" factor, the findings that show that the odds of SJPR as compared to NYPR are two-fold higher to participate in a screening where they have to be interviewed about their alcohol habits has clear implications for screening programs related to cancers associated with alcohol consumption, e.g., oral cancer screenings. Thus, to be effective and successful in the development and implementation of preventive programs geared to increase cancer screening participation, this barrier should be addressed and overcome.

Data gathered in this study provide information that is of paramount importance in the development and implementation of cancer screening campaigns for the Puerto-Rican population. As stated earlier, NYPR appear to be less responsive to cancer screening exams for skin cancer, where there are inquiries about their alcohol drinking habits or when the cancer screenings are conducted by drug companies.

In conclusion, the development of tailor-made can-

Figure 3. Comparison between San Juan Puerto Ricans' and New York Puerto Ricans' willingness to participate in cancer screenings based on what they are asked to do



* Significant difference p≤0.05, unadjusted analysis; ‡ Significant difference p≤0.05, multivariate analysis

cer screening campaigns that are responsive to these observed differences between SJPR and immigrant NYPR in self-reported willingness to participate in cancer screenings could be a fundamental and critical step in the process of increasing cancer screening participation in Puerto-Rican populations.

REFERENCES

1. National Cancer Institute, Cancer Statistics Review 1975-2003, Surveillance Epidemiology and End Results. Bethesda, MD. Based on Nov 2005 SEER data submission, posted to the SEER website, 2006. www.seer.cancer.gov/csr/1975_2003. Accessed 07/05/06.

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence – SEER 9 Regs Public-Use, Nov 2005 Sub (1973-2003), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2006, based on the November 2005 submission.

2. Jemal A, Siegel R, Ward E, et al. Cancer Statistics, 2006. *CA Cancer J Clin.* 2006;56:106-130.

3. Smith RA, Cokkinides V, and Harmon JE. American Cancer Society guidelines for the early detection of cancer, 2006. *CA Cancer J Clin.* 2006;56:11-25.

4. Kerr AR, Changrani JG, Gany FM, et al. An academic dental center grapples with oral cancer disparities: current collaboration and future opportunities. *J Dent Educ.* 2004;68(5):531-541.

5. Sankaranarayanan R, Ramadas K, Thomas G, et al. Trivandrum Oral Cancer-screening Study Group. Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomized controlled trial. *Lancet.* 2005; Jun 21; 365(9475): 1927-1933.

6. Puerto Rico Central Cancer Registry, Preliminary data on death rates-2003, Puerto Rico Department of Health, November 2006.

7. Puerto Rico Central Cancer Registry, Preliminary data on incidence rates-2003, Puerto Rico Department of Health, November 2006.

8. American Cancer Society. Cancer Facts and Figures 2003. Atlanta, GA. www.cancer.org. Accessed 07/07/06.

9. Busquets JM, García HA, Trinidad-Pinedo J, et al. Clinicopathologic Characteristics of Head and Neck Squamous Cell Carcinoma in Puertoricans. *P R Health Sci J.* 2003; September,22(3):259-264.

10. American Cancer Society. Cancer Facts and Figures 2006. Atlanta, GA. www.cancer.org/downloads/STT/CAFF2006PWSecured.pdf. Accessed 01/02/07.

11. Swan J, Breen N, Coates RJ, et al. Progress in cancer-screening practices in the US: results from the 2000 National Health Interview Survey. *Cancer.* 2003;15,97(6):1528-1540.

12. Zambrana RE, Breen N, Fox SA, et al. Use of cancer-screening practices by Hispanic women: analyses by group. *Prev Med.* 1999;29:466-477.

13. Brown DR, Topcu M. Willingness to Participate in Clinical Treatment Research Among Older African Americans and Whites. *Gerontologist.* 2003;43:62-72.

14. McCoy CB, Preyra M, Metsch LR, et al. A community-based breast cancer-screening program for medically underserved women: its effect on disease stage at diagnosis and on hazard of death. *Rev Panam Salud Pública.* 2004;Mar,15(3):160-167.

15. Abraído- Lanza AF, Chao MT, Gammon M. Breast and cervical cancer-screening among Latinas and non-Latina Whites. *Am J Public Health.* 2004; 94:1393-1398.

16. Luquis RR, Villanueva IJ. Knowledge, attitudes, and perceptions about breast cancer and breast cancer-screening among Hispanic women residing in South Central Pennsylvania. *J Community Health.* 2006;Feb,31(1):25-42.

17. Berry JW, & Sam DL. Acculturation and adaptation. In: Berry JW, Segall MH, Kagitcibasi C, eds. Handbook of cross-cultural psychology Vol. 3: Social Behaviour and Applications. Boston: Allyn & Bacon; 1997:291-326.

18. Abraído-Lanza AF, Chao MT, Gate CY. Acculturation and cancer-screening among Latinas: results from the National Health Interview Survey. *Ann Behav Med.* 2005; 29(1):22-28.

19. Hazuda HP, Stern MP, Haffner SM. Acculturation and assimilation among Mexican Americans: scales and population-based data. *Soc Sc Quarterly.* 1988;69:687-706.

20. Lara M, Gamboa C, Kahramanian MI, et al. Acculturation and Latino health in the United States: a review of the literature and its sociopolitical context. *Annu Rev Public Health.* 2005;26:367-397.

21. Franzini L, Ribble JC, Keddie AM. Understanding the Hispanic paradox. *Ethn Dis.* 2001;11:496-518.

22. Suarez L, Pulley L. Comparing acculturation scales and their relationship to cancer-screening among older Mexican American women. *J Natl Cancer Inst Monogr.* 1995:(18):41-47.

23. Solis JM, Marks G, García M, et al. Acculturation, access to care, and use of preventive services by Hispanics: findings from HHANES 1982-84. *Am J Public Health.* 1990;80 (Suppl):11-19.

24. Guevarra JS, Kwate NO, Tang T, et al. Acculturation and its relationship to smoking and breast self-examination, frequency in African-American women. *J Behav Med.* 2005;28(2):191-199.

25. Abraído-Lanza AF, Armbrister AN, Florez KR, et al. Toward a theory-driven model of acculturation in publiv health research. *Am J Public Health.* 2006;96(8):1342-1346.

26. Cantero PJ, Richardson J, Baezconde-Garbanati L, et al. The association between acculturation and health practices among middle-aged and elderly Latinas. *Ethn Dis.*1999;9(2):166-180.

27. Alaniz ML, Treno AJ, Saltz RF. Gender, acculturation, and alcohol consumption among Mexican-Americans. *Subst Use Misuse.* 1999;34:1407-1426.

28. Vega WA, Alderet E, Kolody B, et al. Illicit drug use among Mexicans and Mexican-Americans in California: the effects of gender and acculturation. *Addiction.* 1998;93:1839-1850.

29. Nieri T, Kulis S, Keith VM, et al. Body image, acculturation, and substance abuse among boys and girls in the Southwest. *Am J Drug Alcohol Abuse.* 2005;31(4):617-639.

30. Atchison KA, Der-Martirosian C, Gift H. Components of self-reported oral health and general health in racial and ethnic groups. *J Public Health Dent. 1998;58:301-308.*

31. Spolsky VW, Marcus M, Coulter ID, et al. An empirical test of the validity of the Oral Health Status Index (OHSI) on a minority population. *J Dent Res.* 2000;79:1983-1988.

32. Stewart DC, Ortega AN, Dausey D, et al. Oral health and use of dental services among Hispanics. *J Public Health Dent.* 2002;62:84-91.

33. Smith RA, Von Eschenbach AC, Wender R, et al. American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers. *CA Cancer J Clin.* 2001;51:38-75.

34. Pipitone M, Robinson J, Camara C, et al. Skin cancer awareness in suburban employees: A Hispanic perspective. *J Am Acad Dermatol.* 2002;July, 47(1):118-123.

35. Perez-Stable EJ, Otero-Sabogal R, Sabogal F, et al. Self-reported use of cancer screening tests among Latinos and Anglos in a prepaid health plan. *Arch Intern Med.* 1994;154:1073-1081.

36. Goldman R. Barriers to intimate cancer screening testing among Dominican and Puerto Rican men. www.brown.edu/Research/ICAP/Modules/Mod1 ensHealthy/TeachingMats/Handout3HispanicMales.doc. Accessed Dec.2006

37. Consedine N, Morganstern A, Kudadjie-Gyamfi E, et al. Prostate cancer screening behavior in men from seven ethnic groups: the fear factor. *Cancer Epidemiology, Biomarkers & Prevention.* 2006; February, 15:228-237. ■

JOURNAL OF WOMEN'S HEALTH
Volume 16, Number 10, 2007

DOI: 10.1089/jwh.2007.0483

Intervention Study of Exercise for Depressive Symptoms in Women

LYNETTE L. CRAFT, Ph.D.,[1] KAREN M. FREUND, M.D., M.P.H.,[2]
LARRY CULPEPPER, M.D., M.P.H.,[3] and FRANK M. PERNA, Ed.D., Ph.D.[2]

ABSTRACT

Background and objectives: **Clinical depression affects millions of women annually. Exercise has been studied as a potential antidepressant, with most studies supporting its efficacy. Exercise also has the potential to reduce the risk for physical comorbidities that occur with depression. However, less is known about the types of exercise programs to which women with depressive symptoms will adhere. Our objectives were to (1) compare two exercise programs, varying in their degree of structure, on improvements in physical activity and (2) compare the two exercise interventions on depressive symptoms, body composition, and fitness.**

Methods: **Women with depressive symptoms (physician diagnosed and confirmed with the Beck Depression Inventory) residing in the greater Boston area were recruited for this 3-month intervention study. Continuous enrollment took place between November 2005 and November 2006. Women were randomly assigned to either a clinic-based or home-based exercise intervention, with assessments at baseline and 3-months.**

Results: **Participants ($n = 32$) were predominantly minority (81.4%) and, at baseline, had moderate symptoms of depression (Beck Depression Inventory [BDI], mean = 25.6, SD = 9.3), and were sedentary (mean = 35.8 min/week of moderate and vigorous activity, SD = 31.4). Gain scores for depressive symptoms (clinic-based mean = −11.7, home-based mean = −9.7) and physical activity (clinic-based mean = 65.4, home-based mean = 39.0) indicate strong improvements across time. Intent-to-treat analyses on 3-month data show that both interventions were associated with improvements in time spent in physical activity and depressive symptoms. Neither intervention impacted body composition or fitness.**

Conclusions: **Both exercise programs were associated with reductions in depressive symptoms and increased physical activity participation, suggesting that even a home-based program can benefit women with depressive symptoms.**

INTRODUCTION

ONE IN FIVE WOMEN will be diagnosed with a depressive disorder in her lifetime. Beyond the reduction in mental health and quality of life associated with this disease, depression also increases the risk of developing several chronic diseases,[1–5] such as cardiovascular disease (CVD) and diabetes, amplifying the mortality and financial impact of depressive disorders. Each year, over $80 billion is lost in productivity and medical costs related to clinical depression, mak-

[1]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois.
[2]Women's Health Unit, Section of General Internal Medicine, and Women's Health Interdisciplinary Research Center, and [3]Department of Family Medicine, Boston University School of Medicine, Boston, Massachusetts.
This work was supported by the National Institutes of Health: the Office of Women's Health Research (K-12-HD4244 and the Clinical Research Feasibility Funds (M01RR00533).

ing it one of the most costly illnesses in this country.[6] Physical inactivity and obesity commonly co-occur with depression, further contributing to the development of physical comorbidities in individuals with depression.[1]

Exercise and depression

The relationship between exercise and depression has been studied for over 100 years, and involvement in structured exercise has shown great promise in alleviating symptoms of clinical depression. Although much of the early research in this area was criticized as suffering from a variety of methodological flaws (e.g., small samples, lack of random assignment, lack of control groups), current researchers have addressed these design issues, and there are at present multiple studies using experimental designs or randomized clinical trial (RCT) approaches that demonstrate the antidepressant effects of exercise.[7–14]

Meta-analytic findings indicate that exercise interventions have a large impact on depression, with between-groups effect sizes (ES), ranging from -0.72 to -1.4.[15–18] Further, exercise is effective for a variety of patient subgroups.[15] That is, patients with both moderate and more severe depression appear to benefit similarly. Exercise efficacy is comparable to and not significantly different from psychotherapy and pharmacological treatment.[15,18] Meta-analytic studies suggest that fitness gains are not necessary to achieve reductions in depression and that exercise program characteristics, such as duration, intensity, frequency, and mode of exercise, do not moderate the effect. The length of the exercise program is a significant moderator, however, with programs 9 weeks or longer associated with larger reductions in depression.[15] Conversely, emerging experimental research demonstrates that a dose-response relationship may exist, with more exercise associated with larger reductions in depression, implying that exercise characteristics such as frequency, duration, and intensity may be important.[10,14] Few studies to date have directly examined this, and more research is needed; however, extant findings do suggest that exercise need not be lengthy or intense to promote positive changes in depressive symptoms. It may be the case that individuals who can exercise at public health recommended doses will experience even larger benefits.

Implementing exercise into treatment

A poorly understood area within this body of research relates to how extensive an intervention need be to promote exercise adherence. Researchers examining the antidepressant effects of exercise have predominantly used supervised, hospital-based exercise protocols, with fewer studies using home-based programs. However, in the general population, long-term adherence to exercise programs has been improved with the use of home-based or community-based programs, the addition of exercise adherence counseling, and brief supportive follow-up contact.[19–28] Keeping in mind the time constraints of clinicians, it remains unclear how much information, counseling, and contact are necessary to help women with depressive symptoms initiate and maintain an exercise program as a treatment component for reducing the severity of symptoms. Further, exercise adherence may be a necessary factor not only for a reduction in depressive symptoms but also for the physical health benefits that exercise can provide in reducing risk for the development of physical comorbidities.

Therefore, the primary purpose of the present pilot study was to compare two types of exercise programs, varying in their degree of supervision and structure, on improvements in physical activity participation. A secondary aim was to compare the two exercise interventions on depressive symptoms, body composition, and fitness. Further, because minority enrollment in previous RCTs of exercise and depression has been minimal, we specifically targeted urban minority women with depression from a community hospital for this study. It is predicted that women in both clinic-based and home-based exercise groups will experience a significant reduction in depression. Further, it is hypothesized that a clinic-based exercise intervention that includes cognitive behavioral techniques (e.g., self-monitoring, goal setting) to enhance exercise adherence will be superior to a home-based intervention in promoting physical activity participation, reducing depressive symptoms, decreasing body fat, and increasing cardiovascular fitness at 3-month follow-up.

MATERIALS AND METHODS

This pilot study was a randomized trial of two exercise strategies targeting low-income, urban, mi-

nority women. Participants were stratified by current depression treatment (i.e., no treatment, medication only, therapy only, medication plus therapy) and then randomly assigned to either a clinic-based or home-based exercise intervention with assessments at baseline and 3 months. Assessments consisted of a symptom-limited graded exercise test on a treadmill, body composition analysis, physical activity recall, pedometer counts, and a psychosocial survey.

Participants

Prior to data collection, all participants signed an informed consent statement as approved by the Institutional Review Board at Boston University. Participants were recruited through The Center for Excellence in Women's Health, Family Medicine, and posted fliers on the Boston University Medical Center (BUMC) Campus. Recruitment took place continuously over a 12-month period. Eligibility criteria included (1) English speaking, (2) age between 18 and 55 years, (3) meeting DSM-IV criteria for Major Depression, Dysthymia, or Depressive Disorder-Not Otherwise Specified as diagnosed by a primary care physician and scoring >9 on the Beck Depression Inventory-II (BDI-II), (4) sedentary lifestyle (i.e., exercise less than 3 times/week for >30 minutes per session in last 6 months), and (5) average or below average level of fitness as determined by graded exercise testing. Exclusion criteria included (1) diagnosed coronary heart disease (CHD), (2) insulin-requiring diabetes identified by patient self-report, (3) self-reported comorbid mental health diagnosis (e.g., anxiety disorder, bipolar disorder, schizophrenia, and (4) any contraindication to exercise determined by the physician supervising the exercise stress test. Further, for women taking antidepressant medication or receiving psychotherapy, study entry was delayed until they had been taking their medication or seeing a therapist for at least 8 weeks. Finally, all women in the study were required to obtain medical clearance to exercise and physician confirmation of depression diagnosis prior to enrollment. Enrollment into the study is outlined in Figure 1. Ultimately, 32 women enrolled in this longitudinal pilot study, and sample characteristics are presented in Table 1.

Instruments

Demographic questionnaire. All participants completed a basic demographic questionnaire that assessed variables, including age, marital status, race, family income level, medical history, and current treatment for depression.

Beck Depression Inventory-II (BDI-II). The BDI-II is a reliable, valid, and widely used 21-item self-report inventory designed to assess the severity of clinical depression and changes in depressive symptoms across time.[29] This inventory has demonstrated good internal consistency and is widely used to assess changes in depressive symptoms across time. For each item, the participant is asked to choose, from a list of statements, the one statement that best reflects the way she has been feeling during the past 2 weeks. Statements are organized according to severity and rated on a 4-point scale ranging from 0 to 3, with higher scores indicative of more severe symptoms. Cutoff points for scoring the inventory are: <14, minimal depression; 14–19, mild depression, 20–28, moderate depression, and 29–63, severe depression. A score <9 indicates that the depression is in remission. None of the study participants scored <9 on the BDI-II at study entry. Internal consistency for this sample was acceptable, with Cronbach's $\alpha = 0.83$ at baseline and 0.84 at 3-month follow-up.

Body mass index (BMI). A standard, regularly calibrated hospital scale and stadiometer were used to assess weight and height. Participants were asked to remove their shoes and heavy outer clothing prior to stepping on the scale. BMI was calculated at each visit as [weight (kg)/height (m^2)].

Body composition assessment. To assess body composition, each participant underwent a whole body scan using a Hologic QDR-4500 dual-energy x-ray absorptiometer (DEXA) (Hologic, Waltham, MA). Densitometry scanning provides a quick, noninvasive, highly accurate percentage of body fat and is considered a reliable method of determination of body composition. The test-retest reliability of the DEXA has been demonstrated with coefficients of variation (CVs) over a 5- to 7-day period of 1.6%, which is similar to underwater weighing and bioelectrical impedance. DEXA is also considered a valid assessment of body composition, as it is highly correlated with underwater weighing, the current gold standard in assessing body composition ($r = 0.87$).[30]

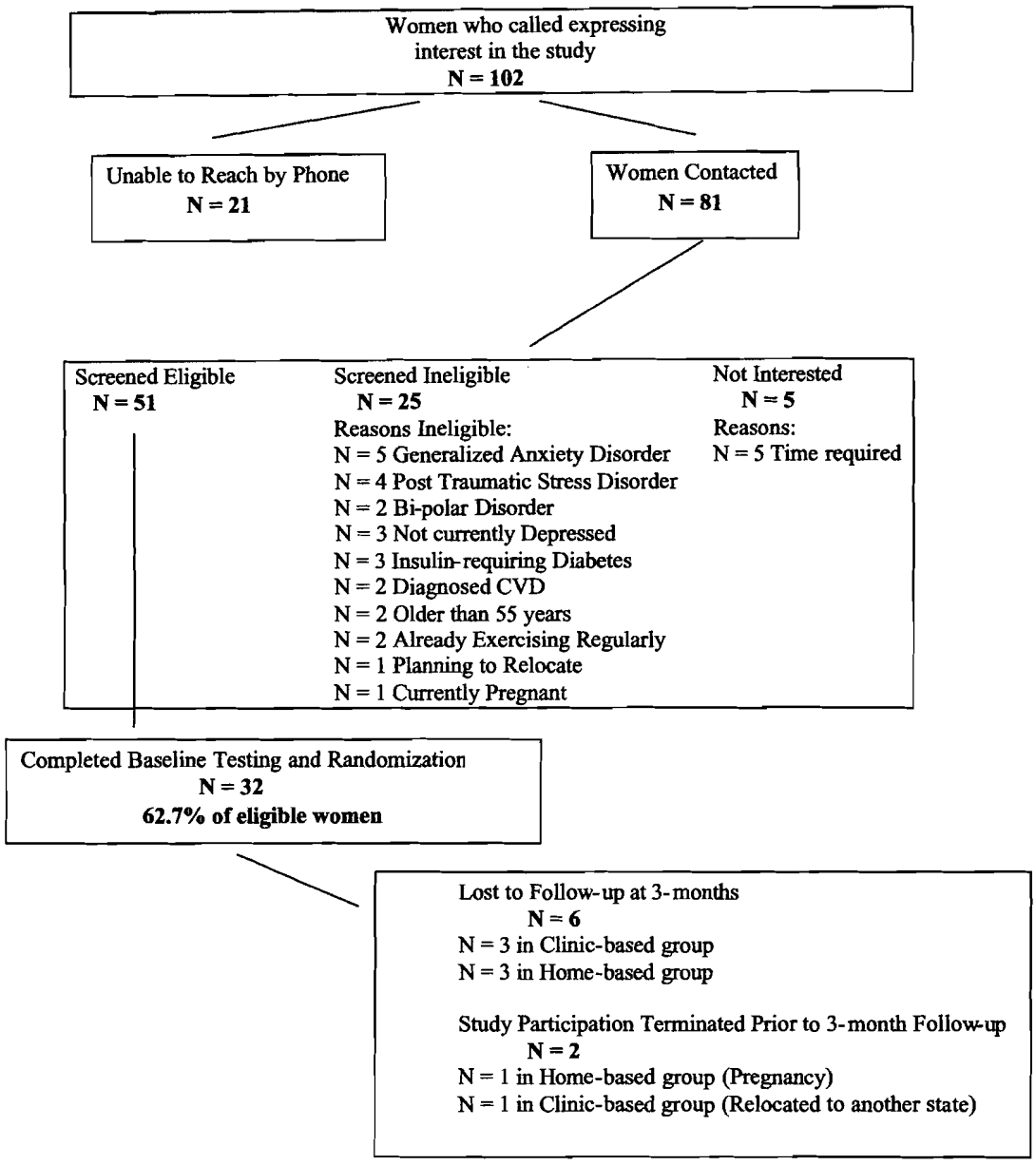


FIG. 1. Patient flow diagram.

Cardiovascular fitness. Participants underwent a physician-monitored graded exercise test on a treadmill using a Balke protocol. Physicians monitoring exercise tests were blinded to participant group assignment. A 12-lead ECG and a Physio-Dyne metabolic cart (Quoque, NY) were used to determine peak heart rate, peak oxygen consumption (Vo_2 peak), total work, time on test, and suitability for exercise. Vo_2 peak served as the primary fitness marker in this study.

The 7-day physical activity recall (PAR). The PAR[31] is a widely used and well-validated brief self-report recall assessing the frequency, duration, and intensity of physical activity, and it yields several physical activity indexes (minutes of exercise at each level of exercise intensity, number of days exercised, and a rough estimate of caloric expenditure over the week). The PAR is administered in a semistructured interview format, with the administrator probing for activity

TABLE 1. BASELINE SAMPLE CHARACTERISTICS

	Pooled (n = 32) Mean (SD)	*Home-based (n = 16) Mean (SD)*	*Clinic-based (n = 16) Mean (SD)*	*p value*
Age, years	40.4 (10.6)	37.4 (10.8)	43.4 (9.9)	0.11
Ethnic background				
African American	68.8%	81.2%	56.3%	0.07
Caucasian	18.7%	6.3%	31.2%	
Latina	12.5%	12.5%	12.5%	
Marital status				
Single	53.1%	56.3%	50.0%	0.42
Married/partnered	18.8%	25.0%	12.5%	
Separated/divorced	28.1%	18.2%	37.5%	
Employment				
Full-time	28.1%	31.3%	25.0%	0.92
Part-time	18.8%	18.2%	18.2%	
Not employed	53.1%	50.0%	56.3%	
Income				
<$20,000	74.2%	81.3%	66.7%	0.64
$21,000–30,000	16.1%	12.5%	20.0%	
$31,000–55,000	9.7%	6.2%	13.3%	
>$55,000	0.0%	0.0%	0.0%	
Education				
Not high school graduate	12.5%	18.8%	6.3%	0.10
High school graduate	31.3%	43.8%	18.7%	
Some college	56.2%	37.4%	75.0%	
Current treatment for depression				
None	56.3%	68.7%	43.8%	0.24
Medication only	15.6%	6.3%	25.0%	
Psychotherapy only	3.1%	6.3%	0.0%	
Medication and therapy	25.0%	18.7%	31.3%	
Percentage of participants with minimal, mild, moderate, and severe depression				
Minimal (BDI score <14)	9.4%	12.5%	6.3%	0.51
Mild (BDI score 14–19)	15.6%	6.3%	25.0%	
Moderate (BDI score 20–28)	46.9%	50.0%	43.8%	
Severe, (BDI score 29–63)	28.1%	31.2%	25.0%	
Reported difficulty sleeping				
None	25.0%	12.5%	37.5%	0.06
<30 minutes interruption	6.2%	0.0%	12.5%	
≥30 or more min interruption	68.8%	87.5%	50.0%	

and clarifying the intensity of activity reported. Two-week test-retest reliability has been reported to range from $r = 0.61$ to $r = 0.99$, and correlations between the PAR and other measures of physical activity have been reported as $r = 0.61$ (Vo_2 max), $r = 0.66$–0.83 (self-reported activity), and $r = 0.28$–0.43 (accelerometer).[32,33] Total minutes spent in moderate and vigorous activity during the previous week was used as the subjective physical activity outcome measure.

Pedometer counts. Pedometers provide an unobtrusive, lightweight, and widely used method to objectively monitor physical activity. The Yamax Digi-Walker SW-701 pedometer (Tokyo, Japan) was used in this study and previously has been found suitable for research purposes.[34] Reliability of the Yamax Digi-Walker is reported at $\alpha = 0.99$, and when compared with actual steps taken at five different speeds on a treadmill, the Yamax Digi-Walker was accurate within ±1% of actual steps taken.[34,35] Each month, participants were contacted and asked to reset the pedometer and to wear the pedometer for 7 consecutive days for data collection purposes. Participants then reported steps taken per day via telephone contact with study staff or in person at the time of their 3-month assessment. Average steps and distance (miles) per day were used as our objective physical activity outcome measures.

Intervention groups

Clinic-based intervention. At study entry, those assigned to the clinic-based exercise group were provided assessment feedback and information on how exercise may help with symptoms of depression. Further, they were given an individualized moderate-intensity walking program based on exercise testing results. They then participated in a 4-week clinic-based training phase during which they came to the medical center twice weekly for exercise training and were asked to complete one exercise session weekly at home. During laboratory training, exercise duration and intensity were gradually increased according to participant comfort, with the ultimate goal of walking 30–40 minutes in a training zone of 60%–80% maximal heart rate (MHR).

Following this 4-week phase, clinic-based exercise group members were transitioned to the home-based phase of the study and were asked to continue the walking program at home for the remaining 8 weeks of the study. Participants received a pedometer and instructions on its use for self-monitoring (i.e., exercise calendar and goal-setting information) and data collection purposes. Once each month, pedometer and physical activity recall measures of physical activity were collected. In addition, women in this group received up to four 30-minute adherence counseling sessions to aid in adherence to the clinic-based exercise, transition to home-based exercise, and maintenance of exercise during follow-up. Participants also received biweekly, brief supportive follow-up contact via telephone to encourage continued exercise participation, set new exercise and pedometer goals, and brainstorm solutions to new barriers encountered. The average number of laboratory-based exercise sessions completed by clinic-based group participants was 6.2 of a possible 8, or 77.5%. There were no adverse events reported by women in the clinic-based group.

Home-based exercise intervention. Those assigned to the home-based exercise group were also provided assessment feedback, information on how exercise may help with symptoms of depression, and a generic exercise prescription reflective of a moderate-intensity walking program. Participants then attended one clinic-based personalized (individual) instructional session at the medical center, during which they were acclimated to walking on a treadmill at various exercise intensities and were instructed on how to use the Rating of Perceived Exertion (RPE) scale to monitor and adjust exercise intensity. Each woman received a pedometer and information on its use for data collection purposes. Members of this group were asked to exercise at home for the duration of the study (home-based exercise only), following the exercise prescription provided. Once each month, these women were contacted to gather pedometer and physical activity recall data. These phone conversations were limited to data collection and brief "Do your best" statements regarding exercise participation. Similar to the clinic-based group, participants in the home-based exercise intervention did not report any adverse events.

Analysis

Descriptive statistics were computed for study variables. To check for preexisting group differences at study entry despite randomization, *t* tests for continuous variables and chi-square analyses for categorical variables were conducted on descriptive variables and outcome measures.

Because some participants missed follow-up assessments and intent-to-treat analyses are the recommended approach, we chose to impute missing values. Mean substitution is the most conservative method of estimating missing data. Replacing the missing value with a group mean (i.e., clinic-based group mean or home-based group mean) is less conservative than using an overall grand mean, but the reduction in within-groups variance can result in spuriously large between-groups differences.[36] A regression approach is more widely applicable and recommended over mean replacement when significant models can be produced. In this method, other variables are used as the independent variables (IV) to predict the missing values or dependent variables (DV). For example, if a 3-month depression score is missing, then baseline depression score, baseline age, ethnicity, and other values are used as predictors (IV) of the 3-month score (DV). Cases with complete values are used to generate the regression equation. The predicted values resulting from the regression equation are imputed for missing values, and all cases are used in a second regression. This continues until the predicted values from one step converge with those of the following step, and the predic-

tions from the last round are used as the estimates of the missing values.[36] The final step is to compare the concordance between these two methods (group mean replacement and regression modeling) of estimation.[36]

These two methods generated estimates that were highly correlated ($r = 0.70$–0.99), and yielded the same result regarding interpretation of statistical significance, with the exception of BMI, percent fat, and cardiovascular fitness data. For those variables, significant findings emerged when using mean replacement but not when using regression modeling estimates. However, regression equations accounted for 99.1%, 96.9%, and 86.9% of the variance in BMI, percent fat, and fitness, respectively. Thus, the regression modeling approach to impute missing values was deemed the more appropriate choice, and reported findings are based on those estimates.

Relationships among variables at study entry and 3-month follow-up were investigated with Pearson's correlations. Overall effect sizes (Cohen's d) for depression for each exercise group were calculated. Gain scores (3 month-baseline) were calculated to demonstrate change in outcomes over time (Table 2). To examine potential group differences at 3-month follow-up on depression, time spent in physical activity, BMI, percent fat, and fitness, analysis of covariance (ANCOVA) was conducted. Baseline scores and three demographic variables (age, ethnicity, and education) served as covariates in these analyses. Finally, pedometers were dispensed with after the baseline assessment; as a result, there are no true baseline pedometer data. However, to investigate group differences in average steps and distance walked per day and to corroborate self-reported activity data at 3-month follow-up, an independent samples t-test was conducted on 3-month pedometer data.

RESULTS

For demographic variables of interest, baseline differences emerged between groups for three variables: ethnicity, education, and sleep difficulty (Table 1). There were no statistically significant differences between groups at baseline on outcome variables of interest: depression, time spent in moderate and vigorous physical activity, BMI, percent body fat, or fitness (all $p > 0.05$).

Means and SDs for dependent variables at baseline and 3-month follow-up are presented in Table 2. Overall ESs for exercise effects on depressive symptoms in this sample were large in both groups (clinic-based -1.3, home-based -0.97). Of the total sample, 46.9% (15 of 32) of participants experienced a ≥50% reduction in depressive symptoms. Further, 31.3% of the sample (10 of 32) achieved remission of their symptoms (BDI score < 9). Gain scores for depressive symptoms (clinic-based mean = -11.7, home-based M = -9.7) and physical activity (clinic-based mean = 65.4, home-based mean = 39.0 min/week) indicated strong improvements across time for both groups (Table 2). Objective monitoring of physical activity by pedometer counts shows that at 3 months, clinic-based group members accumulated more steps and walked further each day (steps: mean = 7036.8, SD = 2259.8; distance: mean = 2.9 miles, SD = 0.9) than those in the home-based group (Steps: mean = 4957.6, SD = 2239.1; distance: mean = 2.1 miles, SD = 1.0). The between-groups difference was significant for both steps and distance (steps: $t(29) = -2.6$, $p < 0.05$; distance: $t(29) = -2.5$, $p < 0.05$).

At baseline and at three months, BMI was significantly correlated with percent body fat ($r = 0.68$ and $r = 0.71$, both $p < 0.001$) and fitness ($r = -0.69$ and $r = 0.70$, both $p < 0.001$). Similarly, percent body fat was inversely correlated with fitness ($r = -0.54$ and $r = -0.53$, both $p < 0.01$). At 3 months, average pedometer steps per day were significantly correlated with BMI ($r = -0.41$, $p < 0.05$), fitness ($r = 0.41$, $p < 0.05$), and time spent in physical activity ($r = 0.46$, $p < 0.01$). Likewise, average distance walked per day, as assessed by the pedometer, was significantly correlated with BMI ($r = -0.41$, $p < 0.05$), fitness ($r = 0.39$, $p < 0.05$), and time spent in activity ($r = 0.45$, $p < 0.05$). All other correlations between depression, time spent in physical activity, fitness, and indicators of body composition at baseline and 3-month follow-up were nonsignificant (all $p > 0.05$).

Based on ANCOVA results, there were no significant differences at 3-month follow-up between groups on BDI-II depression score, after controlling for baseline BDI score and demographic variables [$F(1,30) = 0.01$, $p = 0.93$]. Similarly, groups did not differ significantly on self-reported time spent in moderate and vigorous activity, after controlling for baseline activity level and the demographic variables [$F(1,30) = 1.2$, $p = 0.28$]. ANCOVA for BMI, body fat, and

TABLE 2. MEANS AND SD FOR DEPENDENT VARIABLES AT BASELINE AND 3-MONTH FOLLOW-UP

	Baseline Mean (SD)		*3-Month follow-up Mean (SD)*				*Average gain score based on imputed values*	
			Home-based		*Clinic-based*			
	Home-based	*Clinic-based*	*Actual values*	*Imputed values*	*Actual values*	*Imputed values*	*Home-based*	*Clinic-based*
BDI-II[a]	26.5 (10.0)	24.6 (8.7)	17.7 (11.1)	16.8 (9.7)	15.1 (6.8)	12.9 (7.2)	−9.7	−11.7
BMI	35.6 (13.6)	34.4 (10.3)	39.5 (13.1)	35.4 (13.0)	34.3 (12.1)	34.5 (10.3)	−0.2	−0.1
% Fat	37.9 (9.6)	41.5 (8.1)	41.0 (7.7)	37.9 (9.6)	38.9 (9.3)	40.7 (8.3)	0.0	−0.8
Vo_2 peak (mL/kg/min)	19.8 (7.1)	20.9 (7.6)	19.0 (5.2)	19.4 (5.4)	22.6 (9.0)	21.0 (7.1)	−0.4	0.1
Min/week of moderate and vigorous activity	41.3 (34.4)	40.9 (40.0)	97.5 (100.6)	80.3 (97.2)	143.9 (94.2)	106.3 (87.8)	39.0	65.4
Average pedometer steps/ day at 3-month follow-up			4957.6 (2239.1)	4957.6 (2239.1)	6869.8 (3230.6)	7036.8 (2259.8)		
Average pedometer distance/ day (miles) at 3-month follow-up			2.1 (1.0)	2.1 (1.0)	2.8 (1.3)	2.9 (0.9)		

[a]BDI-II, Beck Depression Inventory-II; BMI, body mass index; Vo_2 peak, peak volume of oxygen consumption.

fitness also revealed no significant between-group differences at 3-month follow-up [BMI: $F(1,30) = 0.29$, $p = 0.59$; percent body fat: $F(1,30) = 1.75$, $p = 0.20$; fitness: $F(1,30) = 0.06$, $p = 0.82$].

DISCUSSION

Participants in both the clinic-based and home-based exercise interventions experienced reductions in their symptoms of depression with preintervention to postintervention ESs ranging from −0.97 to −1.3 for clinic-based and home-based groups, respectively. Both clinic-based and home-based exercise groups similarly improved their self-reported physical activity from baseline to 3-month follow-up, but clinic-based group participants had significantly higher objectively assessed (pedometer) physical activity. We found that pedometer data, but not self-reported physical activity recall, were significantly associated with fitness (V_{O_2}) and body composition (BMI). With the exception of pedometer steps, however, there were no significant between-group effects at 3-month follow-up for depressive symptoms, self-reported physical activity, body composition, or cardiorespiratory fitness.

The findings that the clinic-based, more time-intensive exercise intervention was associated with significantly more objectively assessed physical activity than the home-based condition and that both interventions were associated with increases in time spent in physical activity, but not with improvements in fitness and body composition, warrant further comment. The lack of impact on physical health outcomes in clinic-based participants may be because the leading supervised phase was only 1 month in length. It is interesting to note that during the clinic-based phase, the majority of participants were able to meet the goal of walking 30–45 minutes in a training zone of 60%–85% MHR. Also, whereas clinic-based group members made significant gains in time spent in physical activity, they did not maintain exercise at a duration or intensity necessary to promote positive changes in body composition and fitness once they were transitioned to the home-based phase. Therefore, the observed increase in activity may reflect an incorporation of activity into daily routines rather than an increase in sustained activity intended to increase cardiovascular fitness or promote a reduction in body fat. It is also possible that 3 months is not an adequate length of time to see changes in these specific variables. Continued exercise involvement beyond the 3-month follow-up may cause fitness and body composition improvements.

Similarly, whereas those in the clinic-based group walked significantly more than those in the home-based group (based on pedometer counts), they did not experience significantly larger reductions in the symptoms of their depression. Emerging research suggests a dose-response relationship between exercise and a reduction in depressive symptoms,[10,14] although few studies have directly examined a potential dose-response gradient, and more research in the area is needed. Therefore, although our data do not support a dose-response relationship, our findings are consistent with previous research suggesting that exercise involvement in a nondose-dependent manner is sufficient to reduce depressive symptoms.[7–9,13]

As mentioned previously, in the general population, exercise adherence is improved among those involved in home-based or community-based activities.[21,23,26,27] Therefore, the long-term goal of having women exercise in their communities rather than coming to the hospital to exercise remains an important one. Our findings suggest that a home-based program was as effective in promoting activity involvement and reducing depressive symptoms as a clinic-based intervention. Future researchers should examine whether the initial supervised phase should be lengthened, if additional exercise adherence counseling sessions are required, or if direct supervision in the community for a period of time after the transition is beneficial in promoting compliance.

There are several limitations to the current study that should be considered when interpreting results. First, this was a pilot study, and we employed a small sample size. Consequently, we may not have had adequate statistical power to detect group differences on all outcome measures. Our findings should be replicated in a larger sample. Second, we did not employ a true control group, and, thus, it is impossible to say that the reduction in depressive symptoms observed among the women in this study was caused by exercise. It is plausible that the improvement in depressive symptoms could reflect the passage of time or the effects of other depression treatments. None of the women in the study initiated, changed (dosage of medication,

frequency of therapy), or discontinued either pharmacological or psychological therapies during the trial, and it is impossible to tease out these potential effects. However, our intervention ESs are consistent with the vast majority of results from exercise and depression studies that support exercise as an effective antidepressant. It is also noteworthy that approximately half the sample experienced at least a 50% reduction in their depressive symptoms. Third, although this is the first intervention study of which we are aware that included a large percentage of minority women, the majority of minority women enrolled in this study were of African American descent, and other racial/ethnic groups were underrepresented. Social and cultural factors may differ among various ethnic groups and, as a result, have a differing impact on the course of depression, use of traditional treatments for depression, and interest in exercise as a treatment option. Fourth, we compared only two types of exercise interventions. Programs that include and promote community involvement may be preferred and may be more appropriate for minority women in particular, resulting in better compliance to the exercise prescription.[37]

CONCLUSIONS

Both a home-based and more intensive structured (clinic-based) exercise intervention were associated with improvements in time spent in moderate and vigorous physical activity and a reduction in depressive symptoms at 3-month follow-up. This suggests that basic information and contact, as was provided to the home-based group, can promote the initiation of an exercise program. An encouraging implication of this finding may be that even a minimal intervention can prompt women with depressive symptoms to initiate an exercise program. Home-based participants were provided one primary exercise session, a pedometer, and a monthly call. These efforts required minimal time and may be enough to promote exercise at a sufficient level to reduce depression, although these are likely significant enhancements to the usual standard care. Future research is needed to determine what types of interventions will promote long-term compliance at an appropriate frequency, duration, and intensity that would also result in positive changes in body composition and fitness.

REFERENCES

1. Chapman DP, Perry GS, Strine TW. The vital link between chronic disease and depressive disorders. Preventing Chronic Dis 2005. Available at *www.cdc.gov/pcd/issues/2005/jan/04_0066.htm*
2. Ferketich AK, Frid DJ. Depression and coronary heart disease: A review of the literature. Clin Geriatr 2001;9:1.
3. Massand PS. Tolerability and adherence issues in antidepressant therapy. Clin Ther 2003;23:2289.
4. Nemeroff CB, Musselman DL, Evans DL. Depression and cardiac disease. Depress Anxiety 1998;8:71.
5. Rugulles R. Depression as a predictor for coronary heart disease. A review and meta-analysis. Am J Prev Med 2002;23:51.
6. Greenberg PE, Kessler RC, Birnbaum HG, et al. The economic burden of depression in the United States: How did it change between 1990 and 2000? J Clin Psychiatry 2003;64:1465.
7. Blumenthal JA, Babyak MA, Moore KA, et al. Effects of exercise training on older patients with major depression. Arch Intern Med 1999;159:2349.
8. Craft LL. Exercise and clinical depression: Examining two psychological mechanisms. Psychol Sport Exerc 2005;6:151.
9. Dimeo F, Bauer M, Varahran I, Proest G, Halter U. Benefits from aerobic exercise in patients with major depression: A pilot study. Br J Sports Med 2001;35:114.
10. Dunn AL, Trivedi MH, Kampert JB, Clark CG, Chambliss HO. Exercise treatment for depression: Efficacy and dose response. Am J Prev Med 2005;28:1.
11. Dunn AL, Trivedi MH, O'Neal HA. Physical activity dose-response effects on outcomes of depression and anxiety. Med Sci Sports Exerc 2001;33:S587.
12. Dunn AL, Trivedi MH, Kampert JB, Clark CG, Chambliss HO. The DOSEstudy: A clinical trial to examine efficacy and dose response of exercise as treatment for depression. Control Clin Trials 2002;23:584.
13. McNeil, JK, LeBlanc EM, Joyner M. The effect of exercise on depressive symptoms in the moderately depressed elderly. Psychol Aging 1991;6:487.
14. Singh NA, Stavrinos TM, Scarbek Y, Galambos G, Liber C, Fiatarone-Singh MA. A randomized controlled trial of high versus low intensity weight training versus general practitioner care for clinical depression in older adults. J Gerontol A Biol Sci Med Sci 2005;60:768.
15. Craft LL, Landers DM. The effect of exercise on clinical depression and depression resulting from mental illness: A meta-analysis. J Sport Exerc Psychol 1998;20: 339.
16. Lawlor DA, Hopker SW. The effectiveness of exercise as an intervention in the management of depression: Systematic review and meta-regression analysis of randomized controlled trials. BMJ 2001;322:763.
17. North TC, McCullagh P, Tran ZV. Effects of exercise on depression. Exerc Sport Sci Rev 1990;18:379.
18. Stathopoulou G, Powers MB, Berry AC, Smits JAJ, Otto MW. Exercise interventions for mental health: A

quantitative and qualitative review. Clin Psychol Sci Pract 2006;13:179.
19. Buckworth J, Wallace LS. Application of the transtheoretical model to physically active adults. J Sports Med Phys Fitness 2002;42:360.
20. Duncan KA, Pozehl B. Staying on course: The effects of an adherence facilitation intervention on home exercise participation. Prog Cardiovasc Nurs 2002;17:59.
21. Fisher KJ, Fuzhong L. A community-based walking trial to improve neighborhood quality of life in older adults: A multilevel analysis. Ann Behav Med 2004; 28:186.
22. King AL, Taylor CB, Haskell WL, et al. Strategies for increasing early adherence to and long-term maintenance of home-based exercise training in healthy middle-aged men and women. Am J Cardiol 1998;61: 628.
23. King AC, Kiernan M, Oman RF, Kraemer HC, Hull M, Ahn D. Can we identify who will adhere to long-term physical activity? Signal detection methodology as a potential aid to clinical decision making. Health Psychol 1997;16:380.
24. Lee RE, Goldberg JH, Sallis JF, Hickmann SA, Castro CM, Chen AH. A prospective analysis of the relationship between walking and mood in sedentary ethnic minority women. Women Health 2001;32:1.
25. Lombard DN, Lombard TN, Winett RA. Walking to meet health guidelines: The effect of prompting frequency and prompt structure. Health Psychol 1995; 14:164.
26. Ransdell LB, Taylor A, Oakland D, Schmidt J, Moyer-Mileur L, Shultz B. Daughters and mothers exercising together: Effects of home-and community-based programs. Med Sci Sports Exerc 2003;35:286.
27. Wilbur J, Vassalo A, Chandler P, McDevitt J, Miller AM. Midlife women's adherence to home-based walking during maintenance. Nurs Res 2005;54: 33.
28. Writing Group for the Activity Counseling Trial Research Group. Effects of physical activity counseling in primary care: The ACT: A randomized controlled trial. JAMA 2001;286:677.
29. Beck AT, Steer RA, Brown K. Beck Depression Inventory manual, 2nd ed. San Antonio,TX: Harcourt Brace, 1996.
30. Lohman TG. Dual energy radiography: Total body and regional composition. In: Lohman, TG, ed. Advancement in body composition assessment. Champaign, IL: Human Kinetics Publishers, 1992:25.
31. Blair SN, Haskell WL, Ho P, et al. Assessment of habitual activity by seven-day recall in a community survey and controlled experiments. Am J Epidemiol 1985;122:794.
32. Sallis JF, Haskell W, Wood P, et al. Physical activity assessment methodology in the Five-City Project. Am J Epidemiol 1985;121:91.
33. Hayden-Wade HA, Coleman KJ, Sallis JF, Armstrong C. Validation of the telephone and in-person interview versions of the 7-Day PAR. Med Sci Sports Exerc 2003;35:801.
34. Schneider PL, Crouter SE, Bassett DR. Pedometer measures of free-living physical activity: Comparison of 13 models. Med Sci Sports Exerc 2004;36:331.
35. Crouter SE, Schneider PL, Karabulut M, Bassett DR. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. Med Sci Sports Exerc 2003;35:1455.
36. Tabachnick BG, Fidell LS. Using multivariate statistics, 4th ed. Boston, MA: Allyn and Bacon, 2001:63.
37. Taylor WC, Baranowski T, Young DR. Physical activity interventions in low-income, ethnic minority, and populations with disability. Am J Prev Med 1998;15:334.

Addrepss reprint requests to:
Lynette L. Craft, Ph.D.
Department of Preventive Medicine
Feinberg School of Medicine
Northwestern University
680 N. Lake Shore Drive, Suite 1220
Chicago, IL 60611

E-mail: lynette-craft@northwestern.edu

J Periodont Res 2007; 42; 169–176

JOURNAL OF PERIODONTAL RESEARCH
doi:10.1111/j.1600-0765.2006.00930.x

Oral health-related quality of life of periodontal patients

J. Cunha-Cruz[1], P. P Hujoel[1], N. R. Kressin[2,3]

[1]Department of Dental Public Health Sciences, University of Washington, Seattle, USA, [2]Department of Veterans Affairs, Bedford, MA, USA and [3]Boston University School of Public Health, MA, USA

Cunha-Cruz J, Hujoel PP, Kressin NR. Oral health-related quality of life of periodontal patients. J Periodont Res 2007; 42: 169–176. © 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

Background and Objective: The purpose of this study was to assess the oral health-related quality of life of patients presenting to a periodontal specialist by means of six questions, and to assess the perceived oral health by means of one question. Self-assessments of oral health were associated with clinical characteristics.

Material and Methods: Logistic regression models were used to associate self-assessments with clinical characteristics in a cross-sectional study.

Results: On the six-item questionnaire, close to 20% (295/1480) of the patients reported that teeth, gums or dentures had an impact fairly often or very often on one or more items (eating, relaxing, avoiding going out, feeling self-conscious, pain or discomfort). On the single question requesting a self-assessment of oral health, 42% (628/1468) rated their oral health as fair or poor. Both common oral health-related quality of life problems and worse perceived oral health were associated with having more than eight teeth with >5 mm periodontal pockets (odds ratio = 1.45, 95% confidence interval = 1.01–2.08; and odds ratio = 2.83, 95% confidence interval = 2.08–3.84, respectively), compared with patients who had fewer than three teeth with >5 mm periodontal pockets.

Conclusion: Oral health-related problems in patients presenting to a periodontal specialist office negatively affect their quality of life. If some of the findings of this study can be confirmed in other studies, it could change the perception of chronic periodontitis as a silent disease.

Joana Cunha-Cruz/Philippe Hujoel, University of Washington, Dental Public Health Sciences, B509, 1959 NE Pacific Street, Box 357475, Seattle, WA 98195–7475, USA
Tel: +1 206 5432034
Fax: +1 206 6854258
e-mail: hujoel@u.washington.edu

Key words: periodontal diseases; quality of life; tooth loss; self-assessment

Accepted for publication July 5, 2006

Chronic periodontitis has been reported to be asymptomatic during most of its course, with clinical signs, such as periodontal pockets, being generally painless (1). Because chronic periodontitis is believed to be asymptomatic in its initial stages, it has been suggested that individuals may be unaware of their clinical periodontal status (2–4) and underestimate what treatments are required, as judged by dental professionals (5). In its more advanced stages, chronic periodontitis can be associated with signs and symptoms that are readily perceivable by individuals, such as tooth mobility, pain, eating difficulties, unesthetic loss of anterior interproximal papillae, or discomfort (6,7).

Various tools and methods have been developed to assess the impact of dental diseases on the oral health-related quality of life as perceived by the affected individual. Two such distinct methods include a self-report of oral health, typically assessed by a single question (subsequently referred to as the perceived oral health), and a self-report on oral symptoms and functions as assessed by a battery of questions (subsequently referred to as oral health-related quality of life). Surveys in the community and clinical settings have used these measures to describe the oral health from an individual perspective (7–15). Improvements on both the way individuals rate their oral health and the oral health-related quality of life of patients under treatment are desirable outcomes of dental treatments (16). Such measures are increasingly being used in clinical trials as subjective true end points (17–22).

In terms of these patient-oriented outcomes, little is known about their frequency among periodontal patients or their relationship to clinical signs of periodontal diseases. The aim of this study was to describe the perceived oral health, the oral health-related quality of life and the association of

these two subjective assessments with clinical characteristics among patients presenting to a periodontal specialist.

Subjects and methods

The study population consisted of members of the Washington Dental Service presenting for a comprehensive initial clinical examination by a periodontal specialist. A total of 3617 individuals were invited to participate in the study; 1497 patients consented to participate and mailed back the questionnaire during the period from February 2003 to October 2004 (41% response rate). The protocol of the study was approved by the Institutional Review Board of the University of Washington.

Patients' characteristics were obtained from a mailed questionnaire, which included questions on age, smoking, diabetes status, perceived oral health (one question) and oral health-related quality of life (six questions). Gender was obtained from administrative data. Number of teeth with at least one periodontal pocket deeper than 5 mm and 8 mm, and number of missing teeth were abstracted from the dental charts.

Perceived oral health was measured by a single question 'How would you describe the health of your teeth and gums?' where the possible answers were: excellent, very good, good, fair and poor. Patients reporting fair or poor oral health were considered as having worse perceived oral health.

Oral health-related quality of life was measured by a short oral health-related quality of life questionnaire consisting of six questions about the impact of teeth, gums and dentures on eating, relaxing, avoiding going out, feeling self-conscious or worried, pain and denture discomfort (Table 1). Patients who answered either fairly often or very often; always; or either quite a bit or a great deal to one or more of the six questions were considered as having common oral health-related quality of life problems.

The oral health-related quality of life questionnaire was developed by Kressin and colleagues (unpublished), based on a conceptual model of oral health and quality of life which posited that the dimensions of physical function, social role function, distress, worry, denture discomfort and impairment/disease were the six most important areas that could be impacted by decrements in oral health. They tested this model empirically, using data previously collected from two samples of older male veterans (total $n = 816$), which included three oral health-related quality of life indices – the Oral Health Impact Profile (23), the Geriatric Oral Health Assessment Index (24) and the Oral Health-related Quality of Life measure (25). They assigned each item from the three scales to the dimension it best represented, and then, using factor analysis and multitrait analyses, pared down the number of items by eliminating items whose deletion least affected the internal consistency reliability of the scales. At the same time, they sought to retain the items which they considered, from a conceptual standpoint, to best represent the subscale. Ultimately, the brief oral health-related quality of life questionnaire consisted of a six-item measure representing six oral health-related quality of life dimensions in which four of these items were from the Oral Health Impact Profile questionnaire; one was from the Geriatric Oral Health Assessment Index; and one was from the Oral Health-related Quality of Life measure (Table 1). Reliability and validity of the new questionnaire were tested and the brief oral health-related quality of life questionnaire presented good internal consistency (Cronbach's alpha = 0.80) and convergent validity, as measured by its overall correlation with number of teeth, coronal and root dental caries, and periodontal status from the two samples of older male veterans.

Statistical analysis

The distribution of perceived oral health and each impact (i.e. each item) of the oral health-related quality of life instrument were examined using descriptive statistics. In the main analysis, logistic regression models were used to relate both worse perceived oral health and common oral health-related quality of life problems to number of teeth with pockets deeper than 5 mm (0–2, 3–4, 5–8, 9–30 teeth), number of teeth with pockets deeper than 8 mm (0, 1–2, 3–19 teeth) and number of missing teeth (0–3, 4–7, 8–11, 12–31 teeth). Subgroup analyses were performed for anterior and posterior teeth (see Table 2 for categories). In a secondary analysis, logistic regression was used to relate each impact (item) of the oral health-related quality of life questionnaire with the clinical characteristics of

Table 1. Development of a short-form oral health-related quality of life (OHQoL) questionnaire

Item	Scale	Original instrument	OHQoL dimension
During the past 3 mo, how often have you experienced the following difficulties because of problems with your teeth, mouth or dentures?			
Have you had to avoid eating some foods?	Never, Hardly ever, Occasionally, Fairly often, Very often	OHIP	Physical
Have you found it difficult to relax?	Never, Hardly ever, Occasionally, Fairly often, Very often	OHIP	Distress
Have you avoided going out?	Never, Hardly ever, Occasionally, Fairly often, Very often	OHIP	Social role
Have you felt nervous or self-conscious?	Never, Sometimes, Always	GOHAI	Worry
How much pain or distress have your teeth or gums caused you?	None at all, A little bit, Some, Quite a bit, A great deal	OHQoL	Impairment
Have you had uncomfortable dentures?	Never, Hardly ever, Occasionally, Fairly often, Very often	OHIP	Denture discomfort

GOHAI, Geriatric Oral Health Assessment Index; OHIP, Oral Health Impact Profile; OHQoL, oral health-related quality of life.

the patients. Given the exploratory nature of this latter analysis, we report only those results where the lower limit of the confidence interval is ≥2. In addition to the clinical characteristics, the multivariate models included age groups (35 to < 49, 49 to < 54, 54 to < 60, 60–89 years old), gender (male, female), smoking status (never, former, current), diabetes status (yes, no) and partial denture use (yes, no).

Results

Participants presenting for an initial periodontal examination and consenting to participate were 35–89 years old (mean = 54.8; standard deviation = 8.2); 51.2% were women; 9.5% reported having diabetes; and 21.8% and 42.9% were current and former smokers, respectively. Fourteen per cent of the patients reported using some type of removable dentures. In general, patients had 6.1 teeth with periodontal pockets deeper than 5 mm (standard deviation = 5.6) [1.2 (standard deviation = 2.3) in the anterior teeth and 4.9 (standard deviation = 3.9) in the posterior teeth]. The mean number of teeth with periodontal pockets deeper than 8 mm was 0.7 (standard deviation = 1.4) [0.1 (standard deviation = 0.5) in the anterior teeth and 0.6 (standard deviation = 1.2) in the posterior teeth]. On average, patients had 7.1 missing teeth (standard deviation = 4.8) [0.8 missing anterior teeth (standard deviation = 1.9) and 6.1 missing posterior teeth (standard deviation = 3.0)].

Perceived oral health

Only 1.9% of the patients rated their oral health as excellent; 14.1% rated their oral health as very good, 40.2% as good, 34.5% as fair and 7.4% as poor (28 participants did not answer this question). A total of 41.9% of the sample indicated that their oral health was in the 'worst' category (fair or poor).

Periodontal pockets > 5 mm and worse perceived oral health — The unadjusted odds of worse perceived oral health increased by 19% for patients with 3–4 teeth that had pockets > 5 mm (95% confidence interval = 0.86–1.66); by 40% for patients with 5–8 teeth that had pockets > 5 mm (95% confidence interval = 1.03–1.90); and by 183% for those with ≥9 teeth that had pockets > 5 mm (95% confidence interval = 2.08–3.84), compared with patients who had 0–2 teeth with pockets > 5 mm. After adjustment, only the odds of worse perceived oral health for patients who had ≥9 teeth with pockets > 5 mm remained significantly different from patients who had 0–2 teeth with pockets > 5 mm. When restricting the analyses to anterior or posterior teeth, worse per-

Table 2. Association of worse perceived oral health (reported as fair or poor self-assessed oral health) with clinical characteristics of periodontal patients presenting for an initial examination

	Worse perceived oral health[a]	
	Crude OR (95% CI)	Adjusted OR[b] (95% CI)
No. of teeth PD > 5 mm		
0–2 teeth (reference)		
3–4 teeth	1.19 (0.86 1.66)	1.09 (0.77 1.53)
5–8 teeth	1.40 (1.03 1.90)	1.21 (0.88 1.67)
9 or more teeth	2.83 (2.08 3.84)	2.78 (2.00 3.87)
No. of teeth PD > 8 mm		
0 tooth (reference)		
1–2 tooth	1.25 (0.96 1.62)	1.10 (0.83 1.45)
3 or more teeth	3.28 (2.10 5.14)	3.18 (2.00 5.05)
No. of missing teeth		
0–3 teeth (reference)		
4–7 teeth	1.53 (1.18 1.99)	1.73 (1.31 2.28)
8–11 teeth	2.08 (1.46 2.96)	2.45 (1.66 3.61)
12 or more teeth	1.92 (1.27 2.90)	2.61 (1.65 4.13)
Subgroup analyses		
No. of anterior teeth PD > 5 mm		
0 tooth (reference)		
1–4 teeth	1.48 (1.14 1.91)	1.34 (1.03 1.74)
5 or more teeth	4.47 (2.95 6.78)	3.73 (2.43 5.73)
No. of posterior teeth PD > 5 mm		
0–2 teeth (reference)		
3–8 teeth	1.53 (1.18 1.97)	1.39 (1.07 1.82)
9 or more teeth	2.61 (1.89 3.62)	2.62 (1.84 3.73)
No. of anterior teeth PD > 8 mm		
0 tooth (reference)		
1–2 tooth	1.70 (1.03 2.82)	1.60 (0.95 2.69)
3 or more teeth	4.23 (1.14 15.69)	3.30 (0.86 12.60)
No. of posterior teeth PD > 8 mm		
0 tooth (reference)		
1–2 tooth	1.25 (0.96 1.63)	1.09 (0.83 1.44)
3 or more teeth	3.33 (2.04 5.45)	3.21 (1.93 5.32)
No. of anterior missing teeth		
0 tooth		
1 teeth	1.08 (0.76 1.54)	1.12 (0.77 1.62)
2–4 teeth	1.12 (0.73 1.71)	1.02 (0.65 1.61)
4 or more teeth	1.31 (0.89 1.95)	1.34 (0.88 2.04)
No. of posterior missing teeth		
0–5 teeth (reference)		
6–8 teeth	1.39 (1.09 1.79)	1.52 (1.17 1.98)
9–10 teeth	2.30 (1.51 3.52)	2.51 (1.61 3.93)
11 or more teeth	1.58 (1.05 2.40)	2.03 (1.27 3.23)

CI, confidence interval; OR, odds ratio; PD, pocket depth.

[a]Patients who described a fair or poor health of their teeth and gums as opposed to excellent, very good and good.

[b]Logistic model included age (35 to < 49, 49 to < 54, 54 to < 60 and 60–89 years), gender, smoking (never, former, current smoker), diabetes, partial denture use, number of teeth with pockets deeper than 5 mm and number of missing teeth.

ceived oral health was associated with both anterior and posterior teeth with pockets > 5 mm (Table 2).

Periodontal pockets > 8 mm and worse perceived oral health — The unadjusted odds for worse perceived oral health increased by 228% for patients who had ≥3 teeth with pockets deeper than 8 mm compared to patients without pockets > 8 mm (odds ratio = 3.28; 95% confidence interval = 2.10–5.14). After adjustment, this association remained statistically significant. When restricting the analyses to anterior or posterior teeth, posterior teeth with pockets > 8 mm was associated with worse perceived oral health (Table 2).

Missing teeth and worse perceived oral health — When compared with patients who had 0–3 missing teeth, having 4–7, 8–11 or ≥12 missing teeth increased the unadjusted odds of worse perceived oral health by 53% (95% confidence interval = 1.18–1.99), 108% (95% confidence interval = 1.46–2.96) or 92% (95% confidence interval = 1.27–2.90), respectively. After adjustment for confounding, all levels of missing teeth (4–7, 8–11 and ≥12) remained significantly associated with worse perceived oral health. When restricting the analyses to anterior or posterior teeth, the number of posterior missing teeth was associated with worse perceived oral health (Table 2).

Common oral health-related quality of life problems: one or more out of six problems with a poor rating

Almost 20% of the patients had one or more items endorsed as having problems fairly often or very often with eating, relaxing, avoiding going out, feeling self-conscious, denture discomfort or pain caused by teeth, gums and dentures (i.e. common oral health-related quality of life problems) (Table 3). Of these patients, 11.1% reported one problem, 5.7% reported two problems, 1.8% reported three problems and 1.2% reported four to six problems.

Periodontal pockets > 5 mm and common oral health-related quality of life problems — Having ≥9 teeth with pockets deeper than 5 mm increased by 45% the unadjusted odds of common oral health-related quality of life problems compared with patients who had 0–2 teeth with pockets > 5 mm (odds ratio = 1.45; 95% confidence interval = 1.01–2.08). Compared with patients who had 0–2 teeth with pockets > 5 mm, the unadjusted odds of common oral health-related quality of life problems were not statistically different between patients who had 3–4 teeth with pockets > 5 mm (odds ratio = 1.00; 95% confidence interval = 0.67–1.50) or 5–8 teeth with pockets > 5 mm (odds ratio = 0.86; 95% confidence interval = 0.59–1.27). Adjustment for confounding increased the magnitude of the association between ≥9 teeth with pockets > 5 mm and frequent oral health-related quality of life (odds ratio = 1.59; 95% confidence interval = 1.07–2.35). The number of anterior and posterior teeth with pockets > 5 mm were also associated with common oral health-related quality of life problems (Table 4).

Periodontal pockets > 8 mm and common oral health-related quality of life problems — Compared with no pockets > 8 mm, the unadjusted odds of common oral health-related quality of life problems increased by 67% for patients who had ≥3 teeth with pockets > 8 mm (odds ratio = 1.67; 95% confidence interval = 1.04–2.68). After adjustment for confounding, ≥3 teeth with pockets > 8 mm remained associated with common oral health-related quality of life problems (Table 4). The number of posterior teeth with pockets > 8 mm was also associated with common oral health-related quality of life problems (Table 4).

Missing teeth and common oral health-related quality of life problems — Having 8–11 missing teeth increased by 106% the unadjusted odds of common oral health-related quality of life problems compared with patients who had 0–3 missing teeth (odds ratio = 2.06; 95% confidence interval = 1.38–

Table 3. Impact of oral health on the quality of life of periodontal patients

	OHQoL impacts[a]						
	Eating $n = 1476$	Relaxing $n = 1474$	Avoiding going out $n = 1476$	Self- conscious[b] $n = 1474$	Pain[c] $n = 1475$	Denture discomfort[d] $n = 209$	Common OHQoL problems[e] $n = 1480$
Intensity (%)							
Never	57.0	53.7	82.5	70.0	34.2	52.6	
Hardly ever	19.0	19.4	11.4		35.6	15.3	
Occasionally	18.0	18.7	4.4	25.3	19.6	16.8	
Fairly often	3.9	5.6	1.1		7.5	8.1	
Very often	2.2	2.5	0.7	4.8	3.1	7.2	
Fairly or very often (%)	6.0	8.1	1.8	4.8	10.6	15.3	19.9

[a]Patients were asked to report how often during the past 3 months they had experienced difficulties because of problems with their teeth, mouth or dentures.
[b]Frequencies of the impact 'feeling self-conscious (or nervous)' included never, sometimes and always.
[c]Frequencies of the impact 'feeling pain (or distress)' included none at all, a little bit, some, quite a bit and a great deal.
[d]Patients with at least one partial denture.
[e]Patients with at least one endorsement of a fairly often or very often impact of teeth, gums or dentures on eating, relaxing, avoiding going out, feeling self-conscious, pain or denture discomfort were considered as having common oral health-related quality of life (OHQoL) impact.

3.09). The odds of common oral health-related quality of life problems were not statistically different between patients with 4–7 (odds ratio = 1.01; 95% confidence interval = 0.72–1.41) or ≥12 (odds ratio = 1.48; 95% confidence interval = 0.91–2.43) missing teeth when compared with patients who had 0–3 missing teeth. After adjustment for confounding, the association between having 8–11 missing teeth and common oral health-related quality of life problems remained statistically significant. When restricting the analyses to anterior or posterior teeth, the number of posterior missing teeth was associated with common oral health-related quality of life problems (Table 4).

Table 4. Association of common oral health-related quality of life (OHQoL) problems (at least one item endorsed as having problems fairly or very often with eating, relaxing, avoiding going out, feeling self-conscious, denture discomfort or pain caused by teeth, gums and dentures) with clinical characteristics of periodontal patients

	Common OHQoL problems[a]	
	Crude OR (95% CI)	Adjusted OR[b] (95% CI)
No. of teeth PD > 5 mm		
0–2 teeth (reference)		
3–4 teeth	1.00 (0.67 1.50)	1.04 (0.68 1.58)
5–8 teeth	0.86 (0.59 1.27)	0.88 (0.58 1.32)
9 or more teeth	1.45 (1.01 2.08)	1.59 (1.07 2.35)
No. of teeth PD > 8 mm		
0 tooth (reference)		
1–2 tooth	1.00 (0.72 1.40)	1.00 (0.70 1.41)
3 or more teeth	1.67 (1.04 2.68)	1.80 (1.09 2.98)
No. of missing teeth		
0–3 teeth (reference)		
4–7 teeth	1.01 (0.72 1.41)	1.07 (0.76 1.51)
8–11 teeth	2.06 (1.38 3.09)	2.32 (1.49 3.61)
12 or more teeth	1.48 (0.91 2.43)	1.54 (0.89 2.65)
Subgroup analyses		
No. of anterior teeth PD > 5 mm		
0 tooth (reference)		
1–4 teeth	1.21 (0.88 1.66)	1.18 (0.85 1.64)
5 or more teeth	1.87 (1.22 2.87)	1.70 (1.08 2.67)
No. of posterior teeth PD > 5 mm		
0–2 teeth (reference)		
3–8 teeth	0.95 (0.69 1.31)	0.99 (0.71 1.38)
9 or more teeth	1.46 (1.00 2.13)	1.63 (1.08 2.48)
No. of anterior teeth PD > 8 mm		
0 tooth (reference)		
1–2 tooth	1.28 (0.70 2.31)	1.28 (0.70 2.37)
3 or more teeth	3.04 (0.96 9.65)	2.22 (0.67 7.40)
No. of posterior teeth PD > 8 mm		
0 tooth (reference)		
1–2 tooth	1.00 (0.72 1.40)	0.99 (0.70 1.40)
3 or more teeth	1.66 (0.99 2.79)	1.76 (1.02 3.04)
No. of anterior missing teeth		
0 tooth		
1 teeth	1.06 (0.68 1.66)	0.99 (0.63 1.58)
2–4 teeth	1.32 (0.79 2.18)	1.24 (0.73 2.11)
4 or more teeth	1.30 (0.82 2.08)	1.22 (0.74 2.03)
No. of posterior missing teeth		
0–5 teeth (reference)		
6–8 teeth	1.00 (0.73 1.38)	1.07 (0.76 1.49)
9–10 teeth	2.11 (1.33 3.35)	2.16 (1.32 3.53)
11 or more teeth	1.75 (1.09 2.82)	1.84 (1.07 3.16)

CI, confidence interval; OR, odds ratio; PD, pocket depth.
[a]At least one endorsement of a fairly often or very often impact of teeth, gums or dentures on eating, relaxing, avoiding going out, feeling self-conscious, pain or denture discomfort.
[b]Logistic model included age (35 to < 49, 49 to < 54, 54 to < 60 and 60–89 years), gender, smoking (never, former, current smoker), diabetes, partial denture use, number of teeth with pockets deeper than 5 mm and number of missing teeth.

Specific oral health-related quality of life impacts: exploratory analyses

Pain and difficulty in relaxing were the most frequent impacts, with prevalences of 10.6% and 8.1%, respectively. Six per cent of the periodontal patients reported having eating problems fairly or very often, 4.8% reported feeling always self-conscious and 1.8% reported fairly or very often avoiding going out because of oral problems. Among patients with dentures, 15.3% of the patients reported discomfort fairly or very often (Table 3).

Periodontal pockets > 5 mm and specific oral health-related quality of life impacts — None of the lower limits of the odds ratio's confidence intervals for the association between periodontal pockets > 5 mm and each oral health-related quality of life impact on eating, avoiding going out, relaxing, feeling self-conscious, pain and denture discomfort were ≥2.

Periodontal pockets > 8 mm and specific oral health-related quality of life impacts — Having ≥3 anterior teeth with pockets > 8 mm was associated with feeling pain frequently when compared to no anterior teeth with pockets > 8 mm (odds ratio = 6.43; 95% confidence interval = 2.01–20.57). This association remained significant after adjustment for confounding (Table 5).

Missing teeth and specific oral health-related quality of life impacts — Patients with 8–11 and ≥12 missing teeth had an increased unadjusted odds of reporting frequent self-consciousness as a result of teeth, gums and dentures when compared to patients who had 0–3 missing teeth (odds ratio = 4.70, 95% confidence interval = 2.05–10.75; and odds ratio = 4.25, 95% confidence interval = 1.68–10.72, respectively). After adjustment, 8–11 and ≥12 missing teeth remained associated with feeling self-conscious

Table 5. Exploratory analyses of the association of each oral health-related quality of life (OHQoL) impact (reported fairly or very-often occurrence of problems on eating, relaxing, avoiding going out, feeling self-conscious, denture discomfort or pain caused by teeth, gums and dentures) with clinical characteristics of periodontal patients: only results where the lower limit of the confidence interval is ≥2 are reported

	Crude OR (95% CI)	Adjusted OR[a] (95% CI)
Self-conscious		
No. of missing teeth		
0–3 teeth (reference)		
4–7 teeth	1.89 (0.86 4.12)	2.12 (0.96 4.70)
8–11 teeth	4.70 (2.05 10.75)	6.28 (2.58 15.31)
12 or more teeth	4.25 (1.68 10.72)	5.24 (1.91 14.37)
No. of posterior missing teeth		
0–5 teeth (reference)		
6–8 teeth	1.64 (0.86 3.11)	1.87 (0.97 3.63)
9–10 teeth	2.86 (1.22 6.67)	3.17 (1.30 7.73)
11 or more teeth	4.06 (1.88 8.74)	5.20 (2.14 12.63)
Pain		
No. of teeth PD > 8 mm – anterior		
0 tooth (reference)		
1–2 tooth	1.29 (0.60 2.76)	1.28 (0.59 2.80)
3 or more teeth	6.43 (2.01 20.57)	5.19 (1.53 17.66)
Denture discomfort		
No. of missing teeth		
0–3 teeth (reference)		
4–7 teeth	1.82 (0.35 9.45)	1.69 (0.27 10.75)
8–11 teeth	9.57 (2.01 45.51)	2.80 (0.47 16.69)
12 or more teeth	17.38 (3.70 81.61)	3.64 (0.62 21.53)
No. of anterior missing teeth		
0 tooth (reference)		
1 teeth	5.41 (1.43 20.39)	2.15 (0.49 9.42)
2–4 teeth	19.53 (6.40 59.57)	6.2 (1.74 22.05)
4 or more teeth	10.66 (3.20 35.54)	2.09 (0.56 7.84)
No. of posterior missing teeth		
0–5 teeth (reference)		
6–8 teeth	1.71 (0.55 5.33)	1.45 (0.37 5.61)
9–10 teeth	2.29 (0.46 11.50)	0.68 (0.12 4.03)
11 or more teeth	11.08 (3.86 31.84)	2.48 (0.61 10.16)

CI, confidence interval; OR, odds ratio; PD, pocket depth.
[a]Logistic model included age (35 to < 49, 49 to < 54, 54 to < 60 and 60–89 years), gender, smoking (never, former, current smoker), diabetes, partial denture use, number of teeth with pockets deeper than 5 mm and number of missing teeth.

frequently. When restricting to anterior or posterior missing teeth, posterior missing teeth was associated with frequent self-consciousness (Table 5).

Missing teeth was associated with denture discomfort. Patients with 8–11 and ≥12 missing teeth had higher odds of frequent denture discomfort than patients with 0–3 missing teeth (odds ratio = 9.57, 95% confidence interval = 2.01–45.51; and odds ratio = 17.38, 95% confidence interval = 3.70–81.61, respectively). These associations were not statistically significant after adjustment for confounding. When restricting to anterior and posterior missing teeth, both anterior and posterior missing teeth were associated with frequent denture discomfort (Table 5).

Discussion

The findings of this study indicate that 42% of the patients reported either fair or poor oral health when a simple question was asked with regard to the perception of the conditions of their teeth and gums. In addition, one in five periodontal patients reported frequent episodes of one or more oral health-related quality of life problems when six questions related to their teeth, gums and dentures were posed. The relationships of periodontal pockets and missing teeth with both perceived oral health and oral health-related quality of life self-assessments were not simple linear associations.

Many patients were not satisfied with the health of their teeth and gums; almost half rated their oral health as fair or poor. It is important to note that these periodontal patients are not necessarily rating their oral health worse than the general population. Our findings were similar to national United States estimates, where 36% and 44% of adults 40–64 years and > 65 years rated their oral health as fair or poor, respectively (15).

About one-fifth of this population of periodontal patients reported one or more frequent adverse impacts in their quality of life caused by teeth, gums or dentures. This figure is similar to estimates from national surveys in the UK and Australia, where 16% and 18% of adults reported at least one oral health-related quality of life adverse impact occurring very or fairly often (26). The three most commonly reported problems were pain, difficulty relaxing and denture discomfort. These findings suggest that not only the physical functioning, but also pleasurable life experiences, such as relaxation and social interaction, can be affected by the oral conditions of periodontal patients.

There was not a simple linear relationship between the number of deep pockets, the number of missing teeth and the self-assessed oral health measures. For periodontal pockets, oral health-related quality of life and perceived oral health was only affected when multiple teeth were involved (generalized periodontitis?). In contrast, the presence of a few periodontal pockets (localized periodontitis?) did not influence the oral health-related quality of life or the perceived oral health. In a study in England, oral health-related quality of life was found to be linearly associated with periodontal pockets 5 mm or deeper of patients either at an initial examination or during the maintenance phase of the periodontal treatment (7), a finding we failed to duplicate.

The findings that a few periodontal pockets were not related to common oral health-related quality of life

problems and moderately related to worse perceived oral health suggest that, even though patients being referred to a periodontal specialist may be aware of their periodontal disease status, the consequences of a few periodontal pockets on the oral health-related quality of life of these patients are likely to be small.

In contrast with localized chronic periodontitis, patients with generalized forms of chronic periodontitis may be more likely to have noticeable signs and symptoms, such as tooth mobility and unaesthetic loss of anterior interproximal papillae, which may be driving the observed poor oral health-related quality of life and worse perceived oral health. These findings suggest that chronic periodontitis may interfere with the social life of periodontitis patients and challenge the perception of chronic periodontitis as a silent disease. In addition, the lack of linear correspondence between the current objective measure of periodontal diseases (i.e. periodontal pockets) and oral health-related quality of life or perceived oral health may encourage clinicians and investigators to make use of outcomes more meaningful to their patients by adding together traditional clinical indicators and subjective indicators to assess periodontal needs and evaluate treatments in the periodontal practice.

Number of missing teeth was another clinical characteristic of the periodontal patients associated with both oral health-related quality of life and worse perceived oral health. This relationship was more complex. When people were missing less than one-quarter of their dentition, quality of life was unaffected. Similarly, when people were missing more than one-third of their dentition, quality of life was not substantially affected. It is only within a relatively narrow range of missing teeth (one-quarter to one-third missing) that oral health-related quality of life seems to be affected, as compared to the people with a complete dentition or missing up to 3 teeth. This pattern was also observed for self-consciousness and denture discomfort, two specific questions of the oral health-related quality of life questionnaire. The findings highlight the non-linearity between number of teeth and these subjective assessments of oral health. Missing teeth has been related to oral health-related quality of life in several studies (27–33) and contrasting results have been reported. Our findings suggest that tooth loss may be associated with both a positive or a negative impact on quality of life, depending on whether a patient relates the loss of their teeth to absence of dental pain and swelling or to functional limitations, such as eating and aesthetic appearance.

The strengths of this study include the diversity of patients, who originated from a large number of clinical practices across the north-west of the USA. The weaknesses of this study include the low response rate and the lack of additional information on other specific dental problems. As people who do not respond to mail surveys may be different from those who do, the low response rate may have introduced bias, the direction and magnitude of which cannot be determined. No information was collected on a range of specific dental problems such as cavities, food impaction, receding gums and other specific dental problems. Such data might have assisted in determining to what extent periodontal therapies could improve oral health-related quality of life outcomes.

In summary, patients presenting for an initial periodontal examination have frequent problems related to teeth, gums and denture, and many of these patients perceive their oral health as fair or poor. In addition, lack of a linear association between periodontal pockets and these subjective measures emphasizes the importance of using subjective oral health assessments in the clinical practice as a tangible patient outcome, as they add information (which is not available through the purely clinical indices) about the impact of the disease state on the patient. Finally, the association of specific aspects of oral health-related quality of life and chronic periodontitis should be confirmed or refuted in other studies as it may significantly change the perception of chronic periodontitis as a silent disease.

Acknowledgements

This research was supported by the NIH/NIDCR: R01 DE13192 and the Department of Veterans Affairs (VA) Health Services Research and Development Service, where Dr. Kressin is supported by a Research Career Scientist award (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.

References

1. Newman MG, Takei HH, Carranza FA. *Carranza's Clinical Periodontology*, 9th edn. Philadelphia: W.B. Saunders Co., 2002.
2. Gilbert AD, Nuttall NM. Self-reporting of periodontal health status. *Br Dent J* 1999;**186:**241–244.
3. Pitiphat W, Garcia RI, Douglass CW, Joshipura KJ. Validation of self-reported oral health measures. *J Public Health Dent* 2002;**62:**122–128.
4. Dietrich T, Stosch U, Dietrich D, Schamberger D, Bernimoulin JP, Joshipura K. The accuracy of individual self-reported items to determine periodontal disease history. *Eur J Oral Sci* 2005; **113:**135–140.
5. Tervonen T, Knuuttila M. Awareness of dental disorders and discrepancy between 'objective' and 'subjective' dental treatment needs. *Community Dent Oral Epidemiol* 1988;**16:**345–348.
6. Blicher B, Joshipura K, Eke P. Validation of self-reported periodontal disease: a systematic review. *J Dent Res* 2005; **84:**881–890.
7. Needleman I, McGrath C, Floyd P, Biddle A. Impact of oral health on the life quality of periodontal patients. *J Clin Periodontol* 2004;**31:**454–457.
8. McGrath C, Bedi R. Measuring the impact of oral health on quality of life in Britain using OHQoL-UK (W)((c)). *J Public Health Dentistry* 2003;**63:**73–77.
9. Astrom AN, Haugejorden O, Skaret E, Trovik TA, Klock KS. Oral Impacts on Daily Performance in Norwegian adults: validity, reliability and prevalence estimates. *Eur J Oral Sci* 2005;**113:**289–296.
10. Borrell LN, Taylor GW, Borgnakke WS, Woolfolk MW, Nyquist LV. Perception of general and oral health in White and African American adults: assessing the effect of neighborhood socioeconomic conditions. *Community Dentistry Oral Epidemiol* 2004;**32:**363–373.

11. Fardal O, Johannessen AC, Linden GJ. Patient perceptions of periodontal therapy completed in a periodontal practice. *J Periodontol* 2002;**73:**1060–1066.
12. John MT, Koepsell TD, Hujoel P, Miglioretti DL, LeResche L, Micheelis W. Demographic factors, denture status and oral health-related quality of life. *Community Dent Oral Epidemiol* 2004;**32:**125–132.
13. Locker D, Matear D, Stephens M, Jokovic A. Oral health-related quality of life of a population of medically compromised elderly people. *Community Dent Health* 2002;**19:**90–97.
14. Atchison KA, Gift HC. Perceived oral health in a diverse sample. *Adv Dent Res* 1997;**11:**272–280.
15. Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS). *Perceived Oral Health Among Adults with Teeth in the United States, 1988–94*. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention 2003.
16. Hujoel PP. Endpoints in periodontal trials: the need for an evidence-based research approach. *Periodontology 2000* 2004;**36:**196–204.
17. D'Avila GB, Carvalho LH, Feres EJ, Feres M, Leao A. Oral health impacts on daily living related to four different treatment protocols for chronic periodontitis. *J Periodontol* 2005;**76:**1751–1757.
18. McGrath C, Wong AHH, Lo ECM, Cheung CS. The sensitivity and responsiveness of an oral health-related quality of life measure to tooth whitening. *J Dentistry* 2005;**33:**697–702.
19. Awad MA, Locker D, Korner-Bitensky N, Feine JS. Measuring the effect of intraoral implant rehabilitation on health-related quality of life in a randomized controlled clinical trial. *J Dent Res* 2000;**79:**1659–1663.
20. Awad MA, Lund JP, Shapiro SH *et al*. Oral health status and treatment satisfaction with mandibular implant overdentures and conventional dentures: a randomized clinical trial in a senior population. *Int J Prosthodont* 2003;**16:**390–396.
21. Robinson PG, Pankhurst CL, Garrett EJ. Randomized-controlled trial: effect of a reservoir biteguard on quality of life in xerostomia. *J Oral Pathol Med* 2005; **34:**193–197.
22. Allen PF, McMillan AS. A longitudinal study of quality of life outcomes in older adults requesting implant prostheses and complete removable dentures. *Clin Oral Implants Res* 2003;**14:**173–179.
23. Slade GD, Spencer AJ. Development and evaluation of the Oral Health Impact Profile. *Community Dent Health* 1994; **11:**3–11.
24. Atchison KA, Dolan TA. Development of the Geriatric Oral Health Assessment Index. *J Dent Educ* 1990;**54:**680–687.
25. Kressin N, 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:**416–427.
26. Slade GD, Nuttall N, Sanders AE, Steele JG, Allen PF, Lahti S. Impacts of oral disorders in the United Kingdom and Australia. *Br Dent J* 2005;**198:**489–493.
27. Jones JA, Orner MB, Spiro A III, Kressin NR. Tooth loss and dentures: patients' perspectives. *Int Dent J* 2003;**53:**327–334.
28. Mack F, Schwahn C, Feine JS *et al*. The impact of tooth loss on general health related to quality of life among elderly Pomeranians: results from the study of health in Pomerania (SHIP-O). *Int J Prosthodont* 2005;**18:**414–419.
29. McGrath C, Bedi R. Can dentures improve the quality of life of those who have experienced considerable tooth loss? *J Dent* 2001;**29:**243–246.
30. Rosenoer LM, Sheiham A. Dental impacts on daily life and satisfaction with teeth in relation to dental status in adults. *J Oral Rehabil* 1995;**22:**469–480.
31. Cushing AM, Sheiham A, Maizels J. Developing socio-dental indicators – the social impact of dental disease. *Community Dent Health* 1986;**3:**3–17.
32. Locker D. Clinical correlates of changes in self-perceived oral health in older adults. *Community Dent Oral Epidemiol* 1997;**25:**199–203.
33. Steele JG, Sanders AE, Slade GD *et al*. How do age and tooth loss affect oral health impacts and quality of life? A study comparing two national samples. *Community Dent Oral Epidemiol* 2004;**32:**107–114.

BMC Medical Genetics

BioMed Central

Introduction

The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports

L Adrienne Cupples*[1,2], Heather T Arruda[1], Emelia J Benjamin[1,2,3,4], Ralph B D'Agostino Sr[1,2,5], Serkalem Demissie[1,2], Anita L DeStefano[1,2,3], Josée Dupuis[1,2], Kathleen M Falls[6], Caroline S Fox[1,7], Daniel J Gottlieb[3,8], Diddahally R Govindaraju[1,3], Chao-Yu Guo[1,5], Nancy L Heard-Costa[1,3], Shih-Jen Hwang[1,7], Sekar Kathiresan[1,9,10,11], Douglas P Kiel[12], Jason M Laramie[3,13], Martin G Larson[1,5], Daniel Levy[1,3,7], Chun-Yu Liu[2], Kathryn L Lunetta[1,2], Matthew D Mailman[14], Alisa K Manning[2], James B Meigs[15,16], Joanne M Murabito[1,3], Christopher Newton-Cheh[1,9,11], George T O'Connor[1,3], Christopher J O'Donnell[1,7,11], Mona Pandey[1], Sudha Seshadri[1,3], Ramachandran S Vasan[1,3,4], Zhen Y Wang[14], Jemma B Wilk[1,3], Philip A Wolf[1,3], Qiong Yang[1,2] and Larry D Atwood[1,2,3]

Address: [1]National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA, [2]School of Public Health, Boston University, Boston, MA, USA, [3]School of Medicine, Boston University, Boston, MA, USA, [4]Whitaker Cardiovascular Institute, Boston University, Boston, MA, USA, [5]Department of Mathematics and Statistics, Boston University, Boston, MA, USA, [6]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, [7]National Heart, Lung and Blood Institute, Bethesda, MD, USA, [8]VA Boston Healthcare System, Boston, MA, USA, [9]Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA, [10]Cardiovascular Disease Prevention Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, [11]Cardiology Division, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, [12]Hebrew SeniorLife: Institute for Aging Research and Harvard Medical School, Boston, MA, USA, [13]Bioinformatics Program, Boston University, Boston, MA, USA, [14]National Center for Biotechnology Information, Bethesda, MD, USA, [15]General Medicine Division, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA and [16]Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Email: L Adrienne Cupples* - adrienne@bu.edu; Heather T Arruda - harruda@bu.edu; Emelia J Benjamin - emelia@bu.edu; Ralph B D'Agostino - ralph@bu.edu; Serkalem Demissie - demissie@bu.edu; Anita L DeStefano - adestef@bu.edu; Josée Dupuis - dupuis@bu.edu; Kathleen M Falls - falls@morgan.harvard.edu; Caroline S Fox - foxca@nhlbi.nih.gov; Daniel J Gottlieb - gottlieb@bu.edu; Diddahally R Govindaraju - drgraju@bu.edu; Chao-Yu Guo - chao-yu.guo@childrens.harvard.edu; Nancy L Heard-Costa - nheard@bu.edu; Shih-Jen Hwang - hwangs2@nhlbi.nih.gov; Sekar Kathiresan - skathiresan@partners.org; Douglas P Kiel - kiel@hrca.harvard.edu; Jason M Laramie - laramiej@bu.edu; Martin G Larson - mlarson@bu.edu; Daniel Levy - levyd@nih.gov; Chun-Yu Liu - liuchunyu2002@yahoo.com; Kathryn L Lunetta - klunetta@bu.edu; Matthew D Mailman - mm888r@nih.gov; Alisa K Manning - amanning@bu.edu; James B Meigs - jmeigs@PARTNERS.ORG; Joanne M Murabito - murabito@bu.edu; Christopher Newton-Cheh - CNEWTONCHEH@PARTNERS.ORG; George T O'Connor - goconnor@lung.bumc.bu.edu; Christopher J O'Donnell - codonnell@nih.gov; Mona Pandey - mpandey@bu.edu; Sudha Seshadri - suseshad@bu.edu; Ramachandran S Vasan - vasan@bu.edu; Zhen Y Wang - wangzhe@mail.nih.gov; Jemma B Wilk - jwilk@bu.edu; Philip A Wolf - pawolf@bu.edu; Qiong Yang - qyang@bu.edu; Larry D Atwood - lda@bu.edu

* Corresponding author

Published: 19 September 2007

BMC Medical Genetics 2007, **8**(Suppl 1):S1 doi:10.1186/1471-2350-8-S1-S1

This article is available from: http://www.biomedcentral.com/1471-2350/8/S1/S1

Abstract

Background: The Framingham Heart Study (FHS), founded in 1948 to examine the epidemiology of cardiovascular disease, is among the most comprehensively characterized multi-generational studies in the world. Many collected phenotypes have substantial genetic contributors; yet most genetic determinants remain to be identified. Using single nucleotide polymorphisms (SNPs) from a 100K genome-wide scan, we examine the associations of common polymorphisms with phenotypic variation in this community-based cohort and provide a full-disclosure, web-based resource of results for future replication studies.

Methods: Adult participants (n = 1345) of the largest 310 pedigrees in the FHS, many biologically related, were genotyped with the 100K Affymetrix GeneChip. These genotypes were used to assess their contribution to 987 phenotypes collected in FHS over 56 years of follow up, including: cardiovascular risk factors and biomarkers; subclinical and clinical cardiovascular disease; cancer and longevity traits; and traits in pulmonary, sleep, neurology, renal, and bone domains. We conducted genome-wide variance components linkage and population-based and family-based association tests.

Results: The participants were white of European descent and from the FHS Original and Offspring Cohorts (examination 1 Offspring mean age 32 ± 9 years, 54% women). This overview summarizes the methods, selected findings and limitations of the results presented in the accompanying series of 17 manuscripts. The presented association results are based on 70,897 autosomal SNPs meeting the following criteria: minor allele frequency ≥ 10%, genotype call rate ≥ 80%, Hardy-Weinberg equilibrium p-value ≥ 0.001, and satisfying Mendelian consistency. Linkage analyses are based on 11,200 SNPs and short-tandem repeats. Results of phenotype-genotype linkages and associations for all autosomal SNPs are posted on the NCBI dbGaP website at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007.

Conclusion: We have created a full-disclosure resource of results, posted on the dbGaP website, from a genome-wide association study in the FHS. Because we used three analytical approaches to examine the association and linkage of 987 phenotypes with thousands of SNPs, our results must be considered hypothesis-generating and need to be replicated. Results from the FHS 100K project with NCBI web posting provides a resource for investigators to identify high priority findings for replication.

Background

Cardiovascular diseases are major illnesses among Americans, affecting about a third of the population (79 million with prevalent disease) and resulting in more than 870,000 cardiovascular disease deaths annually [1]. Cardiovascular disease and its risk factors have substantial genetic contributors [2-11]. Numerous reports from the Framingham Heart Study (FHS) have documented that coronary heart disease [12,13], blood pressure [14-16], lipids [17-20], diabetes [21-23] and weight [24,25] have substantial heritability and linkage/association to specific genomic regions. To evaluate the genetic contributors to these phenotypes, the Framingham Heart Study conducted a genome-wide scan of 1345 study participants in two generations, using genotyping from the 100K Affymetrix GeneChip Human Mapping Set.

In this manuscript, we summarize the strategies that we pursued to conduct the 100K genome-wide study, providing an overview for a series of 17 companion manuscripts (Table 1 of the Overview) describing associations with specific collections of traits [26-42]. The primary purpose of this project was to generate hypotheses regarding genetic factors that may contribute to the wide spectrum of phenotypic variables collected in the FHS through a genome-wide approach. More specifically, we primarily hypothesized that common genetic variants contributing to phenotypic variation can be detected through a genome-wide association study (GWAS) and that genetic loci contributing to phenotypic variation can be detected through linkage. Each manuscript also examines whether the 100K analyses replicated previously reported associations with consistent evidence from the literature for some specific traits. The main purpose of this series of publications is to describe the association results made available for investigators and to direct readers to their free availability in the database of Genotype and Phenotype (dbGaP) public repository http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007 at the National Center for Biotechnology Information (NCBI), where these comprehensive results are posted and may be browsed in the context of multiple genomic tracks includ-

ing Entrez Gene, RefSeq, dbSNP, genetic markers, and OMIM. The deposition of these data in a public repository is consistent with the long tradition of publishing preliminary results from the FHS to benefit the wider scientific community.

To organize the evaluation of the rich resource of data collected over nearly 60 years of follow up, we established a set of "Phenotype Working Groups" that included clinicians, epidemiologists, geneticists, and biostatisticians. These groups specified the traits to be studied, along with covariate adjustment and subgroups for analyses. In all, 987 phenotypes were examined for association, 835 for linkage. Some phenotypes are the same trait with different covariate adjustments, at different examinations or evaluated in different subgroups. For example, many traits were evaluated with both age and sex adjustment as well as with additional multivariable adjustments, yielding more than one phenotype for analysis. Each manuscript in this series provides a platform for the web posted results. Not every trait is described in the manuscripts; rather, the purpose of each manuscript is to introduce the trait areas and to present a brief summary of the results. In the present manuscript, we describe the general approach to analysis of the traits, provide an overview of some results, and discuss the limitations of the studies.

Methods

Study sample

The Framingham Heart Study (FHS) began in 1948 with recruitment of 5209 men and women (2336 men and 2873 women) between the ages of 28 and 62 years in the town of Framingham, Massachusetts, about 20 miles west of Boston [43-46]. These individuals were recruited through a two-thirds systematic sample of the households of Framingham, Massachusetts. Although not initially intended as a family study, many households consisted of spouse pairs (1644 pairs). The primary purpose of the Study was to follow individuals over time for development of cardiovascular disease events to evaluate the interplay among multiple risk factors that lead to disease and their individual and joint effects. The participants in the Original Cohort have been examined every two years since.

In 1971, an Offspring Cohort of 5124 men and women, who were adult children of Original Cohort members or were spouses of these offspring, was recruited and has been examined every four to eight years since [47,48]. The subjects in this report are drawn from the largest 310 pedigrees in these two generations. The participants were recruited without regard to phenotypes. Thus, the Offspring Cohort of 5124 (2483 men and 2641 women) was recruited by inviting all offspring of the spouse pairs (2616 and 34 stepchildren), the offspring spouses (1576) and, additionally, those offspring (898) of singleton Original Cohort members with elevated lipid levels. Further information regarding recruitment can be seen in Cupples *et al*. [49] and Dawber [43].

In the late 1980s and through the 1990s, DNA was collected from living study participants. As many of the Original Cohort members were deceased by that time, these DNAs were mostly collected in Offspring Study participants. During the mid- to late-1990s, 1702 DNA samples were genotyped by the Mammalian Genotyping Service in the largest 330 two-generation pedigrees consisting of 2885 Framingham Study participants. These pedigrees were used for linkage analyses of blood pressure [15], lipids [17,50], body mass index [25] and a wide variety of other traits [51-56]. The numbers of relative pairs among the 1345 subjects both genotyped and phenotyped in this study are 435 parent-offspring pairs, 988 sib pairs, 300 avuncular pairs and 634 first-cousin pairs. Among the 1087 Offspring Cohort participants, who were the only participants evaluated in some analyses, there were 936 sib pairs, 63 avuncular pairs and 612 first-cousin pairs.

Original Cohort study subjects return to the Study every two years for a detailed medical history, physical examination and laboratory tests. The Original Cohort subjects are currently in their 29th examination. Participants in the Offspring Cohort return every 4 to 8 years for similar examinations and the 8th examination is currently underway.

In the early 2000s, a family DNA plate set with 1,399 participants from these 330 pedigrees http://www.nhlbi.nih.gov/about/framingham/policies/index.htm was established. Only subjects with lymphoblast cell lines were included on the plate set, although a substantial number of the DNA samples on the plate set were derived from whole blood or buffy coat. The family plate set was used for genotyping of the Affymetrix 100K GeneChip. After cleaning the genotyping data, the study sample comprised 1345 FHS participants, 278 from the Original and 1087 Offspring Cohorts.

Phenotype definition & methods

Given the breadth of phenotypes, we established 8 larger Phenotype Working Groups, overseeing 17 discrete phenotypic domains, as follows: 1. Blood pressure, arterial stiffness, echocardiography, endothelial function and exercise testing; 2. Metabolic traits including anthropometry, glycemic traits and lipids; 3. Pulmonary function and sleep; 4. Systemic biomarkers, including inflammatory and thrombotic factors; 5. Subclinical atherosclerosis; 6. Renal and endocrine function; 7. Longevity and aging, including brain and bone aging phenotypes; 8. Cardiovascular disease outcomes, cancer, electrocardiography and

Table 1: Papers published in the Framingham Heart Study 100K Series

First Author	Title
Ramachandran S. Vasan	Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the Framingham Heart Study
Daniel Levy	Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness
Christopher J. O'Donnell	Genome-wide association study for subclinical atherosclerosis in major arterial territories in the NHLBI's Framingham Heart Study
Martin G. Larson	Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes
Joanne M. Murabito	Genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study
Christopher Newton-Cheh	Genome-wide association study of electrocardiographic and heart rate variability traits: the Framingham Heart Study
Jemma B. Wilk	Framingham Heart Study genome-wide association: results for pulmonary function measures
Daniel J. Gottlieb	Genome-wide association of sleep and circadian phenotypes
Shih-Jen Hwang	A genome-wide association for kidney function and endocrine-related traits in the NHLBI's Framingham Heart Study
Emelia J. Benjamin	Genome-wide association with select biomarker traits in the Framingham Heart Study
Qiong Yang	Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study
Kathryn L. Lunetta	Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study
Douglas P. Kiel	Genome-wide association with bone mass and geometry in the Framingham Heart Study
Sudha Seshadri	Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham study
James B. Meigs	Genome-wide association with diabetes-related traits in the Framingham Heart Study
Sekar Kathiresan	A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study
Caroline S. Fox	Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project

heart rate variability. In addition, we established a statistical and analytical methodology group. These groups were convened by the FHS Genetic Steering Committee to define phenotypes to be evaluated, including the covariates used in analyses, to review results of linkage and association analyses, to foster communication among various Framingham investigators who were working on different traits, and to suggest possible follow-up strategies.

For the 100K genome-wide project, each Working Group defined the phenotypes to be studied. Since most traits have well established factors that contribute to their variation, each group created a set of residuals from multivariable regression models accounting for the primary known covariates, in order to control for confounding from these variables and to increase the ability to detect genetic signals. For quantitative traits, the adjusted standardized residuals were generated using linear regression models. For qualitative traits, we used a variety of approaches including Cox proportional hazards with Martingale residuals for time-to-event (survival) traits and logistic regression with deviance residuals for dichotomous traits. These methods are described below. In some cases, several different covariate adjustments were used for a single trait. Each manuscript describes the specific adjustments that were applied. We used residuals from regression models that included all subjects with traits in each Cohort, rather than limiting analyses to those who were genotyped, to produce residuals based on all subjects with phenotypic values, regardless of availability of genotypic data. This approach avoids potential biases in covariate adjustment based only upon the subset of individuals with both genotype and phenotype data and produces robust estimates of covariate effects.

Genotyping methods

Genomic DNA derived from whole blood or buffy coat was phenol-chloroform extracted and DNA from immortalized lymphoblast cell lines was salt-precipitate extracted. Genotyping of the 100K SNPs in FHS families was performed through an ancillary study to Drs. Michael Christman and Alan Herbert at Boston University School of Medicine in the Department of Genetics and Genomics using the GeneChip Human Mapping 100K set from Affymetrix, following the manufacturer's protocol as previously described [57]. Genotypes were determined using the Dynamic Modeling (DM) algorithm [58]. For linkage analyses, we also included microsatellites that had been genotyped by the NHLBI Mammalian Genotyping Service, Center for Medical Genetics, Marshfield Medical Research Foundation http://research.marshfieldclinic.org/genetics. A set of 401 microsatellite markers)[59], covering the genome at an average density of one marker every 10 cM and with an average heterozygosity of 0.77, were genotyped in 1702 subjects in the mid to late 1990s (Screening Set v. 8))[60]. An additional 190 participants on the Family Plate Set were genotyped later with microsatellites using Screening Set v.13 and some additional microsatellites were also genotyped in the FHS Genetics Laboratory. With the addition of these microsatellites and changes in the marker sets from Set 8, there were 613 microsatellite markers available for analysis.

Statistical analysis methods

Data cleaning

A total of 1380 individuals were successfully genotyped. First, familial relationships were checked using the sib_kin utility in the Aspex software package [61]. Because this study focused on participants of families, nine individuals were excluded as they no longer had biologic relatives in the sample. Twenty-six individuals were excluded due to inconsistencies; the majority of these individuals were found to have an excessive number of Mendelian errors as identified by the software PedCheck, Version 1.1 [62]. Others were excluded for having a relationship inconsistency, for a sex discrepancy or for a low genotyping call rate. Mendelian inconsistencies were resolved by removing the genotypes of all individuals within nuclear families in which the error occurred. These steps left 1345 individuals with genotypes available for analyses.

For Hardy-Weinberg equilibrium (HWE) testing, we randomly selected one individual per family to form a sample of unrelated individuals. Then, for each of the 100K SNPs, the observed genotype frequencies were compared to those expected under HWE using an exact chi-square test statistic [63] implemented in the Genetics package [64] in R, Version 1.2.0 [65]. To guard against a result that might depend upon an unusual selection of individuals, we repeated this process of a random selection of subjects ten times and computed the geometric mean of the ten p-values for the ten random samples of individuals as the final p-value for HWE tests for each SNP. Tests for HWE that indicate a SNP is far from HWE suggest that the SNP may have issues with genotyping error.

We found 38,062 SNPs with MAF <10% (of which 3084 autosomal SNPs were monomorphic), an additional 2346 SNPs with genotyping call rates <80% and still another 1595 with HWE p-value < 0.001. We used a genotyping call rate cutpoint of 80% in part because of the use of the less accurate DM algorithm. SNPs with low MAF, low genotyping call rates and inadequate HWE produced unstable results, so they were excluded from our association results reported in this set of manuscripts, leaving 70,987 SNPs. Results for all autosomal SNPs are reported on the dbGaP website, regardless of MAF, genotyping call rates or HWE p-values; however, filters for these factors are provided on the dbGaP website.

Linkage analyses

Both microsatellites previously genotyped by the Mammalian Genotyping Service and SNPs from the 100K were used to calculate identity by descent probabilities. We constructed genetic maps using all microsatellite NCBI genetic markers with Marshfield genetic location available and whose physical order and genetic order were consistent. Using this NCBI Marshfield map as our skeleton, we applied linear interpolation from physical to genetic distance to obtain approximate genetic locations (in centi-Morgans) for all SNPs in the 100K set with known physical location.

Because current linkage analysis software cannot handle the marker density available from a 100K scan, we selected a subset of 10,592 SNPs to supplement 613 genome scan microsatellite markers available on 1886 members of the largest 330 Framingham families. We selected SNPs to minimize linkage disequilibrium (LD) because current linkage software assumes that markers are in linkage equilibrium, and violation of this assumption has been shown to create spurious linkage evidence in certain contexts [66,67]. Thus, for calculation of identity by descent (IBD) probabilities for linkage analyses we used SNPs with a call rate of at least 85%, HWE p-value > 0.05 and more informative markers with MAF > 5%. We iteratively identified SNP pairs with LD measure D' > 0.5, as estimated from HapMap data, and eliminated the SNP that was least informative for linkage (lowest MAF). We started with SNP pairs most closely located (physical distance) and continued until no pairs of SNPs had a D' measure exceeding 0.5. The final set of 10,592 SNPs combined with the 613 microsatellites were checked for excess recombination using MERLIN, Version 0.10.2 [68], and 4 SNPs and 1 microsatellite were omitted from linkage analyses based on a high number of possible errors, leaving a total of 11,200 markers to perform linkage analysis (10,588 SNPs + 612 short tandem repeats).

Variance component linkage analyses were performed on residuals of up to 1341 individuals in 310 full pedigrees. Four of the 1345 subjects were the only person in a pedigree and were excluded from linkage analyses since they contributed no information. Multipoint probabilities of IBD between relative pairs were computed at each genetic marker location with the program MERLIN, Version 0.10.2. Due to size limitations for exact identity by descent (IBD) multipoint computation in MERLIN software [68], the 310 full pedigrees were broken into 356 smaller pedigrees. The hypothesis of "no linkage at a specific genomic location" was tested by comparing models incorporating an effect of a putative quantitative trait locus (QTL) in complete linkage to the genetic marker, in the form of multipoint IBD sharing probabilities at the locus, to models incorporating only polygenic effects without a QTL effect. At each genetic location, a LOD score was computed as the logarithm to base 10 of the likelihood ratio of the locus-specific model to the polygenic model using the program SOLAR, Version 3.0.4 [69]. Allele frequencies were estimated by simple allele counts. In Framingham family data, which were collected from randomly sampled pedigrees, we have found the allele frequency estimates by simple allele counting

closely match those calculated by maximum likelihood methods accounting for familial correlations.

Association testing

We applied population-based and family-based methods to test for association between the 100K SNPs and residual phenotypes using an additive model unless otherwise specified. We used family-based association test methods, implemented in the program FBAT, Version 1.5.5 [70,71], to test for differences in probability of transmission of a genotype from parents to offspring based on phenotype, as a test of linkage and association. FBAT has limited power because it requires association within families, and many families are non-informative. However, because FBAT examines association only within families, the type-I error rate is not affected by population stratification bias)[72,73]. We did not report results if the number of informative families was fewer than 10.

For the population-based approach, we used generalized estimating equation (GEE) [74] regression models to test for association between the 100K SNPs and each residual phenotype while taking into account the correlation among related individuals. We implemented the GEE approach by breaking families into sibships and used an exchangeable working correlation matrix to account for correlation within each sibship. Parental correlations with their children were not considered in these analyses. The analyses were performed using the gee program package, Version 4.13-10 [75] in R [65]. The GEE association test is a population-based approach that uses all individuals with both genotype and phenotype, regardless of genotype configuration within a family. Therefore, it is expected to be a powerful test of association if population stratification bias is not believed to be an issue, as in the FHS [76].

Results

Participant characteristics

Of the 1345 subjects who satisfied appropriate familial relationships and who were considered in the presentation of results in these manuscripts, 258 were Original Cohort participants (90 men and 168 women) and 1087 were Offspring Cohort participants (527 men and 560 women). Table 2 of the Overview presents descriptive information on these participants at enrollment (examination one). The Offspring and Cohort participants included on the family plates had lower mean age than other examination 1 participants, as these subjects needed to survive to the mid 1990s to provide DNA. We note that we used residuals based upon all subjects, as opposed to only those who were genotyped. Thus, the phenotypes reflect deviations of these subjects based on regressions for the full sample of subjects and are thus representative of the full sample.

Format of the FHS 100K manuscripts

We present 17 manuscripts, each displaying selected results for an epidemiologically related group of traits. Table 1 of the Overview presents the title and first author of each manuscript. The web resource displaying genetic association and linkage results is available at the NCBI dbGaP website, http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007. Each manuscript describes the traits that were studied and presents some results. These manuscripts are not intended to be comprehensive and generally do not include results for all phenotypes and covariate adjustment schemes that were studied and presented on the website. Full listings of all traits evaluated are provided in Additional file 1 (phenotypes for population-based GEE analyses), Additional file 2 (phenotypes for family-based FBAT analyses) and Additional file 3 (phenotypes for linkage analyses), including url links to the corresponding analytical results on the NCBI dbGaP website. To facilitate the reading of these manuscripts, we have used a common format for all manuscripts. Table 1 of each manuscript presents a general description of the phenotypes that were evaluated. Table 2 of each manuscript displays the top results (lowest p-values) from GEE analyses, the top results (lowest p-values) from FBAT analyses and linkage results where the LOD score was 2 or more. Whereas top association results are based solely on p-value rank, the Working Groups also applied various additional strategies to identify SNPs that the group would prioritize to pursue further. For Table 3 of each manuscript, the groups devised schema to summarize results for related traits, grouping phenotypic traits within biologically plausible domains, or traits examined longitudinally. Each manuscript provides a description of the strategy employed and the results for its Table 3. Finally, Table 4 in each manuscript lists some SNPs that are the same as or correlated with genetic variants in genes that have been reported in the literature to be associated with the manuscript's phenotypes and indicates whether our results replicate those reports. Physical locations of the SNPs are provided according to NCBI Build35, whereas the dbGaP website uses a more recent version. Thus, the physical locations reported in the manuscripts may differ from those on the website. Each manuscript provides criteria for choosing which results were reported.

SNP allele frequencies and distribution

Allele frequencies for the 100K Affymetrix GeneChip in the Framingham sample are displayed in Figure 1. About 38% have MAF < 10% and are not considered in the series of manuscripts, although they are included on the dbGaP website. Among SNPs with MAF ≥ 10%, there were large numbers between 10–25% and were somewhat evenly spread over the range from 25–50% MAF. Many SNPs on the Affymetrix Chip are not near genes (Figure 2). About 30,000 with MAF ≥ 10% are within 5 kb of a gene; another

Table 2: Description of Framingham Heart Study Subjects in 100K Genome-Wide Scan. Baseline Data at Exam 1 for Original (1948–1951) and Offspring (1971–1975) Cohorts

	Original Cohort Men N = 90	Original Cohort Women N = 168	Offspring Cohort Men N = 527	Offspring Cohort Women N = 560
Age, years	35 ± 4	35 ± 4	31 ± 10	32 ± 10
(Limits)	(30–46)	(29–48)	(11–62)	(5–59)
Body Mass Index, kg/m^2	25.7 ± 3.2	23.9 ± 3.6	26.0 ± 3.9	23.6 ± 4.3
Obese (BMI 30 ≥kg/m^2), %	7.8	6.0	13.3	8.2
Systolic Blood Pressure, mm Hg	129 ± 13	120 ± 12	124 ± 14	115 ± 14
Diastolic Blood Pressure, mm Hg	81 ± 9	76 ± 8	81 ± 10	75 ± 10
Antihypertensive Medication, %	0	0	1.7	1.4
Hypertension, %	22.2	8.9	15.0	6.8
Current Smoking, %	61.1	42.3	40.0	41.3
Blood Glucose, mg/dL	78 ± 13	78 ± 11	102 ± 10	97 ± 9
Prevalent Diabetes, %	0	0	0.6	0.4
Total Cholesterol, mg/dL	220 ± 44	194 ± 37	192 ± 37	185 ± 36
HDL Cholesterol, mg/dL	N/A	N/A	44.9 ± 11.0	56.4 ± 13.9
Lipid Lowering Medication, %	0	0	0.2	0.5
Hemoglobin, g/dL	14.3 ± 1.2	12.3 ± 1.1	15.5 ± 1.0	13.5 ± 1.0
Forced Vital Capacity, dL	40 ± 7	28 ± 6	45 ± 8	32 ± 5
Prevalent CVD, %	0	0	5.5	2.6

Abbreviations: BMI = body mass index; HDL = high density lipoprotein; CVD = cardiovascular disease.

10,000 with MAF < 10% are within 5 kb. The remaining SNPs are further away from known genes.

P-value distribution

The results displayed on the NCBI dbGaP http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007 include all autosomal SNPs, regardless of genotyping call rate, HWE p-value or MAF; adjustable filters for these factors are provided. In our manuscripts, we present results for SNPs that satisfy genotyping call rates ≥ 80%, HWE p-value ≥ 0.001, and MAF ≥ 10%. The proportion of SNPs satisfying a genotyping call rate of ≥80% is 97.4%; 91% of SNPs satisfied a call rate of ≥90%.

We expect only a small percentage of all tested SNPs to be truly associated with any phenotype. Therefore, to obtain an approximation of the null distribution of p-values, we examined the distribution of p-values for 415 phenotypes from the Metabolic Working Group and 14 CVD event phenotypes. If one assumes that only a few true associations exist for each phenotype, these p-value distributions approximate the null distribution, because only a few SNPs out of the large total number tested would be expected to exceed any critical value due to true associations. Table 3 of the Overview displays the proportion of p-values among all SNPs below specific nominal alpha levels, summarized (mean, minimum and maximum) across all phenotypes in the trait group for GEE and FBAT results.

Many of the phenotypes in the Metabolic Working Group were approximately normally distributed (about 90% had absolute value of skewness <1 and about 80% had absolute value of kurtosis <2) and thus may reflect the situation for which the assumptions of the analytical methods were generally satisfied. We display two sets of results in Table 3 of the Overview for these phenotypes, those used in the publication of the manuscripts with the number of SNPs equal to 70,987 and the larger set of results displayed on the website with the number of SNPs equal to ~100–103 K. The difference in the number of SNPs evaluated for GEE and FBAT results arises from those SNPs that are uninformative for FBAT analyses (those with sufficiently rare minor allele so that fewer than 10 nuclear families were informative for transmission). The p-value distributions suggest that FBAT p-values generally follow the expected null distribution, assuming that nearly all results are false positives, and may actually be somewhat conservative. In contrast, the GEE p-values exhibit an excess of small p-values, especially for smaller nominal alpha levels. For example, for SNPs reported in the manuscripts, the average proportion of SNPs for a phenotype with p-value below specified alpha levels ranged from 1.3 to 19 times greater than the nominal level (1.3 times larger for nominal alpha of 0.01, 19 for nominal alpha of 10^{-7} and 10 for 10^{-8}). The excess is higher for the full set of SNPs reported on the website. Here we found that the average proportion ranged from 1.2 times larger for nominal level of 0.05 to 19 times greater for nominal level of 10^{-5} and 2500 times greater for nominal level 10^{-8}.

The CVD phenotypes represent an extreme case, as the phenotypes were residuals from survival models, were generally bimodal, and do not satisfy general assumptions for normality. We see the same general pattern that we observed for the Metabolic Working Group phenotypes with somewhat conservative FBAT tests and excess num-

Table 3: Proportion of p-values falling below nominal levels for selected traits

SNPs with MAF ≥ 10%, HWE p ≥ 0.001 and Genotyping Call Rate ≥ 80%									
		Nominal alpha level							
	# SNPs	0.05	0.01	0.001	1×10^{-4}	1×10^{-5}	1×10^{-6}	1×10^{-7}	1×10^{-8}
Metabolic Traits*									
Mean† FBAT	70987	0.050	0.010	8.5×10^{-4}	7.3×10^{-5}	5.1×10^{-6}	6.1×10^{-7}	1.4×10^{-7}	0
Min† FBAT		0.046	0.008	4.4×10^{-4}	0	0	0	0	0
Max† FBAT		0.056	0.012	1.5×10^{-3}	2.7×10^{-4}	5.6×10^{-5}	4.2×10^{-5}	2.8×10^{-5}	0
Mean GEE	70987	0.058	0.013	1.6×10^{-3}	2.3×10^{-4}	3.7×10^{-5}	7.9×10^{-6}	1.9×10^{-6}	1.0×10^{-7}
Min GEE		0.051	0.010	9.2×10^{-4}	5.6×10^{-5}	0	0	0	0
Max GEE		0.081	0.022	3.3×10^{-3}	6.6×10^{-4}	1.4×10^{-4}	8.5×10^{-5}	5.6×10^{-5}	1.4×10^{-5}
CVD Events*									
Mean FBAT	70987	0.050	0.009	6.6×10^{-4}	4.4×10^{-5}	4.0×10^{-6}	0	0	0
Min FBAT		0.047	0.006	1.7×10^{-4}	0	0	0	0	0
Max FBAT		0.053	0.010	9.4×10^{-4}	8.5×10^{-5}	1.4×10^{-5}	0	0	0
Mean GEE	70987	0.054	0.012	1.4×10^{-3}	1.8×10^{-4}	2.4×10^{-5}	6.7×10^{-6}	2.7×10^{-6}	6.7×10^{-7}
Min GEE		0.046	0.009	7.9×10^{-4}	4.2×10^{-5}	0	0	0	0
Max GEE		0.062	0.015	2.5×10^{-3}	5.9×10^{-4}	1.3×10^{-4}	5.6×10^{-5}	2.8×10^{-5}	1.4×10^{-5}
All SNPs posted on website http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007									
Metabolic Traits*									
Mean FBAT	100584	0.050	0.009	7.7×10^{-4}	6.3×10^{-5}	4.3×10^{-6}	5.7×10^{-7}	1.2×10^{-7}	0
Min FBAT		0.046	0.008	4.4×10^{-4}	0	0	0	0	0
Max FBAT		0.054	0.011	1.3×10^{-3}	2.2×10^{-4}	5.0×10^{-5}	3.0×10^{-5}	2.0×10^{-5}	0
Mean GEE	101944	0.061	0.015	2.4×10^{-3}	5.5×10^{-4}	1.9×10^{-4}	8.4×10^{-5}	4.3×10^{-5}	2.5×10^{-5}
Min GEE		0.052	0.012	1.5×10^{-3}	1.9×10^{-4}	3.0×10^{-5}	0	0	0
Max GEE		0.082	0.025	1.1×10^{-2}	6.8×10^{-3}	4.3×10^{-3}	2.6×10^{-3}	1.7×10^{-3}	9.4×10^{-4}
CVD Events*									
Mean FBAT	101060	0.048	0.008	5.9×10^{-4}	3.5×10^{-5}	3.3×10^{-6}	0	0	0
Min FBAT		0.043	0.005	1.2×10^{-4}	0	0	0	0	0
Max FBAT		0.052	0.010	9.1×10^{-4}	9.9×10^{-5}	1.0×10^{-5}	0	0	0
Mean GEE	103194	0.059	0.016	3.4×10^{-3}	1.1×10^{-3}	4.7×10^{-4}	2.0×10^{-4}	9.5×10^{-5}	4.4×10^{-5}
Min GEE		0.050	0.011	1.1×10^{-3}	1.3×10^{-4}	2.9×10^{-5}	0	0	0
Max GEE		0.078	0.033	1.4×10^{-2}	7.1×10^{-3}	3.6×10^{-3}	1.6×10^{-3}	7.8×10^{-4}	3.3×10^{-4}

* There were 415 Metabolic Traits and 14 CVD Events from which these descriptive statistics were calculated using the # SNPs indicated. # SNPs = average number of SNPs across all traits in the Trait Group; †Mean = average proportion of SNPs below nominal level across phenotypes in the trait group; Min = minimum proportion below nominal level across phenotypes in the trait group; Max = maximum proportion below nominal level across phenotypes in the trait group.

bers of small p-values for GEE tests. As might be expected, the CVD phenotypes revealed larger excesses of small GEE p-values, with the average proportion of SNPs falling below specified alpha levels ranging from 1.08 (5.4% for nominal 5% with SNPs reported in the manuscripts) to 2.4 times greater for nominal level of 10^{-5} and 67 times greater for nominal 10^{-8} than expected. Thus, GEE p-values need to be interpreted with care, as SNPs in the lowest p-value range have the potential to be especially enriched with false positives.

We also examined the dependence of the p-value distribution for GEE results on the genotyping call rate for the Metabolic Working Group phenotypes. Our sample was not ascertained on trait status; so genotyping failures were likely to be randomly distributed. Therefore, one might expect that the effect of genotyping error on type I error would be more modest than for case-control studies)[77]. As expected, we continued to find an excess of small p-values, despite increasingly stringent call rate thresholds. More importantly, we found that this excess occurred regardless of call rate. For example, for nominal alpha of 0.001 and genotyping call rate > 95%, we found that the ratio of the number of observed to expected significant results ranged from 1.6 for MAF in the range of (0.2, 0.5) to 7.0 for MAF in the range (0, 0.05). Similarly, for call rate less than 80% we found similar ratios of 1.6 to 8.1, respectively. For nominal alpha of 10^{-6} we found this ratio varied from 9.5 to 614 for call rate > 95% and 6.8 to 667 for call rate < 80%. Thus, we used a liberal genotyping call rate of > 80% for presentation of results in our manuscripts to err on the side of including a result rather than not, even though we expect nearly all results to be false positives.

Table 4: Power of the population-based association approach (GEE test) for a SNP with MAF = 0.1

			Nominal Type I Error							
SNP QTL Heritability	Effect Size (SD)*	Model	0.05	0.01	0.001	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}
100% phenotype available (n = 1345)										
1%	0.24	GEE	0.977	0.919	0.78	0.60	0.43	0.27	0.17	0.10
2%	0.33	GEE	1	1.00	0.99	0.97	0.91	0.84	0.72	0.59
3%	0.41	GEE	1	1	1	0.998	0.992	0.985	0.957	0.918
4%	0.47	GEE	1	1	1	1	1	1	0.996	0.991
5%	0.53	GEE	1	1	1	1	1	1	1	1
~80% subjects have phenotype(~20% missing at random): Sample size = 1076										
1%	0.24	GEE	0.934	0.836	0.631	0.427	0.261	0.149	0.074	0.041
2%	0.33	GEE	1.00	0.99	0.97	0.93	0.82	0.66	0.51	0.37
3%	0.41	GEE	1	1	0.998	0.993	0.976	0.938	0.85	0.753
4%	0.47	GEE	0.998	0.998	0.998	0.998	0.991	0.984	0.971	0.949
5%	0.53	GEE	1	1	1	1	1	1	1	0.99
~60% subjects have phenotype(~40% missing at random): Sample size = 783										
1%	0.24	GEE	0.879	0.737	0.48	0.281	0.137	0.055	0.023	0.008
2%	0.33	GEE	0.992	0.96	0.88	0.744	0.591	0.433	0.293	0.186
3%	0.41	GEE	1	1	0.982	0.945	0.874	0.754	0.602	0.467
4%	0.47	GEE	1	1	1	0.998	0.981	0.95	0.884	0.797
5%	0.53	GEE	1	1	0.999	0.997	0.996	0.982	0.964	0.931

*SD: Standard Deviation. The effect size is expressed in unit of standard deviation of the phenotype.

Figure 3 displays observed GEE (blue) and FBAT (red) p-values versus expected p-values (straight line) on a negative logarithm scale for mean fasting plasma glucose and mean high density lipoprotein cholesterol, calculated from Offspring measurements over exams 1 to 7. We see that FBAT p-values tend to be less significant than expected (conservative) whereas GEE p-values tend to be more significant than expected (liberal), especially for smaller expected p-values. While we would expect most p-values to fall on the line if there were no genetic associations, p-values that reflect true associations will be more extreme (smaller) than expected. In looking at the figure for mean fasting glucose, SNPs represented by the blue dots (GEE) far above the expected line on the right hand side of the figure may represent true associations with mean fasting plasma glucose. The plot for mean fasting HDL cholesterol also suggests that there may be some true positives, as even a few FBAT p-values are more extreme than expected.

As in any GWAS, we expect that most results with small p-values are false positives. The p-value distributions support this notion and further suggest that the GEE results may have more false positives than one would expect. Table 2 in each manuscript ranks results by p-value, but each paper also pursues its own strategy to identify which results may be more worthy of follow up in Table 3 of each manuscript, usually by considering evidence from several sources such as correlated traits.

Power estimations for population-based association approach

To assess the power of the population-based association approach with GEE, we simulated a trait following a normal distribution with 30% polygenic heritability in this sample of 1345 subjects. We generated a SNP with MAF 0.10 and assumed that the SNP was the QTL with an additive effect and QTL heritability varying from 1% to 5%. We also varied the proportion of phenotyped individuals from 60% to 100%, as some traits were not available in all subjects genotyped. The phenotype and genotype data were simulated using SOLAR simqtl, Version 3.0.4. We tested the association between the SNP and the trait using GEE. One thousand replicates were performed for each scenario. The results are displayed in Table 4 of the Overview. For a conservative alpha level such as 10^{-8}, we have more than 80% power to detect a SNP explaining 4% or more total phenotypic variation when 60% or more individuals are phenotyped. With higher MAF, the power remains similar for the same QTL heritability (data not shown). Thus, we have sufficient power to detect SNPs explaining ≥4% or more of the phenotypic variance using the population-based GEE association test approach, controlling for multiple testing for a single trait. The effect size for a specific QTL heritability, defined as the increase/

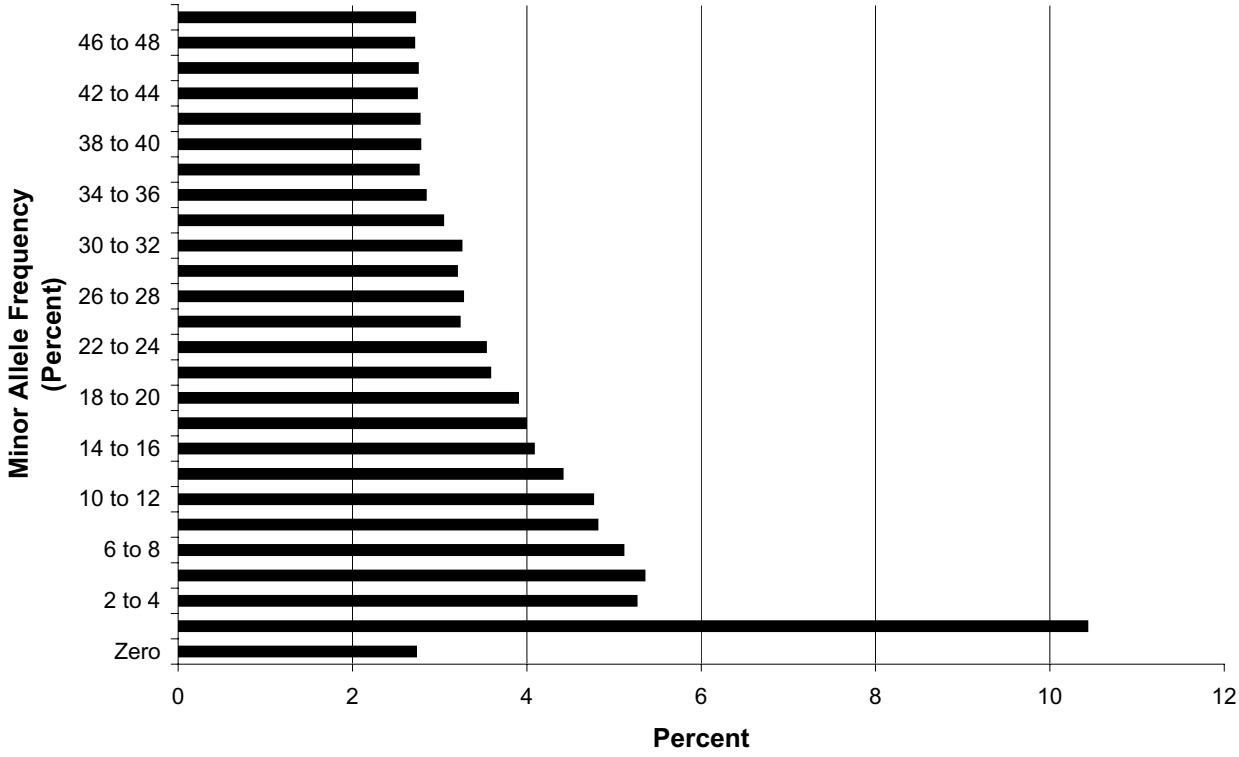


Figure 1

Distribution of Minor Allele Frequency for SNPs in Framingham Sample displayed on dbGaP Website. The percentage of SNPs (X axis) with MAF of zero and in ranges (0,2], (2,4], ..., (48,50] percent (Y axis) in the Framingham sample of 1345 subjects is detailed. For example, approximately 2.5–3% of SNPs in the Affymetrix 100K GeneChip had MAF of zero in the Framingham sample and about 4% had MAF greater than 16 percent and less than or equal to 18 percent. This distribution represents SNPs described on the dbGaP website. The manuscripts only include SNPs with MAF of 10% or more.

decrease of the phenotype value with one copy increment of the allele tested, depends on the MAF of the SNP. For example, for a SNP explaining 4% of the phenotypic variation, the effect size is 0.47 times the phenotypic standard deviation (SD) for a MAF of 0.1, and 0.28 SD for a MAF of 0.5.

Synthetic strategies

Beyond simple examination of individual p-values for single tests, the authors of each manuscript developed their own synthetic strategies to prioritize SNPs that may be worthy of follow up (results shown in Table 3 of each manuscript). Many strategies considered results among similar traits. For example, the Subclinical Working Group created four subgroups of traits: (1) ankle brachial index phenotypes, (2) common carotid IMT phenotypes, (3) internal carotid IMT phenotypes, and (4) multidetector computed tomography (MDCT) coronary calcification phenotypes [29]. Within each trait group, SNPs were ranked according to the proportion of traits with $p < 0.01$ for both FBAT and GEE in the group. Table 3 in the Subclinical manuscript displays the top 5 ranked SNPs with highest proportions of significant traits and lowest mean GEE p-values for each trait group [29]. The Lipids Subgroup of the Metabolic Working Group decided to focus on four phenotypes measured in Offspring subjects: apolipoprotein A-I levels measured at exam 4, small low-density lipoprotein as measured by nuclear magnetic resonance, mean high-density lipoprotein cholesterol levels over 7 exams, and mean log triglyceride levels over 7 exams [39]. Presenting only those results where at least 3 of the 4 traits had GEE p-value < 0.01, the SNPs were ranked according to geometric mean of these four traits [39]. The manuscript presenting results for neurological traits focused on specific phenotypes within subgroups

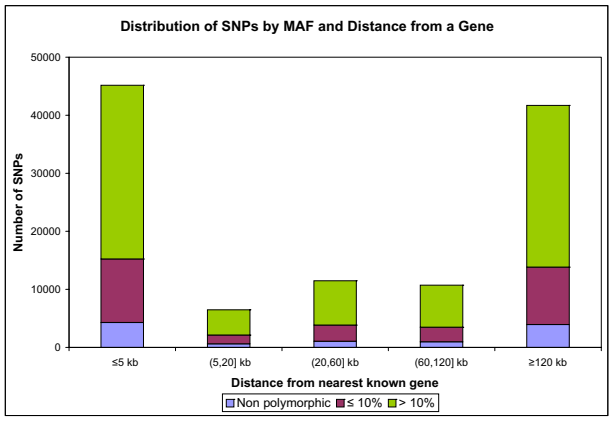


Figure 2
Distribution of Affymetrix 100K GeneChip SNPs by distance from known genes. The X axis is the distance from known genes and the Y axis is the number of SNPs according to each distance. Blue represents monomorphic SNPs, maroon is for SNPs with MAF < 10% and green is for SNPs with MAF ≥ 10%.

with p < 0.001 in FBAT or GEE results and also other traits within the subgroup with p < 0.01 [37].

Some results of interest

The main results for each Working Group are presented in the individual manuscripts of this series. Here we highlight some results that address some of our expectations.

Overlap of linkage and association results

Whereas strategies for genetic studies have been undergoing substantial changes in recent years, partly due to changes in the laboratory, we hypothesized that genomic regions that harbor significant linkage results would also contain significant association results. One example of the concordance of linkage and association results was noted for monocyte chemoattractant protein-1 (MCP1), an inflammatory marker, where a region of linkage (LOD = 4.96 at chromosome 1, 159 Mb) for MCP1 concentrations also contains two SNPs on chromosome 1 within the 1.5 support interval for the linkage peak and within 60 kb of the genes *OR10J1* and *FCER1A* or *OR10J3* (rs4128725 and rs2494250 with p-values in the 10^{-8} range by FBAT, $\leq 10^{-12}$ by GEE) [33]. We found a similar result for C-reactive protein with a LOD score of 3.28 on chromosome 1 and two significant SNPs (rs2794520 ($p = 2.83*10^{-8}$) and rs2808629 ($p = 3.19*10^{-8}$)) within the 1.5 support interval of this peak [33]. Whereas Lp(a) had the highest LOD score (LOD = 23.0) at 159.4 Mb, a SNP (rs1591375) located at 160.7 Mb near this peak had somewhat modest p-values by comparison ($4.37*10^{-06}$ by GEE and 0.0045 by FBAT) [39]. However, there were many instances for which there was no evidence of linkage in the setting of significant association. Additionally, each manuscript describes linkage results that are in accord with previously published linkage results.

SNPs overlapping across phenotypes

We did not expect the same SNPs to appear in many manuscripts, as cardiovascular disease is complex and involves a large and varied number of pathways for its development. In contrast, some manuscripts report on correlated traits. Thus, we examined overlap among the top 500 SNPs associated with the phenotypes across 3 Metabolic Working Groups: glycemic/diabetes phenotypes, lipid phenotypes and obesity phenotypes. Of 11 SNPs found in more than one group, none were found among the top 500 SNPs in all three groups. However, 7 SNPs were found in the glycemia and obesity groups, 2 in glycemia and lipid groups and 2 in lipid and obesity groups [38-40].

Replication of prior associations

In Table 4 of each manuscript, we investigated whether our results replicated previous reports in the literature. The 100K chip does not contain many SNPs in well-known lipid genes, such as *APOE*. On the other hand, we found that SNP rs7007797 in the *LPL* gene was associated with both HDL and triglycerides [39]; we replicated recent findings of association of a SNP in the TCF7L2 gene with diabetes [38]. Strong statistical support was found for the association of factor VII concentrations with SNP rs561241 on chromosome 13 ($4*10^{-16}$) [34], which resides near the factor VII gene and is in complete linkage disequilibrium ($r^2 = 1$) with the Arg/Gln FVII SNP previously shown to account for 9% of the total phenotypic variance [78]. Similarly, we found associations of circulating levels of C-reactive protein with a SNP in the gene encoding C-reactive protein [33]. Two SNPs in *SORL 1*, a gene recently related to the risk of Alzheimer's disease [79], were found to be associated with performance on tests of abstract reasoning (rs1131497; FBAT $p = 3.2 \times 10^{-6}$; rs726601; FBAT $p = 8.2 \times 10^{-4}$) [37]. We found that SNPs (rs2543600 and rs27225364) near the *WRN* gene that causes premature aging are associated with age at death and morbidity-free survival at age 65 years [35]. The LD between these SNPs and those previously reported in the WRN gene is unknown as the previously reported SNPs are not in the HapMap. We found that SNP rs2478518 in the *AGT* gene was associated with both systolic and diastolic blood pressure [28]. The association of common variation at the *NOS1AP* locus with electrocardiographic QT interval duration was replicated with p-values ranging from 0.0001–0.0009 for 4 partially correlated SNPs [27]. In contrast, a number of results reported

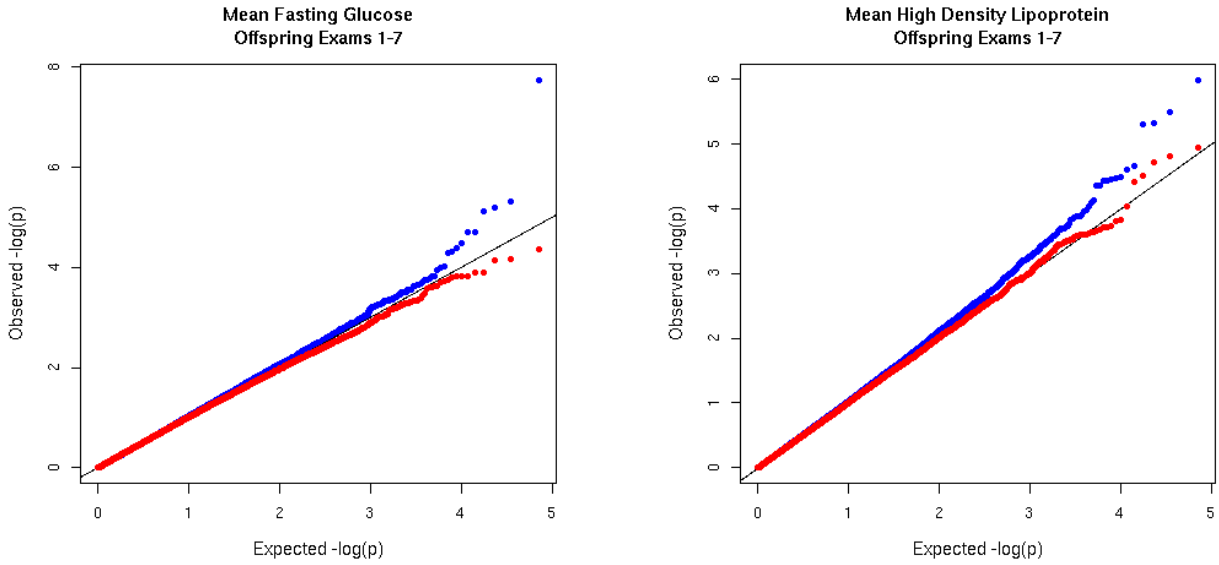


Figure 3
Observed versus Expected p-values (-log base 10 scale) for Mean Fasting Glucose in Offspring Exams 1 to 7 (Left) and for Mean Fasting High Density Lipoprotein in Offspring Exams 1 to 7 (Right). Blue dots are for GEE and red dots for FBAT.

in the literature were not replicated in our results. For example, we did not find significant SNPs in *ACE* associated with blood pressure [28]. *PPARG* P12A (rs1801282) was not associated with diabetes or related traits including body mass index [38]. Some of these 'negative' results may be due to low power or low LD between a SNP reported in the literature and SNPs on the 100K chip.

We have also identified a few results that are biologically compelling, although replication of the SNP association is warranted. For example, we found that a SNP (rs1158167) near the *CST3* gene was highly associated with serum cystatin-C levels (GEE p-value = $8.5*10^{-09}$) [42]. This SNP explained 2.5% of the variation in serum cystatin-C levels in our data and has been previously reported to be associated with cystatin-C. These results are presented in more detail in the Renal Endocrine Working Group manuscript [42].

Replication of results from other genome-wide studies

While we have been preparing this series of manuscripts, several genome-wide studies have been published [80-86]. Some results in our analyses support results reported in these recent studies. For example, we find significant associations for coronary heart disease, cardiovascular disease and coronary artery calcium [29,30] in the same chromosomal region on 9p recently reported to be associated with myocardial infarction by Helgadottir *et al.* [82] and McPherson *et al.* [83]. While our results need to be compared more closely with results being reported by other genome-wide studies, this example provides evidence that our results replicate strong associations from other genome-wide studies.

Discussion

We have presented a brief description of the methods and a few selected results derived from analyses of the 100K Affymetrix GeneChip with a large number of FHS traits, ranging from CVD events and subclinical measures to traditional cardiovascular risk factors of diabetes, lipid levels, blood pressure and also including more novel biomarker measures that reflect modern hypotheses, such as the role of inflammatory pathways in the development of CVD. We have also reported on a number of neurological, renal, cancer and aging traits, including longevity (age at death) and bone mass and structure. None of these manuscripts provide a comprehensive report. Rather, the purpose of this set of manuscripts is to provide a brief

summary of the results and to introduce readers to the data posted on the dbGaP website http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007. We note that the genotypes in this sample have also been evaluated by Drs. Michael Christman and Alan Herbert. Some of their results are reported on line as described by Herbert *et al*. [87].

Several aspects of our investigation merit comment. **First**, the present investigation represents a comprehensive GWAS analysis of numerous phenotypes in a large community-based cohort. To our knowledge, it is the largest GWAS performed in an observational cohort in terms of the number of phenotypes analyzed and web posted. **Second**, we exploited the phenotypic diversity and richness of the Framingham Offspring Study database to analyze a set of phenotypes that were for the most part collected by detailed, direct measurements of study participants. Further, many of the phenotypes are quantitative traits. Phenotypes have been broadly categorized into seventeen different domains for manuscripts in this supplement. It is noteworthy that key risk factor phenotypes, such as blood pressure and lipid levels, were collected at multiple examinations, and thus we were able to conduct analyses using time-averaged traits, maximizing the scientific yield from the longitudinal prospective design of our cohort study. Further, several recently collected phenotypes, in particular biomarkers and imaging measures, were collected using highly reproducible, state-of-the-art modalities. Correlated phenotypes facilitated the assessment of pleiotropy by seeking associations of SNPs with such phenotypes. These investigations occurred primarily among the variables in each individual manuscript. Finally, for most phenotypes, there was evidence for a significant heritable component from FHS or other studies. We acknowledge that some phenotypic domains may represent analytical constructs, rather than truly distinct groups from a biological standpoint.

Third, we have web-posted the results of all analyses on autosomes on dbGaP, including results without statistical evidence of association, so that investigators world-wide can access the data freely and mine them *in silico* for hypothesis generation, inclusion in meta-analysis, and direct comparisons with their own results. In addition to the freely posted aggregate results, participant-specific genotypic and phenotypic data are available for distribution for further analyses to approved scientific investigators world-wide via the NCBI/NHLBI and consistent with Framingham Study data distribution policies (see http://www.nhlbi.nih.gov/about/framingham/policies/index.htm). For the purpose of publication, reference to these analyses may be made by referring to either the appropriate manuscript or the specific URL for web-posted data. **Fourth**, the simultaneous and full-disclosure release (on the web) of all association and linkage results of phenotypes encompassing at least 17 different domains in a cohesive and comprehensive manner signifies the tremendous teamwork of numerous FHS investigators, statisticians, programmers, and others. Most importantly, this effort would not be possible without the full cooperation and commitment of the FHS participants, who continue to attend Study examinations in an effort to further the scientific knowledge of factors that lead to heart disease and other traits.

Fifth, as in any genome-wide association study with a large number of SNPs, most results that are considered statistically significant by a conventional $p < 0.05$ may be falsely positive; so it is difficult to decide what results are important. Not only do we have a large number of statistical tests for each phenotype, but we also have numerous phenotypes. Thus, considering multiple testing in the interpretation of results is of paramount importance. There are several approaches to address the issue of multiple testing, such as Bonferroni correction, permutation testing and false discovery rates. To conduct permutation testing for all of the traits that we considered is prohibitively time-consuming, particularly in preserving heritability of the traits with family data. Further, with correlated traits it is difficult to decide what traits should be included in a permutation testing strategy. One approach to controlling the false-positive rate in genome-wide association studies is to set a stringent threshold for declaring statistical significance. According to the report of the International HapMap Consortium, complete testing of common variants (MAF > 0.05) in each 500 kb is equivalent to performing 150 independent tests in white populations of European descent [88]. Using this guide and given that there are about 3000 Mb in the human genome, we would estimate that there are approximately 900,000–1,000,000 independent tests if testing all common variants in the genome. A conservative Bonferroni correction using this number of tests (0.05/1,000,000) yields an approximate threshold of genome-wide significance to be $5*10^{-8}$. Thus, for a single trait, one could use this threshold. Several results do fall below this threshold (Table 5 of the Overview). In considering these results, we note that our sample size of 1345 biologically related subjects is relatively small for detecting genetic variants of modest effect. We also note that we have a large number of correlated traits, including the same traits with different covariate adjustments. Further, we have already observed that our GEE results have an excess number of small p-values. Thus, we are hesitant to regard any result reported in our manuscripts as significant at a genome-wide level. We believe these findings are best regarded as hypothesis-generating. The determination of what constitutes genome-wide significance is challenged both by theoretical considerations as well as practical ones. Without pursuing more

computationally intensive analyses, it is thus difficult to provide specific advice regarding what SNPs are most important. **It may be safer to assume that most of the small p-values are likely to be false positives and that replication of our results in other independent samples is of critical importance**. We proceed with presentation of full-disclosure results to encourage readers to pursue such studies.

Sixth, we note that use of the 80% genotyping call rate is unusually liberal by today's standards in GWAS. We used this threshold in these manuscripts to be inclusive, rather than exclusive, in a first look such as this. We recognize that this threshold may permit consideration of some results that could be spurious due problems with genotyping. However, a limitation of our genotypes is that the genotype calls were made with the DM algorithm, which is less precise than those that have recently been introduced. At this time, we are unable to apply more accurate, reliable genotyping calls [89], as we do not have access to the source data. Further, we found that the choice of the 80% threshold versus a more conservative one had little effect upon p-value distributions. Finally, all results, regardless of genotyping call rate, are posted on the dbGaP website and thus, investigators can evaluate for themselves what they believe to be the more valid results from this study.

Seventh, in our analyses we found that the GEE results appear to have an excess of significant results. We suspect that one reason is low MAF. Also, given the small sample of at most 1345 subjects, we would expect only 13–14 individuals to have the minor homozygote. Thus, we limited the results that we present in the manuscripts to those SNPs with MAF = 10%. Further analyses have indicated that use of a linear mixed effects model such as incorporating a SNP as a covariate in a regression model with proper correlation structure for the error terms that fully represent the familial correlations remedies this problem and has a valid type I error rate in simulated data.

Eighth, coverage of LD is incomplete with the 100K scan. Nicolae *et al*. report that the Affymetrix 100K GeneChip includes fewer SNPs in coding and more SNPs in intergenic regions than represented on the HapMap [90]. Further, our sample size is modest. These two facts combined likely limit the power for detection of associations with several traits in these data. For instance, while we noted modest to high heritability of numerous phenotypes, underscoring the contribution of additive genetic effects to interindividual variation in these traits, we did not find significant low p-values for several heritable traits in relation to the SNPs evaluated. Factors contributing to this observation included both the limited coverage of the Affymetrix 100K GeneChip as well as the possibility that some of the less significant p-values (example between 0.05 and 10^{-5}) may represent true positive findings. The limited power to detect SNPs of small effect sizes offered by the analysis of our relatively modest sample size of ~1300 participants contributes to this phenomenon as well; we only have high power to detect a SNP explaining 4% or more of the phenotypic variance in the population-based GEE association test; the power of FBAT and variance component linkage analysis is even lower.

Additionally, for several of the analyzed phenotypes we did not observe any overlap between the top SNP-phenotype associations noted in GEE and FBAT analyses. The inherent differences in the two analytical methods especially in the context of the modest sample sizes, particularly for FBAT with small numbers of informative trios, may contribute to this phenomenon. FBAT is limited by the number of informative transmissions and although we suspect that there is little population stratification in our sample [76], GEE is limited by potential bias due to stratification. Furthermore, for several phenotypes the SNPs associated with the top LOD scores in linkage analyses were not among the top 50 SNPs in association analyses (GEE or FBAT).

Table 5: Associations achieving nominal genome wide significance, $p < 5*10^{-8}$ across the 17 phenotype working groups

Phenotype working group/manuscript	Trait	SNP rs ID*	Chr	Physical location (bp)	GEE P-value	FBAT P-value	IN/NEAR gene
Select biomarkers [33]	Monocyte chemoattractant protein-1	rs2494250	1	156,091,324	$1.0*10^{-14}$	$3.5*10^{-8}$	*FCER1A, OR10J3*
	Monocyte chemoattractant protein-1	rs4128725	1	156,219,032	$3.7*10^{-12}$	$3.3*10^{-8}$	*OR10J1*
	C-reactive protein average exams 2,6,7	rs2794520	1	156,491,889	$2.8*10^{-8}$	$4.3*10^{-5}$	*CRP*
	C-reactive protein average exams 2,6,7	rs2808629	1	156,489,869	$3.2*10^{-8}$	$4.8*10^{-5}$	*CRP*
Kidney/Endocrine [42]	Cystatin C	rs1158167	20	23,526,189	$8.5*10^{-09}$	0.006	*CST9L|CST9|CST3*
Diabetes [38]	28-year mean fasting plasma glucose	rs2722425	8	40,603,396	$2.0*10^{-8}$	0.005	*ZMAT4*
Sleep and circadian [26]	Epworth sleepiness scale	rs1823068	5	58,711,806	$2.5*10^{-8}$	0.069	*PDE4D*
Neurology [37]	Total Cerebral Brain Volume (ATCBV)	rs1970546	20	59287333	$4.0*10^{-8}$	0.005	*CDH4*
Hemostatic factors [34]	Factor VII	rs561241	13	112,808,035	$4.5*10^{-16}$	$3.4*10^{-4}$	*F7*

The following phenotype working groups did not have any traits achieving nominal genome-wide significance: echocardiography, flow-mediated dilation and exercise tolerance testing; blood pressure and tonometry; subclinical cardiovascular disease; cardiovascular outcomes; cancer; electrocardiography and heart rate variability; pulmonary function testing; aging; bone; lipids; obesity

Ninth, we were limited in our ability to replicate genetic variants previously reported to be associated with phenotypes in our database because specific coverage of such genetic variation in these candidates was limited in the Affymetrix 100K GeneChip. We view such analyses as more illustrative of the potential utility of our GWAS, rather than as definitive evidence for or against an association described with a putative candidate gene in the published literature.

Our data do suggest several interesting biological candidates among the SNPs most strongly associated with different traits in the various analytical approaches. The strongest and most clear-cut of the associations were for those phenotypes that represent the direct protein product of a gene. Examples include the association of CRP concentrations with SNPs in the *CRP* gene (Benjamin *et al*. in this series [33]) and factor VII levels with SNP rs561241 on chromosome 13 (Yang *et al*. in this series [34]). Thus, while it is difficult to point to any result as definitive, those results for which we find some evidence of replication of associations found in the literature are regarded as worthy of further research.

Finally, the Framingham Study participants were white of European descent and predominantly middle-aged to elderly. Hence, the genetic associations may not be generalizable to other ethnicities/races or to younger individuals.

Conclusion

In summary, the results from the FHS 100K association and linkage studies described herein and posted on the NCBI website provide a GWAS resource for investigators. We have presented a description of the methods and general strategies used for analysis of the 100K Affymetrix GeneChip in relation to a broad range of traits measured in the FHS. Brief descriptions of results of these analyses are provided a series of 17 manuscripts, with results for all autosomal SNPs genotyped successfully displayed at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007. Interested investigators can also access the data through a standing protocol, described at http://www.nhlbi.nih.gov/about/framingham/policies/index.htm. Key to interpretation of these results is replication and evaluation of these results in other cohorts and ultimately, functional studies. We encourage investigators to examine the results and to pursue the genetic signals therein in their own cohorts. In the near future we will provide results and data from approximately 550,000 SNPs on more than 9000 participants from three generations in the FHS SNP Health Association Resource (SHARe) project. Data will be available to qualified investigators through an application process to dbGaP. It is our hope that the results from these two genome-wide association studies will lead to a much deeper understanding of the role of common genetic variation in the development of cardiovascular disease and its risk factors.

Abbreviations

FBAT = family-based association test; GEE = generalized estimating equations; GWAS = Genome wide association study; IBD = identity by descent; LD = linkage disequilibrium; LOD = logarithm of the odds; SNP = single nucleotide polymorphism.

Competing interests

Dr. Meigs currently has research grants from GlaxoSmithKline and Wyeth, and serves on safety or advisory boards for GlaxoSmithKline, Merck, and Lilly.

Authors' contributions

All authors have made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; LAC, EJB, SD, ALD, JD, KF, CSF, DJG, SK, DPK, MGL, DL, KLL, MDM, JBM, CNC, GTO, CJO, SS, RSV, JBW, QY and LDA have been involved in drafting the manuscript or revising it critically for important intellectual content; and All authors have given final approval of the version to be published.

In addition, MDM, ZYW: imported, performed QC, processed, organized, and displayed all results within the NCBI dbGaP.

Additional material

Additional file 1

Phenotypes for population-based GEE analyses.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2350-8-S1-S1-S1.pdf]

Additional file 2

Phenotypes for family-based FBAT analyses.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2350-8-S1-S1-S2.pdf]

Additional file 3

Phenotypes for linkage analyses.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2350-8-S1-S1-S3.pdf]

Acknowledgements

We are most grateful to the Framingham Heart Study participants who have committed so much of their time and devotion to this research. This research was support by a contract from NHLBI N01-HC 25195 (LAC, HA, EJB, RBD, ALD, JD, KF, DRG, CYG, NHC, MGL, CYL, KLL, JMM, MP, RSV, JBW, PAW, QY, LDA), by NHLBI K24 HL 04334 (RSV), by NIA 5R01-

AG08122 (SS, PAW), by NIA 5R01-AG16495 (SS, PAW, LDA), by NINDS 5R01-NS17950 (SS, PAW, ALD, EJB), by NHLBI RO1 HL076784 and NIA 1R01 AG028321 (EJB, JD, KL, MGL, RS, VSR), by NHLBI HL54776 (LAC, SD, AKM) and by an American Diabetes Association Career Development Award (JBM, JD, LAC, AKM, CYL). A portion of the research was conducted using the Boston University Linux Cluster for Genetic Analysis (LinGA) funded by the NIH NCRR (National Center for Research Resources) Shared Instrumentation grant 1S10RR163736-01A1.

This article has been published as part of *BMC Medical Genetics* Volume 8 Supplement 1, 2007: The Framingham Heart Study 100,000 single nucleotide polymorphisms resource. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2350/8?issue=S1.

References

1. Rosamond W, Flegal K, Friday G, Furie K, Go A, Greenlund K, Haase N, Ho M, Howard V, Kissela B, Kittner S, Lloyd-Jones D, McDermott M, Meigs J, Moy C, Nichol G, O'Donnell CJ, Roger V, Rumsfeld J, Sorlie P, Steinberger J, Thom T, Wasserthiel-Smoller S, Hong Y, American Heart Association Statistics Committee and Stroke Statistics Subcommittee: **Heart disease and stroke statistics – 2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee.** *Circulation* 2007, **115(5):**e69-171.
2. Myers RH, Kiely DK, Cupples LA, Kannel WB: **Parental history is an independent risk factor for coronary artery disease: the Framingham Study.** *Am Heart J* 1990, **120(4):**963-969.
3. Lloyd-Jones DM, Nam BH, D'Agostino RB Sr, Levy D, Murabito JM, Wang TJ, Wilson PWJ, O'Donnell CJ: **Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring.** *JAMA* 2004, **291:**2204-2211.
4. Murabito JM, Pencina MJ, Nam BH, D'Agostino RB Sr, Wang TJ, Lloyd-Jones D, Wilson PW, O'Donnell CJ: **Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults.** *JAMA* 2005, **294:**3117-3123.
5. O'Donnell CJ, Chazaro I, Wilson PW, Fox C, Hannan MT, Kiel DP, Cupples LA: **Evidence for heritability of abdominal aortic calcific deposits in the Framingham Heart Study.** *Circulation* 2002, **106(3):**337-341.
6. Fox CS, Polak JF, Chazaro I, Cupples A, Wolf PA, D'Agostino RA, O'Donnell CJ: **Genetic and environmental contributions to atherosclerosis phenotypes in men and women: heritability of carotid intima-media thickness in the Framingham Heart Study.** *Stroke* 2003, **34(2):**397-401.
7. Murabito JM, Guo C-Y, Fox CS, D'Agostino RB Sr: **Genetic Contributions to Peripheral Arterial Disease: Heritability of the Ankle-Brachial Blood Pressure Index in the Framingham Heart Study.** *Am J Epidemiol* 2006, **164(10):**963-968.
8. Fox CS, Parise H, D'Agostino RB Sr, Lloyd-Jones DM, Vasan RS, Wang TJ, Levy D, Wolf PA, Benjamin EJ: **Parental atrial fibrillation as a risk factor for atrial fibrillation in offspring.** *JAMA* 2004, **291:**2851-2855.
9. Shearman AM, Cooper JA, Kotwinski PJ, Miller GJ, Humphries SE, Ardlie KG, Jordan B, Irenze K, Lunetta KL, Schuit SC, Uitterlinden AG, Pols HA, Demissie S, Cupples LA, Mendelsohn ME, Levy D, Housman DE: **Estrogen receptor alpha gene variation is associated with risk of myocardial infarction in more than seven thousand men from five cohorts.** *Circ Res* 2006, **98(5):**590-592.
10. Samani NJ, Burton P, Mangino M, Ball SG, Balmforth AJ, Barrett J, Bishop T, Hall A, BHF Family Heart Study Research Group: **A genomewide linkage study of 1,933 families affected by premature coronary artery disease: The British Heart Foundation (BHF) Family Heart Study.** *Am J Hum Genet* 2005, **77(6):**1011-1020.
11. Wang L, Hauser ER, Shah SH, Pericak-Vance MA, Haynes C, Crosslin D, Harris M, Nelson S, Hale AB, Granger CB, Haines JL, Jones CJ, Crossman D, Seo D, Gregory SG, Kraus WE, Goldschmidt-Clermont PJ, Vance JM: **Peakwide mapping on chromosome 3q13 identifies the kalirin gene as a novel candidate gene for coronary artery disease.** *Am J Hum Genet* 2007, **80(4):**650-663.
12. Shearman AM, Cupples LA, Demissie S, Peter I, Schmid CH, Karas RH, Mendelsohn ME, Housman DE, Levy D: **Association between estrogen receptor alpha gene variation and cardiovascular disease.** *Jama* 2003, **290(17):**2263-2270.
13. Lahoz C, Schaefer EJ, Cupples LA, Wilson PW, Levy D, Osgood D, Parpos S, Pedro-Botet J, Daly JA, Ordovas JM: **Apolipoprotein E genotype and cardiovascular disease in the Framingham Heart Study.** *Atherosclerosis* 2001, **154(3):**529-537.
14. O'Donnell CJ, Lindpaintner K, Larson MG, Rao VS, Ordovas JM, Schaefer EJ, Myers RH, Levy D: **Evidence for association and genetic linkage of the angiotensin-converting enzyme locus with hypertension and blood pressure in men but not women in the Framingham Heart Study.** *Circulation* 1998, **97(18):**1766-1772.
15. Levy D, DeStefano AL, Larson MG, O'Donnell CJ, Lifton RP, Gavras H, Cupples LA, Myers RH: **Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study.** *Hypertension* 2000, **36(4):**477-483.
16. DeStefano AL, Larson MG, Mitchell GF, Benjamin EJ, Vasan RS, Li J, Corey D, Levy D: **Genome-wide scan for pulse pressure in the National Heart, Lung and Blood Institute's Framingham Heart Study.** *Hypertension* 2004, **44(2):**152-155.
17. Shearman AM, Ordovas JM, Cupples LA, Schaefer EJ, Harmon MD, Shao Y, Keen JD, DeStefano AL, Joost O, Wilson PW, Housman DE, Myers RH: **Evidence for a gene influencing the TG/HDL-C ratio on chromosome 7q32.3-qter: a genome-wide scan in the Framingham study.** *Hum Mol Genet* 2000, **9(9):**1315-1320.
18. Lai CQ, Demissie S, Cupples LA, Zhu Y, Adiconis X, Parnell LD, Corella D, Ordovas JM: **Influence of the APOA5 locus on plasma triglyceride, lipoprotein subclasses, and CVD risk in the Framingham Heart Study.** *J Lipid Res* 2004, **45(11):**2096-2105.
19. Yang Q, Lai CQ, Parnell L, Cupples LA, Adiconis X, Zhu Y, Wilson PW, Housman DE, Shearman AM, D'Agostino RB, Ordovas JM: **Genome-wide linkage analyses and candidate gene fine mapping for HDL3 cholesterol: the Framingham Study.** *J Lipid Res* 2005, **46(7):**1416-1425.
20. Demissie S, Cupples LA, Shearman AM, Gruenthal KM, Peter I, Schmid CH, Karas RH, Housman DE, Mendelsohn ME, Ordovas JM: **Estrogen receptor-alpha variants are associated with lipoprotein size distribution and particle levels in women: the Framingham Heart Study.** *Atherosclerosis* 2006, **185(1):**210-218.
21. Meigs JB, Cupples LA, Wilson PW: **Parental transmission of type 2 diabetes: the Framingham Offspring Study.** *Diabetes* 2000, **49(12):**2201-2207.
22. Meigs JB, Panhuysen CI, Myers RH, Wilson PW, Cupples LA: **A genome-wide scan for loci linked to plasma levels of glucose and HbA(1c) in a community-based sample of Caucasian pedigrees: The Framingham Offspring Study.** *Diabetes* 2002, **51(3):**833-840.
23. Panhuysen CI, Cupples LA, Wilson PW, Herbert AG, Myers RH, Meigs JB: **A genome scan for loci linked to quantitative insulin traits in persons without diabetes: the Framingham Offspring Study.** *Diabetologia* 2003, **46(4):**579-587.
24. Atwood LD, Heard-Costa NL, Fox CS, Jaquish CE, Cupples LA: **Sex and age specific effects of chromosomal regions linked to body mass index in the Framingham Study.** *BMC Genet* 2006, **7:**7.
25. Atwood LD, Heard-Costa NL, Cupples LA, Jaquish CE, Wilson PW, D'Agostino RB: **Genomewide linkage analysis of body mass index across 28 years of the Framingham Heart Study.** *Am J Hum Genet* 2002, **71(5):**1044-1050.
26. Gottlieb DJ, O'Connor GT, Wilk JB: **Genome-wide association of sleep and circadian phenotypes.** *BMC Med Genet* 2007, **8(Suppl 1):**S9.
27. Newton-Cheh C, Guo C-Y, Wang TJ, O'Donnell CJ, Levy D, Larson MG: **Genome-wide association study of electrocardiographic and heart rate variability traits: the Framingham Heart Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S7.
28. Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang SJ, Vasan RS, Mitchell GF: **Framingham Heart Study 100K Project: Genome-wide associations for blood pressure and arterial stiffness.** *BMC Med Genet* 2007, **8(Suppl 1):**S3.
29. O'Donnell CJ, Cupples LA, D'Agostino RB Sr, Fox CS, Hoffmann U, Hwang S-J, Ingellson E, Liu C-Y, Murabito JM, Polak JF, Wolf PA, Demissie S: **Genome-wide association study for subclinical**

atherosclerosis in major arterial territories in the NHLBI's Framingham Heart Study. *BMC Med Genet* 2007, **8(Suppl 1):**S4.
30. Larson MG, Atwood LD, Benjamin EJ, Cupples LA, D'Agostino RB Sr, Fox CS, Govindaraju DR, Guo CY, Heard-Costa NL, Hwang SJ, Murabito JM, Newton-Cheh C, O'Donnell CJ, Seshadri S, Vasan RS, Wang TJ, Wolf PA, Levy D: **Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes.** *BMC Med Genet* 2007, **8(Suppl 1):**S5.
31. Murabito JM, Rosenberg CL, Finger D, Kreger BE, Levy D, Splansky GL, Antman K, Hwang S-J: **A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S6.
32. Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT: **Framingham Heart Study genome-wide association: results for pulmonary function measures.** *BMC Med Genet* 2007, **8(Suppl 1):**S8.
33. Benjamin EJ, Dupuis J, Larson MG, Lunetta KL, Booth SL, Govindaraju DR, Kathiresan S, Keaney JF Jr, Keyes MJ, Lin JP, Meigs JB, Robins SJ, Rong J, Schnabel R, Vita JA, Wang TJ, Wilson PWF, Wolf PA, Vasan RS: **Genome-wide association with select biomarker traits in the Framingham Heart Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S11.
34. Yang Q, Kathiresan S, Lin JP, Tofler GH, O'Donnell CJ: **Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S12.
35. Lunetta KL, D'Agostino RB Sr, Karasik D, Benjamin EJ, Guo C-Y, Govindaraju R, Kiel DP, Kelly-Hayes M, Massaro JM, Pencina MJ, Seshadri S, Murabito JM: **Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S13.
36. Kiel DP, Demissie S, Dupuis J, Lunetta KL, Murabito JM, Karasik D: **Genome-wide association with bone mass and geometry in the Framingham Heart Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S14.
37. Seshadri S, DeStefano AL, Au R, Massaro JM, Beiser AS, Kelly-Hayes M, Kase CS, D'Agostino RB Sr, DeCarli C, Atwood LD, Wolf PA: **Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S15.
38. Meigs JB, Manning AK, Fox CS, Florez JC, Liu C-Y, Cupples LA, Dupuis J: **Genome-wide association with diabetes-related traits in the Framingham Heart Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S16.
39. Kathiresan S, Manning AK, Demissie S, D'Agostino RB Sr, Surti A, Guiducci C, Gianniny L, Burtt NP, Melander O, Orho-Melander M, Arnett D, Peloso GM, Ordovas JM, Cupples LA: **A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S17.
40. Fox CS, Heard-Costa N, Cupples LA, Dupuis J, Vasan RS, Atwood LD: **Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project.** *BMC Med Genet* 2007, **8(Suppl 1):**S18.
41. Vasan RS, Larson MG, Aragam J, Wang TJ, Mitchell GF, Kathiresan S, Newton-Cheh C, Vita JA, Keyes MJ, O'Donnell CJ, Levy D, Benjamin EJ: **Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the Framingham Heart Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S2.
42. Hwang S-J, Yang Q, Meigs JB, Pearce EN, Fox CS: **A Genome-Wide Association for Kidney Function and Endocrine-Related Traits in the NHLBI's Framingham Heart Study.** *BMC Med Genet* 2007, **8(Suppl 1):**S10.
43. Dawber T: **The Framingham Study: The Epidemiology of Atherosclerotic Disease.** *Cambridge, Massachusetts: Harvard University Press*; 1980.
44. Dawber TRMG, Moore FE: **Epidemiologic approaches to heart disease: the Framingham study.** *Am J Public Health* 1951, **41:**279-286.
45. Dawber TRKW, Lyell LP: **An approach to longitudinal studies in a community: the Framingham Heart Study.** *Ann N Y Acad Sci* 1963, **107:**539-556.
46. Cupples LA, D'Agostino RB: **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 *The Framingham Heart Study: An Epidemiological Investigation of Cardiovascular Disease Volume NIH publication 87-203*. Edited by: *Kannel WB, Wolf PA, Garrison RJ. Washington, DC: Government Printing Office*; 1987.
47. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP: **The Framingham Offspring Study: design and preliminary data.** *Prev Med* 1975, **4:**518-525.
48. 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.
49. Cupples LA, Yang Q, Demissie S, Copenhafer D, Levy D: **Description of the Framingham Heart Study data for Genetic Analysis Workshop 13.** *BMC Genet* 2003, **4(Suppl 1):**S2.
50. Yang Q, Chazaro I, Cui J, Guo CY, Demissie S, Larson M, Atwood LD, Cupples LA, DeStefano AL: **Genetic analyses of longitudinal phenotype data: a comparison of univariate methods and a multivariate approach.** *BMC Genet* 2003, **4(Suppl 1):**S29.
51. Fox CS, Yang Q, Cupples LA, Guo CY, Larson MG, Leip EP, Wilson PW, Levy D: **Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study.** *J Am Soc Nephrol* 2004, **15(9):**2457-2461.
52. Fox CS, Cupples LA, Chazaro I, Polak JF, Wolf PA, D'Agostino RB, Ordovas JM, O'Donnell CJ: **Genomewide linkage analysis for internal carotid artery intimal medial thickness: evidence for linkage to chromosome 12.** *Am J Hum Genet* 2004, **74(2):**253-261.
53. Karasik D, Hannan MT, Cupples LA, Felson DT, Kiel DP: **Genetic contribution to biological aging: the Framingham Study.** *J Gerontol A Biol Sci Med Sci* 2004, **59(3):**218-226.
54. Karasik D, Myers RH, Cupples LA, Hannan MT, Gagnon DR, Herbert A, Kiel DP: **Genome screen for quantitative trait loci contributing to normal variation in bone mineral density: the Framingham Study.** *J Bone Miner Res* 2002, **17(9):**1718-1727.
55. Dupuis J, Larson MG, Vasan RS, Massaro JM, Wilson PW, Lipinska I, Corey D, Vita JA, Keaney JF Jr, Benjamin EJ: **Genome scan of systemic biomarkers of vascular inflammation in the Framingham Heart Study: evidence for susceptibility loci on 1q.** *Atherosclerosis* 2005, **182(2):**307-314.
56. Murabito JM, Yang Q, Fox CS, Cupples LA: **Genome-wide linkage analysis to age at natural menopause in a community-based sample: the Framingham Heart Study.** *Fertil Steril* 2005, **84(6):**1674-1679.
57. Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, Wichmann HE, Meitinger T, Hunter D, Hu FB, Colditz G, Hinney A, Hebebrand J, Koberwitz K, Zhu X, Cooper R, Ardlie K, Lyon H, Hirschhorn JN, Laird NM, Lenburg ME, Lange C, Christman MF: **A common genetic variant is associated with adult and childhood obesity.** *Science* 2006, **312(5771):**279-283.
58. Di X, Matsuzaki H, Webster TA, Hubbell E, Liu G, Dong S, Bartell D, Huang J, Chiles R, Yang G, Shen MM, Kulp D, Kennedy GC, Mei R, Jones KW, Cawley S: **Dynamic model based algorithms for screening and genotyping over 100K SNPs on oligonucleotide microarrays.** *Bioinformatics* 2005, **21(9):**1958-1963.
59. Weber JL, May PE: **Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction.** *Am J Hum Genet* 1989, **44:**388-396.
60. Yuan BVD, Weber JL, Beck J, Sheffield VC: **Improved set of short-tandem-repeat polymorphisms for screening the human genome [letter].** *Am J Hum Genet* 1997, **60:**459-460.
61. Hinds DA, Risch N: **The ASPEX package: affected sib-pair exclusion mapping.** 1996 [http://aspex.sourceforge.net/]. *Hinds, DA: web site*
62. O'Connell JR, Weeks DE: **PedCheck: a program for identification of genotype incompatibilities in linkage analysis.** *Am J Hum Genet* 1998, **63:**259-266.
63. Emigh T: **On the number of observed classes from a multinomial distribution.** *Biometrics* 1983, **39:**485-491.
64. Warnes G, Leisch F: **Genetics: Population Genetics. R package Version 1.1.4.** 2005.
65. R_Development_Core_Team: **R: A language and environment for statistical computing.** *Vienna, Austria: R Foundation for Statistical Computing*; 2005.
66. Huang Q, Shete S, Amos CI: **Ignoring linkage disequilibrium among tightly linked markers induces false-positive evidence**

of linkage for affected sib pair analysis. *Am J Hum Genet* 2004, **75(6):**1106-1112.
67. Huang Q, Shete S, Swartz M, Amos CI: **Examining the effect of linkage disequilibrium on multipoint linkage analysis.** *BMC Genet* 2005, **6(Suppl 1):**S83.
68. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: **Merlin – rapid analysis of dense genetic maps using sparse gene flow trees.** *Nat Genet* 2002, **30:**97-101.
69. Almasy L, Blangero J: **Multipoint quantitative-trait linkage analysis in general pedigrees.** *Am J Hum Genet* 1998, **62(5):**1198-1211.
70. Rabinowitz D, Laird N: **A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information.** *Hum Hered* 2000, **50(4):**211-223.
71. Laird NM, Horvath S, Xu X: **Implementing a unified approach to family-based tests of association.** *Genet Epidemiol* 2000, **19(Suppl 1):**S36-42.
72. Ewens WJ, Spielman RS: **What is the significance of a significant TDT?** *Hum Hered* 2005, **60(4):**206-210.
73. Ewens WJ, Spielman RS: **The transmission/disequilibrium test: history, subdivision, and admixture.** *Am J Hum Genet* 1995, **57(1):**455-464.
74. Liang KY, Zeger SL: **Longitudinal Data Analysis using Generalized Linear Models.** *Biometrika* 1986, **73:**13-22.
75. Carey VJ, ported to R by Thomas Lumley (versions 3.13 and 4.4) and Brian Ripley (version 4.13): **gee: Generalized Estimation Equation solver. R package Version 4.13-10.** 2002.
76. Wilk JB, Manning AK, Dupuis J, Cupples LA, Larson MG, Newton-Cheh C, Demissie S, DeStefano AL, Hwang SJ, Liu C, Yang Q, Lunetta KL: **No evidence of major population substructure in the Framingham Heart Study. Abstract 171 from the 14th Annual Meeting of the International Genetic Epidemiology Society.** *Genet Epidemiol* 2005, **29:**286.
77. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, Nutland S, Howson JM, Faham M, Moorhead M, Jones HB, Falkowski M, Hardenbol P, Willis TD, Todd JA: **Population structure, differential bias and genomic control in a large-scale, case-control association study.** *Nat Genet* 2005, **37(11):**1243-1246.
78. Kathiresan S, Yang Q, Larson MG, Camargo AL, Tofler GH, Hirschhorn JN, Gabriel SB, O'Donnell CJ: **Common genetic variation in five thrombosis genes and relations to plasma hemostatic protein level and cardiovascular disease risk.** *Arterioscler Thromb Vasc Biol* 2006, **26(6):**1405-1412.
79. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, Katayama T, Baldwin CT, Cheng R, Hasegawa H, Chen F, Shibata N, Lunetta KL, Pardossi-Piquard R, Bohm C, Wakutani Y, Cupples LA, Cuenco KT, Green RC, Pinessi L, Rainero I, Sorbi S, Bruni A, Duara R, Friedland R, Inzelberg R, Hampe W, Bujo H, Song Y, Andersen O, Graff-Radford N, Petersen R, Dickson D, Der SD, Fraser PE, Schmitt-Ulms G, Younkin S, Mayeux R, Farrer LA, St George-Hyslop P: **The sortilin-related receptor SORL1 is genetically associated with Alzheimer's Disease.** *Nat Genet* 2007, **39(2):**168-177.
80. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT LU, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S: **Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.** *Science* 2007, **316(5829):**1331-1336.
81. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: **A genome-wide association study identifies novel risk loci for type 2 diabetes.** *Nature* 2007, **445(7130):**881-885.
82. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson D, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K: **A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction.** *Science* 2007, **316(5830):**1491-1493.
83. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC: **A Common Allele on Chromosome 9 Associated with Coronary Heart Disease.** *Science* 2007, **316(5830):**1488-1491.
84. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorradottir S, Bjarnason H, Ng MC, Hansen T, Bagger Y, Wilensky RL, Reilly MP, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Doumatey A, So WY, Ma RC, Andersen G, Borch-Johnsen K, Jorgensen T, van Vliet-Ostaptchouk JV, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Rotimi C, Gurney M, Chan JC, Pedersen O, Sigurdsson G, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K: **A variant in CDKAL1 influences insulin response and risk of type 2 diabetes.** *Nat Genet* 2007, **9(6):**770-775.
85. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: **A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.** *Science* 2007, **5829(5829):**1341-1345.
86. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, AS D, Wellcome Trust Case Control Consortium (WTCCC), Burton PR, Clayton DG, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, Ouwehand WH, Samani NJ, Todd JA, Donnelly P, Davison D, Easton D, Evans D, Leung HT, Spencer CC, Tobin MD, Attwood AP, Boorman JP, Cant B, Everson U, Hussey JM, Jolley JD, Knight AS, Koch K, Meech E, Nutland S, Prowse CV, Stevens HE, Taylor NC, Walters GR, Walker NM, Watkins NA, Winzer T, Jones RW, McArdle WL, Ring SM, Strachan DP, Pembrey M, Breen G, St Clair D, Caesar S, Gordon-Smith K, Jones L, Fraser C, Green EK, Grozeva D, Hamshere ML, Holmans PA, Jones IR, Kirov G, Moskvina V, Nikolov I, O'donovan MC, Owen MJ, Collier DA, Elkin A, Farmer A, Williamson R, McGuffin P, Young AH, Ferrier IN, Ball SG, Balmforth AJ, Barrett JH, Bishop DT, Iles MM, Maqbool A, Yuldasheva N, Hall AS, Braund PS, Dixon RJ, Mangino M, Stevens S, Thompson JR, Bredin F, Tremelling M, Parkes M, Drummond H, Lees CW, Nimmo ER, Satsangi J, Fisher SA, Forbes A, Lewis CM, Onnie CM, Prescott NJ, Sanderson J, Mathew CG, Barbour J, Mohiuddin MK, Todhunter CE, Mansfield JC, Ahmad T, Cummings FR, Jewell DP, Webster J, Brown MJ, Lathrop GM, Connell J, Dominiczak A, Braga Marcano CA, Burke B, Dobson R, Gungadoo J, Lee KL, Munroe PB, Newhouse SJ, Onipinla A, Wallace C, Xue M, Caulfield M, Farrall M, Barton A, Bruce IN, Donovan H, Eyre S, Gilbert PD, Hider SL, Hinks AM, John SL, Potter C, Silman AJ, Symmons DP, Thomson W, Worthington J, Dunger DB, Widmer B, Newport M, Sirugo G, Lyons E, Vannberg F, Hill AV, Bradbury LA, Farrar C, Pointon JJ, Wordsworth P, Brown MA, Franklyn JA, Heward JM, Simmonds MJ, Gough SC, Seal S, Stratton MR, Rahman N, Ban M, Goris A, Sawcer SJ, Compston A, Conway D, Jallow M, Rockett KA, Bumpstead SJ, Chaney A, Downes K, Ghori MJ, Gwilliam R, Hunt SE, Inouye M, Keniry A, King E, McGinnis R, Potter S, Ravindrarajah R, Whittaker P, Widden C, Withers D, Cardin NJ, Ferreira T, Pereira-Gale J, Hallgrimsdottir IB, Howie BN, Su Z, Teo YY, Vukcevic D, Bentley D, Compston A, Ouwehand NJ, Samani MR, Isaacs JD, Morgan AW, Wilson GD, Ardern-Jones A, Berg J, Brady A, Bradshaw N, Brewer C, Brice G, Bullman B, Campbell J, Castle B, Cetnarsryj R, Chapman C, Chu C, Coates N, Cole T, Davidson R, Donaldson A, Dorkins H, Douglas F, Eccles D, Eeles R, Elmslie F, Evans DG, Goff S, Goodman S, Goudie D, Gray J, Greenhalgh L, Gregory H, Hodgson

SV, Homfray T, Houlston RS, Izatt L, Jackson L, Jeffers L, Johnson-Roffey V, Kavalier F, Kirk C, Lalloo F, Langman C, Locke I, Longmuir M, Mackay J, Magee A, Mansour S, Miedzybrodzka Z, Miller J, Morrison P, Murday V, Paterson J, Pichert G, Porteous M, Rahman N, Rogers M, Rowe S, Shanley S, Saggar A, Scott G, Side L, Snadden L, Steel M, Thomas M, Thomas S, McCarthy MI, Hattersley AT: **Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.** *Science* 2007, **316(5829):**1336-1341.

87. Herbert A, Lenburg ME, Ulrich D, Gerry NP, Schlauch K, Christman MF: **Open-access database of candidate associations from a genome-wide SNP scan of the Framingham Heart Study.** *Nat Genet* 2007, **39(2):**135-136.
88. The International HapMap Consortium: **A haplotype map of the human genome.** *Nature* 2005, **437(7063):**1299-1320.
89. Rabbee N, Speed TP: **A genotype calling algorithm for Affymetrix SNP arrays.** *Bioinformatics* 2006, **22:**7-12.
90. Nicolae DL, Wen X, Voight BF, Cox NJ: **Coverage and Characteristics of the Affymetrix GeneChip Human Mapping 100K SNP Set.** *PLoS Genetics* 2006, **2(5):**e67.

Communication During Brief Intervention, Intention to Change, and Outcome

Jean-Bernard Daeppen, MD
Nicolas Bertholet, MD
Gerhard Gmel, PhD
Jacques Gaume, MA

SUMMARY. *Objectives*: To explore the relationship between patient's intention to change regarding future alcohol consumption following brief alcohol intervention (BAI) and changes in alcohol consumption 12-months later and the communication characteristics between patient and counselor during BAI.

Design, Setting and Subjects: Data from 367 patients (experimental arm) of a pragmatic randomized controlled trial were used to assess the effectiveness of BAI among hazardous drinkers attending an Emergency Department (Lausanne University Hospital, Lausanne, Switzerland). Alcohol outcome measures at baseline and 12 months follow-up included usual number of drinks per week, monthly frequency of heavy episodic drinking (5 or more standard drinks for men; 4 or more for women), and the Alcohol Use Disorders Identification Test (AUDIT) score. In addition, the communication characteristics between patient and counselor were analyzed via tape recordings using the Motivational Interviewing Skill Code (MISC) from 97 participants. Patient readiness and importance to change on a 10-point Likert scale (readiness/importance to change ruler) was asked during BAI, and patient intention to change alcohol consumption (yes/no) was asked at the last step. Differences in alcohol outcome at follow-up between the 367 patients who did or did not have an intention to change consumption at baseline were compared, as were differences between these two groups in communication characteristics for the 97 who completed tape recordings.

Results: Patients with an intention to decrease alcohol consumption reduced alcohol use and related problems more often, and reported higher levels of importance and readiness to change than did their counterparts. Analyses of MISC-coded data showed a significantly higher use of MI-consistent skills among those with a moderation intention, but no group differences on the 8 other counselor communication skills measures were found. Analyses of patient speech during the intervention indicated that those with an intention to change their alcohol consumption significantly more often self-explored personal ambivalence towards alcohol, expressed more intensely their ability, commitment, desire, need and reason to change their alcohol use than did those in the no decrease group.

Conclusions: The intention expressed by hazardous drinkers when concluding BAI is associated with both patient change talk during BAI and drinking outcome 12 months later, but is mainly

Jean-Bernard Daeppen, Gerhard Gmel, and Jacques Gaume are affiliated with the Alcohol Treatment Center, Department of Community Medicine and Health, Lausanne University Hospital, Lausanne, Switzerland.

Nicolas Bertholet is affiliated with the Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA.

Address correspondence to: Dr. Jean-Bernard Daeppen, CTA, Mont-Paisible 16, 1011 Lausanne, Switzerland (E-mail: jean-bernard.daeppen@chuv.ch).

[Haworth co-indexing entry note]: "Communication During Brief Intervention, Intention to Change, and Outcome." Daeppen, Jean-Bernard et al. Co-published simultaneously in *Substance Abuse* (The Haworth Medical Press, an imprint of The Haworth Press, Inc.) Vol. 28, No. 3, 2007, pp. 43-51; and: *Alcohol/Drug Screening and Brief Intervention: Advances in Evidence-Based Practice* (ed: Richard Saitz, and Marc Galanter) The Haworth Medical Press, an imprint of The Haworth Press, Inc., 2007, pp. 43-51. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

Available online at http://suba.haworthpress.com

doi:10.1300/J465v28n03_05

independent of counselor communication skills. This intention may be an important clinical indicator of which hazardous drinkers are most likely to improve after BAI. doi:10.1300/J465v28n03_05

KEYWORDS. Brief alcohol intervention, intention, patient adherence, communication characteristics, motivational interviewing skills code, counselor skills

INTRODUCTION

Several systematic reviews and meta-analyses support the belief that brief alcohol intervention (BAI) is effective in reducing hazardous alcohol use in a variety of patients.[1-5] However, this aggregated effect might help obscure the fact that not all patients gain from BAI. Following BAI, some patients (independently of their drinking patterns) will demonstrate major changes in their alcohol use, stay the same, or even increase drinking. Little is known about the patient or counselor characteristics associated with behavior changes. Several authors have looked at moderators (characteristics) of BAI efficacy that influence the degree to which behavior change occurs, such as patient gender,[1] age,[6,7] severity of alcohol use disorder,[8] counselor empathy and directiveness[3,9] and type of BAI (e.g., duration, number of sessions[3,10]).

Patient readiness to change[11] may also moderate behavior change resulting from BAI. Gains made by some patients might be due to contemplated changes among those who have high levels of readiness to change. Here, even brief exposure may trigger relatively important and long-lasting behavior changes, where for other patient's interventions that are much more intensive may be of little help.[12] One study[13] concluded, after examining the relative value of readiness or underlying motivational dimensions of perceived desire, ability, or reasons to change alcohol or drug use, that commitment was the final common pathway in the process of change. Many patients might have high levels of readiness, but lack firm commitment to change, thus showing poorer drinking outcomes in many of the BAI studies. Since the final step of BAI often incorporates asking patients their intention, those who commit to moderation are probably more likely to actually reduce their alcohol use by follow-up.

BAI generally incorporates motivational interviewing (MI) style and techniques for use in time-limited health care settings.[14-17] In MI-style BAI, characteristics related to patient commitment, expressed by an objective of drinking moderation or abstinence, might influence drinking outcome. The Motivational Interviewing Skill Code (MISC) quantifies communication characteristics and behaviors of counselors and patients through analyses of video or tape recordings[18-20] and includes a measure of the patient's commitment to change. Since BAI works differentially for patients, it seems worthwhile to explore the relationships between the communication characteristics during BAI, the intention at the end of BAI, and the evolution of alcohol use between baseline and follow-up.

This paper has the goal of estimating whether or not an objective of alcohol use reduction set by patients at the end of BAI is associated with actual reduction 12 months later. However, this estimation would not allow many conclusions to be drawn regarding the content of BAI and how it influences patients' intention to change their alcohol consumption. For example, counselors might perform better during BAI if they believe that a patient is ready to change; conversely, a superior performance of counselors could lead to a stronger patient commitment to change by the end of BAI. Thus, a secondary goal is to evaluate the relationship between patients' intention to change and the communication characteristics during BAI using the Motivational Interviewing Skill Code (MISC) on the 97 subjects with taped BAI.

METHODS

Sample

This study was part of a randomized controlled trial in an Emergency Department (ED) aimed at evaluating the efficacy of BAI in decreasing hazardous alcohol consumption during a 12-month follow-up, as described in detail in another paper (Daeppen et al., submitted). Briefly, there were 8,833 patients attending the ED who were screened; of these 1,366 were positive for hazardous alcohol consumption, defined according to NIAAA standards as men under 65 years old who drink more than 14 drinks per week or more than 4 drinks on a single occasion in the past 30 days, and men over 65 years old and women who drink more than 7 drinks per week or more than 3 drinks on a single occasion. One drink was defined as a regular glass of wine, a regular beer, or a single shot spirit, straight or mixed in a soft drink. After signing informed consent, 486 patients were randomized to a BAI group and 880 to two control groups; of these 367 (75.5%) of the BAI patients and 688 (78.2%) of the control patients were successfully followed-up at 12 months. The analyses reported focus on the 367 patients who completed both BAI and the follow-up in order to evaluate the difference in alcohol use from baseline to 12-months later. The local ethics committee approved the entire protocol.

Counselor

Counselors were six masters level psychologists and one experienced nurse, all with a minimum of one year of clinical practice. A senior physician and a psychologist experienced in teaching motivational interviewing and BAI trained the counselors, first with a 2-day workshop on motivational interviewing containing exercises aimed at improving performance using an active, empathic listening style that minimizes confrontation[21] and second with a 5-day workshop focused on trial information procedures as well as on practice of the 6-step standardized BAI. Supervision was ongoing during the project, either in the presence of patients or through tape recordings of the BAI.

Research Procedures

A single 15-minute motivational style BAI consisting of six steps was performed. At the fifth step, counselors asked patients a question about whether or not they felt ready (based on what happened in BAI) to set an objective regarding their drinking. Patients were then dichotomized into either the group who expressed an objective to decrease or quit drinking (N = 134) or the group who did not (N = 233).

Alcohol use was measured by the weekly drinking amount calculated as the number of standard drinks usually consumed per week and by the frequency of heavy drinking measured by the average occurrence of having 5 standard drinks or more on one occasion for males under 65 years old (4 standard drinks or more on one occasion for males over 65 years old and for females) per month over the last 12-months. Patient alcohol consumption was assessed at baseline and at the 12-months follow-up. The AUDIT score was computed at baseline and at twelve months to serve as a measure of severity of alcohol use disorder.[22] The importance and readiness to change at baseline were estimated using a readiness ruler assessing importance and readiness to change on a 1-10 Likert scale, one being "not important at all–not ready at all–to change my drinking now" and 10 being "very important–ready–to change my drinking now."[23] Finally, the counselor subjectively evaluated the effectiveness of BAI immediately after BAI and assigned patients to a group that deemed BAI "rather or very effective" or to another group of "all others."

In order to explore the relationship between patients' intention to change and the communication characteristics during BAI, 166 consecutive BAI sessions were tape-recorded with patient permission and 97 of these were eligible for coding and analyses. Excluded were 33 lost to follow-up, 25 with incomplete records, 7 with mismatched identification codes, 3 who were not sufficiently fluent in French, and 1 whose wife intruded during the session. Two masters level psychologists independently completed the Motivational Interviewing Skills Code (MISC) coding, blinded to data from assessment and follow-up; both were

trained in motivational interviewing and in using the MISC, Version 2.0.[19]

The MISC is comprised of global ratings and behavior counts. Two passes were made through each taped session: the first to assess global ratings on 7-point Likert scales and the second to assign behavior counts. For the current analyses, 9 items of counselor talk and 7 items of patient talk were selected for their clinical relevance to BAI. Counselor items were (1) *Acceptance*, (2) *Empathy*, and (3) *MI Spirit* global ratings; (4) frequency of *MI-consistent skills* (*Advise with permission*, *Affirm*, *Emphasize control*, *Open question*, *Simple* and *Complex Reflections*, and *Support*); (5) frequency of *MI-inconsistent skills* (i.e., the addition of the behavior counts *Advise without permission*, *Confront*, *Direct*, *Raise concern without permission*, and *Warn*); (6) *MI-consistent skills percentage* as the proportion of the sum of MI-consistent skills and MI-inconsistent skills; (7) *Reflections/Questions ratio*; (8) *Open questions percentage* as the proportion of the sum of open and closed questions; and (9) *Complex reflections percentage* as the proportion of the sum of complex and simple reflections.[18] Patient items were (1) *Self exploration* global rating and 6 behavior counts expressed as the average strength of inclination measured on a scale from −5 (strong inclination away from change) to + 5 (strong inclination toward change) for 6 kinds of "Change talk": (2) *Ability or inability to change*, (3) *Commitment to change or not to change*, (4) *Desire to change or not to change*, (5) *Need to change versus lack of need for change, or a need not to change*, (6) *Reasons to change or reasons not to change*, and (7) *Taking steps toward or away from change*.

Data Analyses

Socio-demographic characteristics, alcohol use and related characteristics and the selected MISC variables were compared at baseline and at follow-up using Pearson's Chi Square test for categorical variables, and Wilcoxon's non-parametric test for continuous variables since asymmetric distributions and heterogeneity of variances for the two groups were frequently seen.

Intra-class correlation coefficients (ICC) of selected MISC variables between the two coders were high, suggesting fair to excellent inter-rater reliability, as reported elsewhere (Gaume et al., submitted).

RESULTS

Table 1 shows baseline demographics and alcohol consumption data on the 367 hazardous drinkers included in the BAI arm of the study. The sample consists mostly of men in their late thirties, of whom about two-thirds are employed and about one-third have college or equivalent education degrees. Their weekly alcohol use, heavy drinking frequency and AUDIT scores are comparable to profiles of average "low-grade hazardous drinkers" (average 13 drinks per week, 4 heavy drinking occasions per month and an AUDIT score of 9). They had relatively low readiness ruler's scores (importance and readiness to change drinking habits) and counselors considered that BAI would be effective for less than 50% of them.

Table 1 also compares the characteristics of the 97 patients with taped sessions to those of the 270 non-coded BAI patients and lists the relevant statistical outcomes in order to determine if findings could be appropriately generalized to the larger sample. The groups had similar demographic characteristics, alcohol use and AUDIT scores. However, the MISC-coded BAI group was significantly higher on the readiness ruler and showed a trend ($p = .06$) towards higher counselor ratings of probable BAI effectiveness than did the non-coded counterparts.

Table 2 compares patients with or without an intention to decrease their alcohol use and AUDIT scores at baseline and 12-month follow-up, as well as on importance and readiness to change and subjective counselor ratings at baseline. Patients with a decrease objective demonstrated significantly greater reductions in weekly drinking amounts, heavy drinking episodes per month and AUDIT scores from baseline to follow-up than did the no decrease counterparts. They also had significantly higher levels of importance and readiness to change, as well as a higher percentage of counselor ratings of BAI effectiveness.

Table 2 also reveals that the decrease objective group initially drank significantly more and had higher AUDIT scores than those in the

TABLE 1. Socio-Demographic Information and Alcohol Consumption–Comparison Between MISC-Coded and Non-Coded BAI in Hazardous Drinkers with 12-Month Follow-Up (N = 367).

	Total (N=367)	MISC-Coded BAI (N=97)	Non-Coded BAI (N=270)	p
Demographic data				
Age, Mean (SD)	39.3 (17.0)	38.4 (17.1)	39.7 (16.9)	0.42 (w)
Sex, % Male (N)	75.7 (278)	80.4 (78)	74.1 (200)	0.21 (c)
Education, % College degree (N)	35.1 (124)	30.5 (29)	36.8 (95)	0.27 (c)
Occupation, % Employed (N)	61.3 (225)	57.7 (56)	62.6 (169)	0.40 (c)
Weekly drinking amount				
Baseline, Mean (SD)	13.0 (11.4)	13.4 (10.2)	12.8 (11.7)	0.45 (w)
12-month follow-up, Mean (SD)	10.9 (11.8)	11.8 (13.3)	10.5 (11.3)	0.72 (w)
Baseline to follow-up difference, Mean (SD)	2.1 (10.8)	1.5 (11.5)	2.3 (10.5)	0.98 (w)
Heavy drinking episodes per month				
Baseline, Mean (SD)	4.3 (6.6)	4.3 (6.2)	4.4 (6.8)	0.65 (w)
12-month follow-up, Mean (SD)	3.6 (6.0)	4.7 (7.5)	3.3 (5.3)	0.30 (w)
Baseline to follow-up difference, Mean (SD)	0.8 (7.0)	-0.2 (5.6)	1.1 (7.4)	0.18 (w)
AUDIT Score				
Baseline; mean (SD)	9.1 (4.9)	9.5 (4.5)	9.0 (5.0)	0.10 (w)
12-month follow-up, Mean (SD)	7.3 (4.6)	7.4 (4.3)	7.2 (4.7)	0.62 (w)
Baseline to follow-up difference, Mean (SD)	1.8 (4.5)	2.1 (4.5)	1.7 (4.5)	0.20 (w)
Readiness Rulers Scores (baseline)				
Importance [1-10], Mean (SD)	3.7 (2.4)	4.4 (2.5)	3.5 (2.3)	0.01 (w)
Readiness [1-10], Mean (SD)	4.2 (2.8)	4.9 (3.0)	3.9 (2.7)	0.04 (w)
Counselor subjective evaluation of BAI effectiveness (baseline)				
% "rather effective" or "very effective" (N)	44.0 (160)	52.1 (50)	41.0 (110)	0.06 (c)

Note: (c) = Pearson's Chi Square test, (w) = Wilcoxon W test.
MISC : Motivational interviewing skill code
BAI : Brief alcohol intervention

other group. The AUDIT scores of the decrease objective patients, though much lower at follow-up, remained significantly higher than those found in the no decrease patients.

Table 3 examines the communications of counselor and patients during BAI by drinking intention among the 97 taped and MISC-coded patients. Of the nine counselor speech elements, no differences existed between decrease and no decrease patients on 8 of them; the single exception being that counselors used significantly more MI-consistent skills within the decrease group. Of the 7 patient speech elements, all but 2 (ability/inability to change and taking steps toward change) had significantly higher frequencies in the decrease objective group. Here, the patients did more self-exploration of ambivalence towards alcohol and expressed with more intensity their ability, commitment, desire, need and reason to change their drinking, compared to counterparts.

DISCUSSION

Results indicate that an objective of alcohol use reduction set by the patient near the end of BAI is associated with more patient self-exploration and change talk during BAI, higher baseline importance and readiness to change, and reduced alcohol use and problems 12 months later. This suggests that the patient intention to

TABLE 2. Alcohol Use and Other Characteristics According to Patient Intention Immediately Following BAI (N = 367).

	Intention		
	Decrease (N=134)	No decrease (N=233)	p value
Weekly drinking amount			
Baseline, Mean (SD)	15.6 (14.0)	11.5 (9.2)	0.01 (w)
12-month follow-up, Mean (SD)	12.5 (16.1)	9.9 (8.4)	0.97 (w)
Baseline to 12-month difference, Mean (SD)	3.1 (13.3)	1.5 (9.0)	0.02 (w)
Heavy drinking episodes per month			
Baseline, Mean (SD)	6.0 (7.9)	3.4 (5.6)	0.00 (w)
12-month follow-up, Mean (SD)	3.9 (6.4)	3.5 (5.7)	0.83 (w)
Baseline to 12-month difference, Mean (SD)	2.3 (8.4)	-0.1 (6.0)	0.00 (w)
AUDIT Score			
Baseline, Mean (SD)	11.2 (6.1)	7.9 (3.5)	0.00 (w)
12-month follow-up, Mean (SD)	8.3 (6.1)	6.6 (3.4)	0.03 (w)
Baseline to 12-month difference, Mean (SD)	2.8 (5.7)	1.3 (3.5)	0.00 (w)
Readiness Rulers Scores (baseline)			
Importance [1-10], Mean (SD)	5.2 (2.5)	2.9 (1.9)	0.00 (w)
Readiness [1-10], Mean (SD)	5.7 (2.4)	3.2 (2.6)	0.00 (w)
Counselor subjective evaluation of BAI effectiveness (baseline)			
% "rather effective" or "very effective" (N)	71.6 (96)	27.8 (64)	0.00 (c)

Note: (c) = Pearson's Chi Square test, (w) = Wilcoxon W test.

change might be a moderator of BAI effectiveness, perhaps reflecting patient commitment to change, which then predicts favorable alcohol use outcome. Commitment has been known to be a major influence in behavioral change processes.[13] Understanding the development of patient intention (commitment) could help maximize BAI strategies. A few studies have investigated the moderators of effectiveness in BAI trials. In chemical dependence treatment studies,[12] a modest relationship has been found between components of 12-step programs that include commitment to abstinence (which could be considered a particular form of intention to change) and outcome. Other ingredients of MI-style BAI that seem to be useful in behavioral therapy paradigms were an increase of self-efficacy,[25] engagement in the process,[26] and readiness to change.[11]

In the present study, counselor skills and communication characteristics of patients were both expected to be instrumental in changing drinking outcomes. It was surprising that only one of the 9 elements (i.e., MI-consistent skills) was significant. Perhaps counselors tend to be more empathic and motivational when they view a patient as less resistant to the whole process and more willing to change.

These findings suggest that BAI might be modified such that they could increase the proportion of patients who at the end of a BAI have a drinking objective to cut down or quit, i.e., a revision of BAI with a special focus on change talk or a modification in BAI training with the particular aim of providing counselors with MI-consistent skills. An alternative would be a case finding strategy that selects the patients who are more likely to benefit from BAI, for example, those with initial higher level of intention to change.

Interestingly, data reported also indicate that patients who set an objective of alcohol use re-

TABLE 3. Communication Characteristics (MISC) During a Brief Alcohol Intervention According to Patient Intention (N = 97).

		Intention		
		Decrease N=36	No decrease N=61	p (Wilcoxon)
Counselor	Acceptance [1-7]	5.9 (0.7)	5.9 (0.7)	0.91
	Empathy [1-7]	5.5 (0.7)	5.5 (0.7)	0.84
	MI Spirit [1-7]	5.6 (0.7)	5.4 (0.7)	0.31
	MI-consistent skills [Freq]	34.9 (8.9)	30.5 (11.1)	0.01
	MI-inconsistent skills [Freq]	1.2 (1.7)	1.2 (1.4)	0.76
	% MI-consistent skills [a]	96.7 (4.1)	96.3 (4.0)	0.48
	Ratio Reflections/Questions [b]	1.1 (0.5)	1.0 (0.5)	0.59
	% Open questions [c]	51.2 (12.5)	53.6 (13.1)	0.48
	% Complex reflections [d]	45.5 (11.2)	45.8 (13.2)	0.80
Patient	Self-exploration [1-7]	5.5 (0.8)	5.0 (0.7)	0.00
	Ability/inability to change [+5 to -5]	0.8 (1.3)	0.4 (0.9)	0.08
	Commitment to change or not to change [+5 to -5]	1.0 (1.5)	-0.7 (1.6)	0.00
	Desire to change or not to change [+5 to -5]	-1.0 (1.6)	-1.9 (1.0)	0.01
	Need to change / Lack of need to change or a need not to change [+5 to -5]	1.0 (1.2)	-0.8 (1.5)	0.00
	Reasons to change or not to change [+5 to -5]	0.3 (0.6)	-0.5 (0.7)	0.00
	Taking steps toward change / away from change [+5 to -5]	1.2 (1.1)	1.1 (0.9)	0.64

[a] % MI-consistent skills: MI-consistent skills / (MI-consistent skills + MI-inconsistent skills)
[b] Ratio Reflections/Questions: Total reflections / Total questions
[c] % Open questions: Total open questions / Total questions (open + closed)
[d] % Complex reflections: Total complex reflections / Total reflections (complex + simple)
MISC : Motivational interviewing Skill Code

duction consumed more alcohol, had more heavy drinking episodes and higher AUDIT scores at baseline. This suggests that objectives to decrease alcohol use reduction are more likely to occur in individuals having higher levels of alcohol use and problems, while patients with less severe hazardous drinking feel less pressure to set drinking reduction objectives and to change their consumption habits. BAI may be most useful after patients reach a certain threshold of hazardous drinking and are more amenable to seeking and receiving help.

Despite some methodological strengths in this study, such as efforts to maximize the internal validity of BAI used (specification of the theoretical model, i.e., motivational interviewing; uniformity of counselor training and attention to fidelity of interventions during the trial), and high agreement between MISC raters, some limitations to the generalizability of these findings should be acknowledged. Most of the socio-demographic and alcohol consumption data of taped and MISC-coded versus non-coded BAI were similar, yet there were differences in readiness ruler scores and counselor subjective evaluations of the efficacy of BAI. This suggests that somehow the recording process itself modified the course of BAI, perhaps because patients ready to change their drinking would be more likely to accept to be recorded. Sample selectivity arising from counselor choice of patient to record can probably be ruled out, since consecutive patients were asked to do this; in any case, these group differences preclude generalizing findings to the non-coded group. There were a relatively small number (97) of MISC-coded BAI patients; this may present a statistical power problem. The confounding of higher alcohol consumption and greater readiness to change among patients who set decrease objectives makes it difficult to determine which of the two provide the greatest impetus to reduce drinking at follow-up.

Several other potential limitations should be mentioned. The MISC was designed for analyses of MI, not BAI, although common skills and

common target (behavior change) of MI and BAI suggest that the MISC might well be used to study BAI content. Using the MISC in French also might be a concern as it is a linguistic coding instrument, developed in English and so far not validated in French. Another limitation might arise due to the nature of ED, where time and medical constraints are quite different than those found in other settings (e.g., primary care), and might lead to specific behaviors and attitudes that unduly influence BAI. Self-selection bias may have been introduced by assessing only patients who were willing to participate and have their data recorded, and to cooperate with a follow-up one year later. Since these findings are in a relatively small sample observed prospectively, it would be helpful to repeat and expand such study to allow use of more complex statistical methods to study mechanisms of behavior change, which is currently a NIAAA priority. Future BAI research should try to probe the interaction effects of higher alcohol consumption/higher readiness to change scores among patients with an intention to reduce alcohol use and identify predictors of BAI effectiveness that could be assessed in advance of any intervention attempts. Even with these caveats, the present preliminary results point to the feasibility of optimizing effectiveness and reducing the costs of BAI in ED settings by focusing on patients who are willing to change.

REFERENCES

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

2. Dunn C, Deroo L, Rivara FP: The use of brief interventions adapted from motivational interviewing across behavioral domains: a systematic review. Addiction 2001; 96: 1725-1742.

3 Bien TH, Miller WR, Tonigan JS: Brief interventions for alcohol problems: a review. Addiction 1993; 88: 315-335.

4 D'Onofrio G, Degutis LC: Preventive care in the emergency department: screening and brief intervention for alcohol problems in the emergency department: a systematic review. Acad Emerg Med 2002; 9: 627-638.

5. Emmen MJ, Schippers GM, Bleijenberg G, Wollersheim H: Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: systematic review. BMJ 2004; 328: 318.

6. Grossberg PM, Brown DD, Fleming MF: Brief physician advice for high-risk drinking among young adults. Ann Fam Med 2004; 2: 474-480.

7. Fleming MF, Manwell LB, Barry KL, Adams W, Stauffacher EA: Brief physician advice for alcohol problems in older adults: a randomized community-based trial. J Fam Pract 1999; 48: 378-384.

8 Fleming MF, Barry K, Manwell L, Johnson K, London R: Brief physician advice for problem alcohol drinkers. JAMA 1997; 13: 1039-1045.

9. Karno MP, Longabaugh R: An examination of how therapist directiveness interacts with patient anger and reactance to predict alcohol use. J Stud Alcohol 2005; 66: 825-832.

10. Woolard RH, Carty K, Wirtz P, Longabaugh R, Nirenberg TD, Minugh PA: Research fundamentals: follow-up of subjects in clinical trials: addressing subject attrition. Acad Emerg Med 2004; 11: 859-866.

11. DiClemente CC, Schlundt D, Gemmell L: Readiness and stages of change in addiction treatment. Am J Addict 2004; 13: 103-119.

12. Barnett NP, Lebeau-Craven R, O'Leary TA, Colby SM, Rohsenow DJ, Monti PM: Predictors of motivation to change after medical treatment for drinking-related events in adolescents. Psychol Addict Behav 2002; 16: 106-112.

13. Amrhein PC, Miller WR, Yahne CE, Palmer M, Fulcher L: Client commitment language during motivational interviewing predicts drug use outcomes. J Consult Clin Psychol 2003; 71: 862-878.

14. Noonan WC, Moyers TB: Motivational interviewing: a review. Journal of Substance Misuse 1997; 2: 8-16.

15. Dunn C, Deroo L, Rivara FP: The use of brief interventions adapted from motivational interviewing across behavioral domains: a systematic review. Addiction 2001; 96: 1725-1742.

16. Burke BL, Arkowitz H, Menchola M: The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. Journal of Consulting & Clinical Psychology 2003; 71: 843-861.

17. Vasilaki EI, Hosier SG, Cox WM: The efficacy of motivational interviewing as a brief intervention for excessive drinking: a meta-analytic review. Alcohol Alcohol 2006; 41: 328-335.

18. Miller WR. Motivational Interviewing Skill Code (MISC) Coder's Manual. http://casaa unm edu/download/misc1 pdf [2000

19. Miller WR, Moyers TB, Ernst D, Amrhein PC: Manual for the Motivational Interviewing Skill Code (MISC) Version 2.0. http://casaa unm edu/download/misc pdf [2003

20. Moyers TB, Miller WR, Hendrickson SM: How does motivational interviewing work? Therapist interpersonal skill predicts client involvement within motivational interviewing sessions. J Consult Clin Psychol 2005; 73: 590-598.

21. Baer JS, Rosengren DB, Dunn CW, Wells EA, Ogle RL, Hartzler B: An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. Drug Alcohol Depend 2004; 73: 99-106.

22. Allen JP, Litten RZ, Fertig JB, Babor T: A review of research on the Alcohol Use Disorders Identification Test (AUDIT). Alcohol Clin Exp Res 1997; 21: 613-619.

23. Hesse M: The Readiness Ruler as a measure of readiness to change poly-drug use in drug abusers. Harm Reduct J 2006; 3:3.

24. Morgenstern J, Frey RM, McCrady BS, Labouvie E, Neighbors CJ: Examining mediators of change in traditional chemical dependency treatment. J Stud Alcohol 1996; 57(1):53-64.

25. Morgenstern J, Longabaugh R: Cognitive-behavioral treatment for alcohol dependence: a review of evidence for its hypothesized mechanisms of action. Addiction 2000; 95: 1475-1490.

26. Dearing RL, Barrick C, Dermen KH, Walitzer KS: Indicators of client engagement: influences on alcohol treatment satisfaction and outcomes. Psychol Addict Behav 2005; 19: 71-78.

doi:10.1300/J465v28n03_05

The New Medicare Part D Prescription Drug Benefit: An Estimation of Its Effect on Prescription Drug Costs in a Medicare Population with Atrial Fibrillation

Carmella Evans-Molina, MD, Susan Regan, PhD,† Lori E. Henault, MPH,‡ Elaine M. Hylek, MD, MPH,‡ and Gregory R. Schwartz, MD, JD†*

(See editorial comments by Dr. Richard Stefanacci on pp 1134–1136)

OBJECTIVES: To compare prescription drug cost savings under the most commonly selected Medicare Part D prescription plan in 2006 with savings under the Medicare standard benefit and with drug costs assuming no coverage in an elderly cohort of patients.

DESIGN: Inception cohort study.

SETTING: An academic medical center.

PARTICIPANTS: Four hundred seventy-two patients aged 65 and older who were followed as part of a larger study assessing stroke prevention in patients with atrial fibrillation.

MEASUREMENTS: Prescription drug expenditures were calculated for each patient in the cohort under three conditions: the 2006 AARP-endorsed prescription drug plan, the Medicare standard benefit, and no prescription drug coverage.

RESULTS: Total prescriptions drug costs were lower under the AARP plan, yet patients paid a similar percentage of total costs under the AARP plan and the Medicare standard benefit. Using different cost assessments, 27% to 46% of patients entered the "doughnut hole" in the AARP plan, and 3% to 11% emerged to receive catastrophic coverage.

CONCLUSION: Both the AARP-sponsored and standard Medicare Part D prescription drug benefit programs offer significant savings to enrollees. A greater savings is achieved under the private AARP drug insurance plan, largely due to greater discounts reflected in the negotiated drug prices. A substantial portion of enrollees enter but do not emerge from the coverage gap. **J Am Geriatr Soc 55:1038–1043, 2007.**

Key words: Medicare Part D; insurance; pharmaceutical services; prescription; drug; atrial fibrillation

In 2006, Medicare patients were given the opportunity to purchase insurance coverage for prescription drugs under Medicare Part D (enacted as the Medicare Prescription Drug Improvement and Modernization Act of 2003 (MMA)).[1,2] Given that many seniors, more than 25% according to some studies, previously lacked drug coverage,[3] the new benefit promised to provide urgently needed relief for many Medicare beneficiaries.

When enacted in 2003, the legislation envisioned a standard set of benefits provided by private insurance companies. Under the standard benefit, patients would pay the first $250 of drug expenses, with this expenditure serving as a deductible. Of the next $2,000 of patient drug costs (up to a patient outlay of $2,250), Medicare would cover 75% (or $1,500), leaving the patient to pay 25% ($500) as coinsurance. At this point, a beneficiary would encounter the "doughnut hole" gap in coverage; there would be no coverage between total outlays of $2,250 and $5,100. After incurring more than $5,100 in total drug costs, a patient would have 95% of their costs covered by Medicare and would pay 5% as coinsurance (Table 1). To enroll in the plan, patients would pay a monthly premium, which at the time of the bill's passage in 2003 was estimated to be approximately $35 per month. In 2005, the Centers for Medicare and Medicaid Services (CMS) estimated that the premium would be $32 per month, or $384 annually. Therefore, under the standard Part D benefit, a patient would pay $3,600 ($250 deductible plus $500 coinsurance plus $2,850 in the doughnut hole) of drug costs up to $5,100, for a total of $3,984 including the annual premium of $384.[2]

From the *Department of Medicine, Division of Endocrinology and Metabolism, University of Virginia, Charlottesville, Virginia; †Department of Medicine, General Medicine Division, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; and ‡Department of Medicine, Section of General Internal Medicine and Clinical Epidemiology Unit, Boston University School of Medicine, Boston, Massachusetts.

Address correspondence to Gregory R. Schwartz, Department of Medicine, General Medicine Division, Massachusetts General Hospital, Harvard Medical School, 50 Staniford Street, Suite 300, Boston, MA 02114. E-mail: gschwartz@partners.org

DOI: 10.1111/j.1532-5415.2007.01285.x

0002-8614/07/$15.00

Table 1. Structure of the Medicare Standard Benefit and the AARP UnitedHealth Group Medicare Prescription Drug Plan*

	Medicare Standard Benefit	AARP
Deductible	$250	None
Initial coverage	Patients pay 25% of drug costs until total patient outlay of $2,250	Patients pay according to tier categorization until total drug costs reach $2,250. Tier 1 drugs have a $5 copayment, Tier 2 drugs have a $28 copayment, Tier 3 drugs have a $56 copayment, and Specialty Tier drugs cost 25% of the NDP
Annual coverage gap	No coverage until total costs reach $5,100	After $2,250 in total drug costs and until $3,600 in true out-of-pocket costs, patients pay 100% of the NDP for each medication
Catastrophic protection	Over $5,100 in total drug cost, patients pay 5% of drug costs	Over $3,600 in true out-of-pocket costs, patients pay a $2 copayment or 5% of the NDP, whichever is greater, for a Tier 1 medication; a $5 copayment or 5% of the NDP, whichever is greater for a Tier 2 medication; a $5 copayment or 5% of the NDP, whichever is greater, for a Tier 3 medication; and 5% of the NDP for a Specialty Tier medication

Note: Adapted from the AARP website: http://www.aarpmedicarerx.com.[14]
NDP = network discounted price.

As an incentive to develop enhanced benefits for participants, a variety of differently structured plans have been offered, each with a unique premium structure, deductible policy, formulary, and method of entering and exiting the coverage gap, or doughnut hole.[4–6] There are four basic plan structures: a standard benefit, based on the original legislation; actuarily equivalent plans, which adhere to the standard benefit with respect to deductibles and the doughnut hole but offer enhanced cost sharing; basic alternative plans, in which deductibles and cost sharing are altered; and enhanced alternative plans, which exceed the defined standard benefit.[7]

The economic benefits of offering these private plans have been estimated based on hypothetical patients.[1,8] In this study, we report the estimated benefit one private plan would have in a prospectively identified cohort of actual patients with the predefined cardiovascular diagnosis of atrial fibrillation (AF). The prevalence of this condition increases with age, and it is estimated that 7.5 million Americans will have AF by 2020.[9–12] For the purposes of the analysis, the 2006 version of the basic alternative plan endorsed by the AARP and administered by UnitedHealth Group was used. As of September 2006, 3.4 million people were enrolled in this plan, which was available in all 50 states and had the leading market share, with 21% of total enrollees.[8]

METHODS

Study Participants

Consecutive patients were prospectively identified as part of a larger study assessing stroke prevention in individuals with AF during January 2001 to June 2003 from daily searches of electronic hospital discharge summaries and patient referrals to an anticoagulation clinic. To be eligible, patients had to be aged 65 and older and have AF verified according to electrocardiogram. Demographic data and diagnoses were extracted from the medical record.

Medication Ascertainment

At the time of enrollment, a complete list of current medications was recorded and verified in the electronic medical record. Dosages were not recorded. During the analysis, to provide the most conservative estimate of annual prescription costs, short-term medications such as antibiotics and medications taken on an as-needed basis were not included. Injectable medications, other than insulin, and medications that have been withdrawn from the market were also excluded. Because dosages were not recorded, two doses were analyzed for each medication: the usual starting (minimum) dose and the maximum acceptable dose.

Prescription Plan Cost Comparisons

Total annual drug costs were calculated for each patient under three conditions: the AARP plan, the Medicare Part D standard benefit (Table 1), and no prescription coverage. For each condition, calculations were performed twice, using minimum and maximum cost assumptions. Under the minimum assumption, the minimum dose was assumed for all drugs. Under the maximum assumption, the maximum dose was assumed. For the majority of medications, the minimum dose was associated with the lowest cost. For a small number of drugs, the minimum dose was priced higher. Under this circumstance, the minimum dose was designated as the maximum cost, with the maximum dose being designated as the minimum cost. The portion paid by the patient, in the form of premiums, copayments, and nonreimbursed out-of-pocket payments, were also determined using the relevant drug prices and plan rules. The amount paid by the plan was calculated by subtracting patient payments from total aggregate drug costs.

For the standard-benefit and no-coverage conditions, retail pricing information was obtained from drugstore.com, an on-line mail-order pharmacy, during March through July 2006. To avoid overestimation of costs, prices from drugstore.com were compared with those of a national retail chain and the adjusted average wholesale price (AWP) for the 20 most commonly prescribed medications. For 80% of medications, the drugstore.com price was the lowest. For the remaining medications, the drugstore.com price varied less than 6% from the lowest price. Using the drugstore.com prices, annual retail medication costs were calculated for each patient in the cohort. If a particular medication was not available through drugstore.com, a price was obtained using the adjusted AWP. Based on a previously published analysis, 4% was added to the AWP

for brand medications, and 36% was subtracted from the AWP for generic medications.[13] If available, the generic price was always used preferentially.

The 2006 AARP plan had three phases of coverage: the initial coverage phase, annual coverage gap (doughnut hole), and catastrophic protection phase. The details of this plan are outlined in Table 1. During the initial phase, patients pay a copayment for medications according to a tier categorization. When drug costs reach $2,250, patients enter the annual coverage gap and remain there until they pay $3,600 in true out-of-pocket costs (defined as the amount the patient pays in co-insurance during the initial coverage period and the amount paid in the coverage gap). While in the coverage gap, patients pay 100% of the network discounted price (NDP) for each prescription until they reach $3,600 in out-of-pocket costs. At that point, patients move into the catastrophic protection phase, in which they pay a copayment or percentage of the NDP (Table 1).[14] Study investigators obtained the NDP for the minimum and maximum doses of each medication (N = 456) through phone calls to the AARP telephone customer service center.

To examine the effects of the doughnut hole structure of the AARP plan, each patient's medication costs were calculated on a monthly basis, and the amount paid by the patient and plan each month was determined. If a patient's drug costs reached $2,250 during a month (the patient entered the doughnut hole), or the out-of-pocket expenses exceeded $3,600 (the patient exited the hole), the drug costs were added in a sequence that maximized the benefit.

The institutional review board at Massachusetts General Hospital approved the study. The nature of the study did not require written informed consent.

RESULTS

Patient Clinical Characteristics

During the study period, 472 patients were enrolled, of whom 47% were female. Forty-two percent of patients were aged 65 to 74, 25% were aged 75 to 79, and 33% were aged 80 and older The most common medical diagnoses (in addition to AF) were hypertension (75%), coronary artery disease (35%), heart failure (27%), and diabetes mellitus (22%). Patients were taking a median of five prescription medications daily (range 1 to 15, Figure 1). There was no difference in the number of medications across age groups. The most commonly prescribed medication classes, besides warfarin, which every participant was prescribed, were beta-blockers, diuretics, anti-arrhythmics, and anti-hyperlipidemics.

Patient Drug Costs

Overall, generic prices were assigned for 76% of medications. Diuretics, beta-blockers, anti-arrhythmics, and angiotensin-converting enzyme inhibitors had the lowest monthly costs, and proton-pump inhibitors, anti-hyperlipidemics, anti-platelet, and osteoporosis medications had the highest monthly costs.

Under the AARP plan, 69% of medications for the cohort were categorized as Tier 1 medications. Only 3% of medications were in the most expensive Specialty Tier or not covered at all. The AARP plan did not cover 62 medications, with the majority being benzodiazepines.

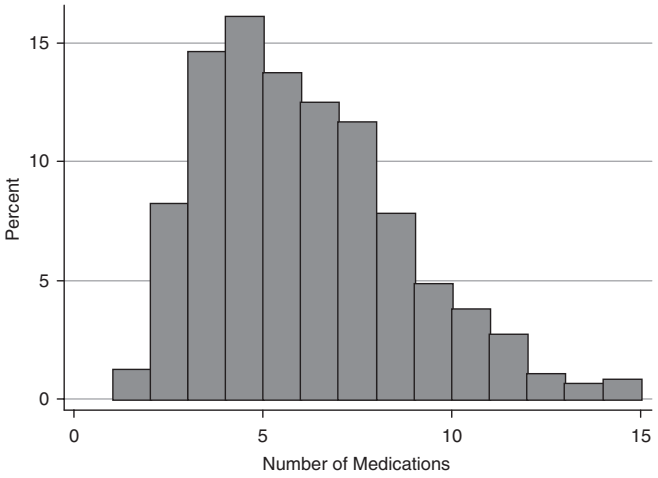


Figure 1. Median number of prescriptions per patient. As shown on the *x*-axis, the range of monthly prescriptions used by patients in the cohort was 1 to 15. The *y*-axis shows the percentage of patients in the cohort and their associated monthly number of prescriptions.

Table 2 shows the cost comparison between the AARP plan, the standard benefit, and no prescription drug coverage. Under the minimum projection, total annual drug costs under the AARP plan for the cohort were $813,124, compared with $937,225 under the standard benefit and $937,225 in the no coverage condition. Using the maximum projection, total costs under the AARP plan were $1,237,741, compared with $1,487,858 under the standard benefit and $1,487,858 in the no coverage condition. Total annual drug costs were the same for the standard benefit and no-coverage conditions, because prices were the same in both conditions. There was a 13% savings under the minimum cost projection and a 17% savings under the maximum projection. Under the AARP plan, the patients paid 71% of drug costs under the minimum projection and 62% under the maximum projection. The median annual cost per patient was $908 under the minimum projection and $1,092 under the maximum projection. Under the standard benefit, the patients paid 69% of total drug costs under the minimum projection and 64% under the maximum projection. The annual median cost per patient under the standard benefit was $956 under the minimum projection and $1,456 under the maximum projection. Under conditions of no prescription drug coverage, patients paid 100% of costs, and the median cost per patient was $1,539 under the minimum projection and $2,572 under the maximum projection.

A compelling characteristic of the MMA is the coverage gap, or doughnut hole, which represents the period after a patient accrues $2,250 in total drug costs. Under the AARP plan, using the minimum projection, 73% of patients in the cohort remained in the initial coverage phase for the entire year, whereas 27% of patients in the cohort entered the coverage gap. Of those who entered the coverage gap, the majority (89%) ended the year in the gap, whereas 11% entered and exited the doughnut hole. Under the maximum projection, 46% of patients in the cohort entered the hole.

Table 2. Annual Cost Comparison of the AARP UnitedHealth Group Medicare Prescription Drug Plan, the Medicare Standard Benefit, and No Prescription Drug Coverage*

	Cost Amount					
	AARP		Medicare Standard Benefit		No Prescription Drug Coverage	
Analysis of Costs	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Total cost of drugs, per patient/total, $	1,723/813,124	2,622/1,237,741	1,986/937,225	3,152/1,487,858	1,986/937,225	3,152/1,487,858
Paid by patient						
Drug costs, per patient/total, $	947/447,093	1,329/627,370	989/466,666	1,628/768,535	1,986/937,225	3,152/1,487,858
Premium, per patient/total, $	284/134,180	284/134,180	384/181,248	384/181,248	0	0
Total, per patient/total, $	1,232/581,273	1,613/761,550	1,373/647,914	2,012/949,783	1,986/937,225	3,152/1,487,858
Median per patient (range), $	908 (344–4,228)	1,092 (344–4,488)	956 (552–4,222)	1,456 (552–4,560)	1,539 (168–9,854)	2,572 (168–16,623)
Percentage of total cost	71	62	69	64	100	100
Paid by plan						
Drug costs, per patient/total, $	775/366,031	1,293/610,372	997/470,563	1,524/719,326	—	—
Premium, per patient/total, $	− 284/134,180	− 284/134,180	− 384/181,248	384/181,248	—	—
Total, per patient/total, $	491/231,851	1009/476,192	613/289,315	1140/538,078	—	—
Percentage of total cost	29	38	31	36	—	—

* The cost of medications used by the cohort was calculated under each plan assuming that patients continued their baseline medications for 12 months. Because actual doses taken were not available, minimum (and maximum) costs were calculated assuming that the patient took the least (or most) expensive dose for every drug. The network discounted prices were obtained from the AARP customer service center and used for calculations under the AARP plan; under the standard benefit and no coverage categories, prices were obtained from drugstore.com

Of the 46% of patients who entered the gap, 76% remained in the gap, whereas 24% entered and exited. Thus, under the minimum projection, 3% of the total cohort ended the year in the catastrophic coverage phase, and under the maximum projection, 11% of the total cohort ended the year in the catastrophic coverage phase. The median monthly expenditure for patients who never entered the doughnut hole (using the minimum cost projection) was $65. For those who entered, the median monthly payment was $134 (range $61–280) before entering the hole and $294 (range $105–780) while in the hole. The median increase in monthly payments upon entering the hole was $166 (range $0–578) or 125% (range 5–483%). Three patients entered the hole during the 12th calendar month but did not experience an increase in cost because several of their medications were priced below the Tier 1 copayment. Payments dropped considerably for those who exited the hole. The median monthly payment was reduced to $56 (range $48–248). On average, those who ended the year in the hole arrived there in approximately calendar month 9, whereas those who entered and exited arrived in approximately calendar month 5 and exited in approximately calendar month 11. Under the minimum cost assessment, 490 person-months were spent in the coverage gap, and under the maximum cost assessment, 996 person-months were spent in the coverage gap, or doughnut hole.

DISCUSSION

An assessment of the effect of the Medicare prescription drug plan on an elderly cohort of patients recruited as part of a larger study to assess stroke prevention in individuals with AF is presented here. Drug costs incurred under the 2006 AARP-endorsed plan were compared with costs incurred under the Medicare standard benefit and with cost projections assuming no prescription drug coverage. This comparison revealed that patients in this cohort benefit significantly from the AARP plan and the standard benefit, with a slightly greater benefit from the AARP plan. Under the AARP plan, median annual patient expenditures decreased 41% under the minimum projection and 57% under the maximum projection. This is compared with 38% under the minimum projection and 43% under the maximum projection with the standard benefit. These numbers are less than those estimated in a recent CMS analysis of 16 hypothetical patients with a variety of comorbid conditions, in which an average 60% savings was predicted, with a maximum savings of 72%.[15]

Regulations have allowed private insurance companies to offer plans with actuarially equivalent or better benefits. This analysis of the AARP plan, which had the leading market share in 2006, with more than 3.4 million enrollees, or 21% of total Part D participants, suggests that patients have lower median annual drug expenditures under this plan than under the standard benefit. Total drug costs were 13% less in the AARP plan under the minimum projection than with the standard benefit and 17% less under the maximum projection.

There are limitations in the analysis that may underestimate the savings seen with the AARP plan. First, the analysis used pricing information from drugstore.com to calculate drug costs under the standard benefit and when patients had no coverage. Drugstore.com often offers lower prices than neighborhood retail pharmacies, so even greater savings might be realized for patients using a local pharmacy. CMS also uses pricing information from drugstore.com, as well as from Costco.com, another on-line mail-order pharmacy, for their analyses.[15] Second, annual costs were analyzed based on the medication list of each participant upon enrollment. Medications taken on an

as-needed basis or for an acute illness were not included. This assumption could lead to lower estimates for drug costs in this cohort. By using only the medication list at the time of enrollment, it was also not possible to capture savings that would occur as a result of patient and physician behavioral changes. In reality, patients could switch to alternative medications within a therapeutic class to achieve greater savings in copayments with the AARP plan. Seventy-six percent of the medications used by this cohort were available as generics, and the majority of medications were categorized in Tier 1.

The finding that patients save more with the AARP plan than with the standard benefit shows that the government subsidy is not the only driver of savings. A substantial portion of patients' savings appears to come from the lower network prices that plans are able to obtain for their members. In the aggregate, under the minimum projection, total drug costs dropped from $937,225 with the standard benefit (based on drugstore.com prices) to $813,124 with the AARP plan (based on the NDP) (Table 2). Thus, nearly $125,000 of savings was due solely to the price discount obtained by the private plan through negotiated savings from drug manufacturers. Because actual prices paid by private insurers to pharmaceutical manufactures are not available, it is impossible to calculate the true percentage of costs assumed by the plan.[5] Despite this, both plans, under the minimum projection, fall short of previous calculations that predict a subsidy of $1,092 per person and short of the CMS projections using hypothetical patients.[15,16] The median subsidy in the minimum cost projection is $631 and $583 for the AARP and standard benefit, respectively. Under the maximum projection, though, this median subsidy is $1,486 (AARP) and $1,116 (standard benefit), highlighting the basic plan structure that those who spend more save more.

An essential component of Medicare Part D is the coverage gap, or doughnut hole. Although this hole makes the overall benefit less costly to the government, it raises important clinical and safety concerns. The financial burden of medications can lead to nonadherence, as evidenced in a recent analysis of Medicare Advantage patients, whose pharmacy benefits were capped at $1,000. The presence of this cap led to worse adherence and clinical outcomes, including poorer control of blood pressure, lipid, and blood glucose levels. Patients exceeding the cap had higher rates of nonelective hospitalizations, visits to the emergency department, and death. The cost of these hospitalizations and emergency department visits offset the prescription drug savings.[17]

In the current study's cohort, 27% of patients in the minimum projection and 46% of patients in the maximum projection fell into the doughnut hole. Only 3% of patients in the minimum projection and 11% of patients in the maximum projection emerged from the hole with catastrophic coverage by the end of the year; the rest remained in the hole, paying full price for their medications. The median increase in monthly payments upon entering the hole was 125%.

Savings under Part D come with a cost. The AARP plan, like most private plans, limits patients' choices of medication through formularies and tiered copayments. More than 30% of the medications for the cohort were not categorized as Tier 1, and 62 medications, the majority of them benzodiazepines, were not covered at all. Tiered copayments are designed to shift patients' drug choices to lower-cost alternatives, but each time a change is made, a potential for a disruption in care is introduced.

Despite all the criticisms of the MMA, these data suggest that it has the potential to reduce patient drug costs significantly. This analysis of the effect of Medicare Part D on real patients with common comorbidities underscores the significant savings achieved by allowing private insurers to negotiate steep price discounts with drug manufacturers. The AARP plan is just one example; as the most popular plan in 2006, its savings suggest that other private plans with negotiated prices are also likely to provide savings to patients. This analysis shows that the percentage of patients affected by the coverage gap ranges from 25% to 50%, underscoring the importance of creating alternative solutions to this lapse in coverage.

Continued research is needed to determine the effect of Medicare Part D on health outcomes, physician prescribing practices, and drug usage patterns in elderly populations.

ACKNOWLEDGMENTS

The authors would like to thank Mr. T. Molina for his assistance in data collection.

Financial Disclosure: This study was funded by the Robert Wood Johnson Foundation Generalist Physician Faculty Scholars Program Grant039174. Dr. Schwartz received funding support from the John D. Stoeckle Center for Primary Care Innovation at Massachusetts General Hospital. Dr. Evans-Molina is supported in part by a National Research Service Award (F32 DK076450-01) from the National Institutes of Health. Elaine M. Hylek has served on an Advisory Board for Bristol Myers Squibb and has received research support from AstaZeneca and Bristol-Myers Squibb.

Author Contributions: Carmella Evans-Molina, Susan Regan, and Gregory R. Schwartz: study concept and design, data analysis, interpretation of data, and preparation of manuscript. Lori E. Henault: acquisition of subjects, data analysis, interpretation of data, and preparation of manuscript. Elaine M. Hylek: study concept and design, acquisition of subjects, data analysis, interpretation of data, and preparation of manuscript.

Sponsor's Role: The sponsor had no role in the design, methods, subject recruitment, data collections, analysis, or preparation of the manuscript.

REFERENCES

1. Altman DE. The new medicare prescription-drug legislation. N Engl J Med 2004;350:9–10.
2. Bach PB, McClellan MB. A prescription for a modern medicare program. N Engl J Med 2005;353:2733–2735.
3. Safran DG, Neuman P, Schoen C et al. Prescription drug coverage and seniors: Findings from a 2003 national survey. Health Aff (Millwood) 2005;(Suppl Web Exclusives):W5-152–W5-166.
4. Medicare Prescription Drug Coverage. U.S. Department of Health and Human Services [on-line]. Available at www.medicare.gov Accessed November 2006.
5. Avorn J. Part "D" for "defective"—The medicare drug-benefit chaos. N Engl J Med 2006;354:1339–1341.
6. Two health firms lead drug-plan signup. Seattle Times 2006; A6.
7. Bach PB, McClellan MB. The first months of the prescription-drug benefit—a CMS update. N Engl J Med 2006;354:2312–2314.

8. Centers for Medicaid and Medicare Services Enrollment Data for Medicare Part D [on-line]. Available at www.cms.hhs.gov/PrescriptionDrugCovGenIn/02_EnrollmentData.asp#TopOfPage Accessed November 2006.
9. Miyasaka Y, Barnes ME, Gersh BJ et al. Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. Circulation 2006;114:119–125.
10. Furberg C, Psaty BM, Manolio TM et al. Prevalence of atrial fibrillation in elderly subjects (the Cardiovascular Health Study). Am J Cardiol 1994;74:236–241.
11. Feinberg WM, Blackshear JL, Laupacis A et al. Prevalence, age distribution, and gender of patients with atrial fibrillation. Analysis and implications. Arch Intern Med 1995;155:469–473.
12. Go AS, Hylek EM, Phillips KA et al. Prevalence of diagnosed atrial fibrillation in adults: National implications for rhythm management and stroke prevention. The Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. JAMA 2001;285:2370–2375.
13. Cubanski J, Frank RG, Epstein AM. Savings from drug discount cards: Relief for medicare beneficiaries? Health Aff (Millwood) 2004;(Suppl Web Exclusives):W4-198–W4-209.
14. AARP Medicare Prescription Drug Plan [on-line]. Available at www.aarpmedicarerx.com/ Accessed November 2006.
15. Center for Medicaid and Medicare Services Press Release [on-line]. Available at www.cms.hhs.gov/apps/media/press/release.asp?Counter=1946 Accessed November 2006.
16. Winter J, Balza R, Caro F et al. Medicare prescription drug coverage: Consumer information and preferences. Proc Natl Assoc Soc 2006;103:7929–7934.
17. Hsu J, Price M, Huang J et al. Unintended consequences of caps on medicare drug benefits. N Engl J Med 2006;354:2349–2359.

The American Journal of Medicine (2007) 120, 700-705

ELSEVIER

THE AMERICAN JOURNAL of MEDICINE®

CLINICAL RESEARCH STUDY

Death and Disability from Warfarin-Associated Intracranial and Extracranial Hemorrhages

Margaret C. Fang, MD, MPH,[a] Alan S. Go, MD,[a,b,c] Yuchiao Chang, PhD,[d] Elaine M. Hylek, MD, MPH,[e] Lori E. Henault, MPH,[e] Nancy G. Jensvold, MPH,[c] Daniel E. Singer, MD[d]

[a]The Department of Medicine, University of California, San Francisco; [b]Department of Epidemiology and Biostatistics, University of California, San Francisco; [c]Division of Research, Kaiser Permanente of Northern California, Oakland, Calif; [d]Clinical Epidemiology Unit, Massachusetts General Hospital, Boston, Mass; [e]General Internal Medicine and Clinical Epidemiology Unit, Boston University School of Medicine, Boston, Mass.

ABSTRACT

OBJECTIVES: Little is known about the outcomes of patients who have hemorrhagic complications while receiving warfarin therapy. We examined the rates of death and disability resulting from warfarin-associated intracranial and extracranial hemorrhages in a large cohort of patients with atrial fibrillation.

METHODS: We assembled a cohort of 13,559 adults with nonvalvular atrial fibrillation and identified patients hospitalized for warfarin-associated intracranial and major extracranial hemorrhage. Data on functional disability at discharge and 30-day mortality were obtained from a review of medical charts and state death certificates. The relative odds of 30-day mortality by hemorrhage type were calculated using multivariable logistic regression.

RESULTS: We identified 72 intracranial and 98 major extracranial hemorrhages occurring in more than 15,300 person-years of warfarin exposure. At hospital discharge, 76% of patients with intracranial hemorrhage had severe disability or died, compared with only 3% of those with major extracranial hemorrhage. Of the 40 deaths from warfarin-associated hemorrhage that occurred within 30 days, 35 (88%) were from intracranial hemorrhage. Compared with extracranial hemorrhages, intracranial events were strongly associated with 30-day mortality (odds ratio 20.8 [95% confidence interval, 6.0-72]) even after adjusting for age, sex, anticoagulation intensity on admission, and other coexisting illnesses.

CONCLUSIONS: Among anticoagulated patients with atrial fibrillation, intracranial hemorrhages caused approximately 90% of the deaths from warfarin-associated hemorrhage and the majority of disability among survivors. When considering anticoagulation, patients and clinicians need to weigh the risk of intracranial hemorrhage far more than the risk of all major hemorrhages.

KEYWORDS: Atrial fibrillation; Death; Disability; Hemorrhage; Intracranial hemorrhage; warfarin

Atrial fibrillation is the most common clinically significant cardiac arrhythmia and increases the risk for ischemic stroke 4- to 5-fold.[1] Anticoagulation therapy with warfarin can reduce the risk for ischemic stroke by 68% but also increases the risk for major hemorrhagic complications.[2] Clinical decision-making regarding the appropriateness of warfarin therapy for patients with atrial fibrillation has generally relied on balancing the risk of ischemic strokes without warfarin therapy with the risk of all major hemorrhage with warfarin therapy.

Rates of ischemic stroke in patients with atrial fibrillation who are not taking warfarin can be as high as 12% per year,[2,3] and the proportion of patients who have major functional disability after an atrial fibrillation-related ischemic stroke is substantial, as high as 59%.[4] Yet multiple

This work was supported by Public Health Services research grant AG15478 from the National Institute on Aging, the Eliot B. and Edith C. Shoolman Fund of Massachusetts General Hospital, and a Hartford Geriatrics Health Outcomes Research Scholars Award from the AGS Foundation for Health in Aging.

Requests for reprints should be addressed to Margaret C. Fang, MD, MPH, Department of Medicine, Hospitalist Group, University of California, San Francisco, 533 Parnassus Ave, Box 0131, San Francisco, CA 94143.

E-mail address: mfang@medicine.ucsf.edu

0002-9343/$ -see front matter
doi:10.1016/j.amjmed.2006.07.034

studies demonstrate that patients with atrial fibrillation are frequently not prescribed warfarin despite its benefits in ischemic stroke risk reduction.[5-7] The primary deterrent to prescribing warfarin is the fear of inducing life-threatening hemorrhage.[8]

Although concerns about hemorrhage prevent many patients from receiving warfarin, there are relatively little data about the actual rates of death and disability resulting from warfarin-associated hemorrhage, particularly among patients with atrial fibrillation and using current targets of anticoagulation intensity. Although intracranial hemorrhages are widely considered more severe than extracranial hemorrhages, few studies have been able to quantify the actual differences in outcomes between intracranial and extracranial types of hemorrhage. Most studies lacked sufficient numbers of hemorrhagic events to examine the differences in mortality and morbidity; for example, a pooled analysis of 5 randomized trials of warfarin for atrial fibrillation reported a total of only 6 intracranial hemorrhages occurring with warfarin therapy.[2] Our objective was to quantify the risk of death and major disability occurring from warfarin-associated intracranial and extracranial hemorrhages. Such information should allow for more rational choices about the appropriateness of warfarin therapy in patients with atrial fibrillation.

CLINICAL SIGNIFICANCE

- Warfarin-associated intracranial hemorrhages are uncommon but substantially more lethal than other types of warfarin-associated bleeding.
- Because extracranial hemorrhages rarely lead to death or significant disability, the risk for extracranial hemorrhages should be less emphasized when considering the use of oral vitamin K antagonists to prevent stroke in patients with atrial fibrillation.
- Our results highlight the critical need to find better predictors of intracranial hemorrhage and improve risk stratification in individual patients.

METHODS

Cohort Assembly and Ascertainment of Patient Characteristics

The AnTicoagulation and Risk Factors In Atrial Fibrillation (ATRIA) Study is a cohort of 13,559 adults with diagnosed nonvalvular atrial fibrillation who received care within Kaiser Permanente of Northern California, a large integrated health care delivery system. Details of the cohort assembly have been described.[9] Cohort members were identified by searching automated inpatient, outpatient, and electrocardiographic databases for physician-assigned *International Classification of Diseases, Ninth Revisions, Clinical Modification* diagnoses of atrial fibrillation (427.31) between July 1, 1996, and December 31, 1997. Because we were interested in nontransient, nonvalvular atrial fibrillation, we excluded patients with atrial fibrillation and diagnosed mitral stenosis, valvular repair or replacement, transient postoperative atrial fibrillation, or concurrent hyperthyroidism.

The cohort was followed through August 31, 1999, providing a median follow-up of 2.4 years (interquartile range 1.8-2.8 years). Data on patient age and sex were obtained from administrative databases, and data on comorbid medical conditions were collected from validated clinical inpatient and ambulatory databases.[9] To specifically address warfarin-associated hemorrhage, we restricted our analyses to periods of time in which patients were exposed to warfarin, as determined using a previously validated algorithm.[9]

Hemorrhagic Events

We identified potential hemorrhagic events by searching comprehensive automated hospitalization and billing databases for primary and secondary discharge diagnoses of intracranial hemorrhage, and for primary diagnoses of gastrointestinal and other nonintracranial hemorrhages.[10] Warfarin exposure was confirmed through review of the admission medical record. We searched only for primary diagnoses of extracranial hemorrhage, because in a review of 110 randomly selected hospitalizations with only a secondary discharge diagnosis of extracranial hemorrhage, only 4.4% of these were valid major hemorrhagic events. Intracranial hemorrhages were categorized as intracerebral, subdural, or other (eg, subarachnoid) on the basis of radiology reports obtained from the medical records. Intracranial hemorrhages caused by significant head trauma (eg, from antecedent neurosurgery) were excluded. We defined major extracranial hemorrhage as fatal, requiring transfusion of 2 or more units of packed red blood cells, or hemorrhage into a critical anatomic site, such as the retroperitoneum. To restrict analyses to the most serious hemorrhages, events not leading to hospitalization or death were excluded. Because Kaiser Permanente is an integrated health care network, cohort members admitted to facilities outside of the network were still identified by our search strategy.

The emergency department and hospitalization medical records for each potential event were reviewed by a 3-physician Clinical Outcomes committee. Each event was independently validated by 2 physicians on the committee using a formal study protocol, and disagreements were resolved by a consensus of all 3 committee members. The international normalized ratio (INR) at presentation and before reversal of anticoagulation was obtained from medical records, emergency department records, or the inpatient health plan laboratory database.

Each patient's functional disability at the time of discharge was determined on the basis of a physician's review of the medical records, including review of documentation from physician, nursing, physical/occupational therapy, and social work services. Functional disability at the time of discharge was categorized using a modified Rankin scale[11]; the categories were fatal inpatient event, major disability (ie, deficit that prevented independent living), minor disability (ie, residual deficit that did not interfere with independent living), and no disability. This assessment of functional disability has been strongly associated with subsequent death within 30 days in patients admitted with atrial fibrillation-associated ischemic stroke.[4] For 13 patients, the available clinical information was insufficient to determine the extent of functional impairment, and these patients were excluded from the analysis of disability. Deaths during the hospitalization and at 30 days after the event were determined through reviewing medical charts, health plan databases, and the comprehensive California State death certificate registry.[12] Data on mortality were complete for all patients.

Statistical Analyses

Incidence rates of intracranial and major extracranial hemorrhage were calculated as the number of events occurring per 100 person-years of follow-up. We compared functional disability at discharge between patients with intracranial versus extracranial hemorrhagic events using the chi-squared test for categorical variables. We then used multivariable logistic regression to quantify the relative impact of intracranial versus extracranial hemorrhage on 30-day mortality, adjusted for age, INR, and comorbid conditions. To avoid potential overfitting of the model, we included in the multivariable model only comorbid conditions that differed between patients with intracranial and extracranial hemorrhages at a P value of less than .1 on bivariate analyses.

All analyses were conducted using SAS statistical software, version 9.1 (Cary, NC). This study was approved by institutional review boards of the collaborating institutions.

RESULTS

During the study period, we identified 72 patients hospitalized with a validated warfarin-associated intracranial hemorrhage during 15,370 person-years of follow-up on warfarin therapy (unadjusted annualized rate 0.47%, 95% confidence interval [CI], 0.37%-0.59%). Of these, 51 (71%) were intracerebral, 15 (21%) were subdural, and 6 were other or unknown types of intracranial hemorrhages. We also identified 98 patients with a validated warfarin-associated major extracranial hemorrhage during 15,306 person-years of follow-up (annualized rate 0.64%, 95% CI, 0.53%-0.78%). Of these, 87 (89%) involved the gastrointestinal tract. The median age did not differ significantly for patients with intracranial hemorrhage compared with extracranial hemorrhage (Table 1). Compared with patients sustaining an intracranial hemorrhage, more patients with extracranial hemorrhages were female, had a history of diabetes and prior gastrointestinal bleed, and had higher median INRs at presentation, although these comparisons did not reach statistical significance (Table 1).

Outcomes After Major Hemorrhage

Although the overall mortality rate from warfarin-associated hemorrhage was low, the majority of deaths occurred from intracranial hemorrhage. Of the 40 patients who died from warfarin-associated hemorrhage, 35 (88%) died of intracranial hemorrhage (annualized mortality rate of 0.23% [95% CI, 0.16%-0.32%]) and only 5 died of major extracranial hemorrhage (annualized mortality rate 0.03% [95% CI, 0.01%-0.08%]) ($P < .001$, Figure 1). Intracranial hemorrhage continued to be strongly associated with 30-day mortality compared with extracranial hemorrhage even after adjusting for differences in other covariates (age, sex, diabetes mellitus, prior gastrointestinal bleed, and INR at presentation), with an adjusted odds ratio of 20.8 [95% CI, 6.0-72.2]. Among patients sustaining an intracranial hemorrhage, intracerebral hemorrhages resulted in a higher case fatality than subdural hemorrhage (60.1% vs 26.7%, respectively, $P = .04$).

At the time of hospital discharge, patients with intracranial hemorrhage had far more severe functional deficits than

Table 1 Comparison of Clinical Characteristics of Patients with Nonvalvular Atrial Fibrillation Who Were Hospitalized for Intracranial Hemorrhage vs Major Extracranial Hemorrhage While Receiving Warfarin Therapy

Characteristic	Intracranial (n = 72)	Extracranial (n = 98)	*P* value
Median age (interquartile range)	77 (69-82)	77 (71-79)	.6
Female (%)	32	45	.09
Prior stroke (%)	25	19	.4
Prior gastrointestinal bleed (%)	5	14	.07
Congestive heart failure (%)	42	54	.1
Hypertension (%)	67	68	.8
Coronary artery disease (%)	35	46	.2
Diabetes mellitus (%)	19	33	.06
Median INR (interquartile range)	2.7 (2.3-4.0)	3.4 (2.2-5.1)	.1

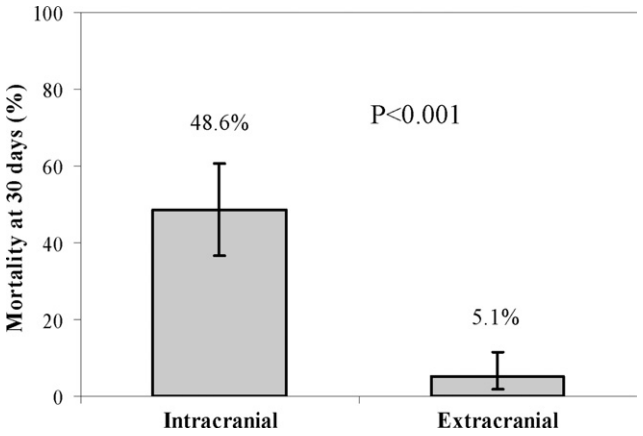


Figure 1 Risk of death 30 days after hospitalization for warfarin-associated intracranial hemorrhage versus major extracranial hemorrhage; 95% confidence intervals (CIs) (*vertical bars*). *P* value refers to the chi-square comparison of mortality rate of intracranial versus extracranial hemorrhage.

did patients with major extracranial hemorrhage (Figure 2). With the exclusion of 13 patients whose discharge disability was unavailable, 26 patients (42%) with intracranial hemorrhage had a fatal inpatient event compared with only 2 patients (2%) with extracranial hemorrhages. Among the 129 survivors at the time of discharge, 21 patients (61%) with an intracranial hemorrhage had major functional disability compared with only 1 patient (1%) with a major extracranial hemorrhage ($P < .001$). Moreover, among the 21 patients who had major disability from intracranial hemorrhage at the time of discharge, 7 died within 30 days.

DISCUSSION

Intracranial hemorrhages accounted for approximately 90% of the deaths caused by warfarin-associated hemorrhage, and most of the functional disability among survivors in this cohort of patients taking warfarin for atrial fibrillation. Although the overall rate of hemorrhagic events in the cohort was relatively low, of those patients who sustained a complication from warfarin, intracranial hemorrhage was the primary determinant of death and disability.

The approximately 50% mortality rate from warfarin-associated intracranial hemorrhage in this study is comparable to that observed in other settings.[13-16] Although intracranial hemorrhages are generally rare events, the adverse consequences resulting from them are considerably higher than those observed from the more common extracranial hemorrhage events. Extracranial hemorrhages, primarily of gastrointestinal origin in this study, are more likely to be remediable events, making residual deficits uncommon. The overall rate of fatal hemorrhage in this present study approximates the rates reported in a recent study from the Leiden Anticoagulation clinic, although that study was not restricted to patients with atrial fibrillation, used higher INR target ranges, and did not examine bleeding-related disability.[17]

Data on the expected outcomes from warfarin-associated hemorrhage are vital to improving how clinicians and patients decide on the use of antithrombotic therapy for atrial fibrillation. Atrial fibrillation-related ischemic stroke is associated with a significant risk of death and major disability, with worse outcomes if the strokes occur while the patient is not taking warfarin.[4] These risks must be balanced against the risk and impact of warfarin-associated hemorrhage. Previous studies generally weighed the risk of ischemic stroke against the risk of all major hemorrhages. Our data demonstrate that it is intracranial hemorrhage that overwhelmingly determines poor outcomes from warfarin, and as a result, the risk of non-intracranial hemorrhage should have a relatively small effect on decisions about warfarin therapy in atrial fibrillation. Although extracranial hemorrhages certainly result in hospitalizations, invasive procedures, and discontinuation of antithrombotic therapy,[18] the longer-term functional impact of such hemorrhages is generally much less than that of intracranial hemorrhages or ischemic strokes. Investigations by Devereaux and colleagues[19] reveal that patients are much more willing than physicians to accept a far higher risk of gastrointestinal hemorrhage in return for an associated reduction in the risk of stroke.

The rates of intracranial hemorrhage on warfarin observed in our study were still considerably lower than the rates of ischemic stroke while the patient was not taking warfarin. As previously reported, the rate of thromboembolism occurring without warfarin therapy was 2.5 per 100 person-years in the overall ATRIA cohort[10] and even higher in other cohorts[2,20-22]; these rates are reduced by more than 50% by warfarin therapy.[10,23] This benefit exceeds the additional risk of warfarin-associated intracranial hemorrhage (0.47 per 100 person-years with warfarin therapy compared with 0.29 per 100 person-years without warfarin therapy),[10,24] and the balance of risk to benefit seems to favor the use of

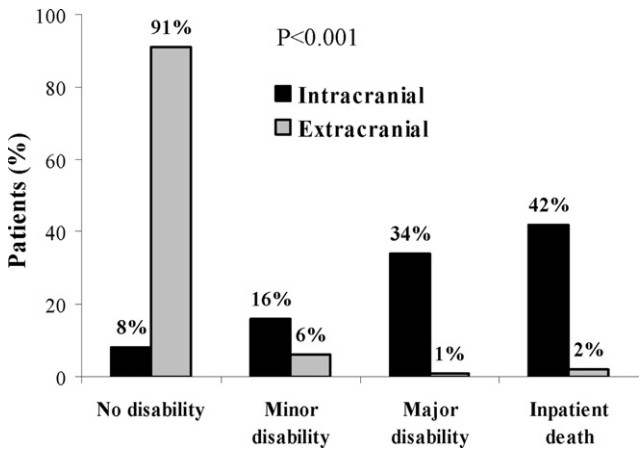


Figure 2 Functional deficit at the time of discharge resulting from warfarin-associated intracranial hemorrhage versus major extracranial hemorrhage in hospitalized patients with atrial fibrillation. *P* value refers to chi-square comparison of discharge deficit between intracranial and extracranial hemorrhage events. Analysis excludes 13 patients whose discharge deficit was not known.

warfarin for the typical patient with atrial fibrillation.[2] Nevertheless, there are likely to be subgroups of patients with atrial fibrillation for whom this balance is less favorable, making further work to identify these subgroups critical to optimizing the use of warfarin at the individual and population levels.

The infrequency of intracranial hemorrhage makes a priori prediction challenging for the individual patient. Older age, elevated INR level, and history of ischemic stroke have all been identified as risk factors for intracranial hemorrhage, but because older age and prior stroke are also risk factors for ischemic stroke, additional investigation into unique risk factors for intracranial hemorrhage is clearly needed.[2,15,24] Modifiable measures, such as maintaining anticoagulation intensity within the therapeutic range of INR 2.0 to 3.0, can reduce but will not entirely remove the risk of bleeding associated with warfarin.[15,25] A study of Medicare patients with atrial fibrillation showed that a high fall risk was associated with an increased risk of intracranial hemorrhage,[16] although the high rate of ischemic stroke (13.7 per 100 person-years) among this subset of patients with a high fall risk made warfarin the preferred option for most patients. Some evidence suggests that testing for the apolipoprotein E genotype or detecting cerebral amyloid angiopathy and leukoaraiosis on brain imaging studies might help refine existing prediction rules for hemorrhage.[26-28] Future validation of the effectiveness of incorporating such predictors into clinical practice is clearly needed to help improve current risk stratification for intracranial hemorrhage.

Our study was strengthened by including the largest number of validated warfarin-related hemorrhages among patients with atrial fibrillation reported to date. Patients were managed using current INR targets, and there was comprehensive assessment of relevant clinical features and subsequent outcomes. However, our study also had several limitations. As an observational study of actual clinical practice, clinicians in our study may have selectively avoided prescribing warfarin to patients at the highest risk for poor outcomes from extracranial hemorrhages. Our observations took place in a health care setting where anticoagulation was predominantly managed by centralized anticoagulation clinics, which may better maintain anticoagulation levels in a therapeutic range.[29,30] However, it is unlikely that these limitations would invalidate the markedly more severe impact of intracranial compared with extracranial hemorrhages. Finally, it is possible that our screening strategy, which identified only hospitalizations for a primary diagnosis of extracranial hemorrhage, could have missed some extracranial hemorrhage events, although prior validation studies of the ATRIA cohort indicate that these exclusions missed only a small fraction of the total number of major extracranial bleeds.[10]

CONCLUSION

Intracranial hemorrhages are the primary determinant of poor outcomes from warfarin-associated hemorrhage, resulting in substantially higher rates of death and disability than major extracranial hemorrhages. As a consequence, rather than basing the anticoagulation decision on a patient's risk for hemorrhage of all types, clinicians should depend primarily on a comparison of the patient's ischemic stroke risk without warfarin therapy with the risk of intracranial hemorrhage with warfarin therapy. Further work is needed to develop better ways to identify patients at the greatest risk for intracranial hemorrhage and more effective methods to mitigate its severe consequences.

References

1. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke.* 1991;22:983-988.
2. 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;154:1449-1457.
3. European Atrial Fibrillation Trial Study Group. Optimal oral anticoagulant therapy in patients with nonrheumatic atrial fibrillation and recent cerebral ischemia. *N Engl J Med.* 1995;333:5-10.
4. Hylek EM, Go AS, Chang Y, et al. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. *N Engl J Med.* 2003;349:1019-1026.
5. Gage BF, Boechler M, Doggette AL, et al. Adverse outcomes and predictors of underuse of antithrombotic therapy in medicare beneficiaries with chronic atrial fibrillation. *Stroke.* 2000;31:822-827.
6. Brass LM, Krumholz HM, Scinto JM, Radford M. Warfarin use among patients with atrial fibrillation. *Stroke.* 1997;28:2382-2389.
7. Fang MC, Stafford RS, Ruskin JN, Singer DE. National trends in antiarrhythmic and antithrombotic medication use in atrial fibrillation. *Arch Intern Med.* 2004;164:55-60.
8. Bungard TJ, Ghali WA, Teo KK, et al. Why do patients with atrial fibrillation not receive warfarin? *Arch Intern Med.* 2000;160:41-46.
9. Go AS, Hylek EM, Borowsky LH, et al. Warfarin use among ambulatory patients with nonvalvular atrial fibrillation: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. *Ann Intern Med.* 1999;131:927-934.
10. Go AS, Hylek EM, Chang Y, et al. Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? *JAMA.* 2003;290:2685-2692.
11. Bamford J, Sandercock P, Dennis M, et al. A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project 1981-1986. *J Neurol Neurosurg Psychiatry.* 1990;53:16-22.
12. Arellano MG, Petersen GR, Petitti DB, Smith RE. The California Automated Mortality Linkage System (CAMLIS). *Am J Public Health.* 1984;74:1324-1330.
13. Hylek E, Singer D. Risk factors for intracranial hemorrhage in outpatients taking warfarin. *Ann Intern Med.* 1994;120:897-902.
14. Berwaerts J, Dijkhuizen RS, Robb OJ, Webster J. Prediction of functional outcome and in-hospital mortality after admission with oral anticoagulant-related intracerebral hemorrhage. *Stroke.* 2000;31:2558-2562.
15. Fang MC, Chang Y, Hylek EM, et al. Advanced age, anticoagulation intensity, and risk for intracranial hemorrhage among patients taking warfarin for atrial fibrillation. *Ann Intern Med.* 2004;141:745-752.
16. Gage BF, Birman-Deych E, Kerzner R, et al. Incidence of intracranial hemorrhage in patients with atrial fibrillation who are prone to fall. *Am J Med.* 2005;118:612-617.

17. Torn M, Bollen WLEM, van der Meer FJM, et al. Risks of oral anticoagulant therapy with increasing age. *Arch Intern Med.* 2005;165: 1527-1532.
18. DiMarco JP, Flaker GC, 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.
19. Devereaux PJ, Anderson DR, Gardner MJ, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ.* 2001;323:1218-1222.
20. Hart R, Pearce L, McBride R, et al. Factors associated with ischemic stroke during aspirin therapy in atrial fibrillation: analysis of 2012 participants in the SPAF I-III clinical trials. *Stroke.* 1999;30:1223-1229.
21. Gage BF, Waterman AD, Shannon W, et al. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA.* 2001;285:2864-2870.
22. Wang TJ, Massaro JM, Levy D, et al. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. *JAMA.* 2003;290:1049-1056.
23. Singer DE, Albers GW, Dalen JE, et al. Antithrombotic therapy in atrial fibrillation: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest.* 2004;126:429S-456S.
24. Fang MC, Go AS, Hylek EM, et al. Age and the risk of warfarin-associated hemorrhage: the anticoagulation and risk factors in atrial fibrillation study. *J Am Geriatr Soc.* 2006;54:1231-1236.
25. Rosand J, Eckman MH, Knudsen KA, et al. The effect of warfarin and intensity of anticoagulation on outcome of intracerebral hemorrhage. *Arch Intern Med.* 2004;164:880-884.
26. O'Donnell H, Rosand J, Knudsen K, et al. Apolipoprotein E genotype and the risk of recurrent lobar intracerebral hemorrhage. *N Engl J Med.* 2000;342:240-245.
27. Rosand J, Hylek EM, O'Donnell HC, Greenberg SM. Warfarin-associated hemorrhage and cerebral amyloid angiopathy. *Neurology.* 2000; 55:947-957.
28. Hart RG, Tonarelli SB, Pearce LA. Avoiding central nervous system bleeding during antithrombotic therapy: recent data and ideas. *Stroke.* 2005;36:1588-1593.
29. Chiquette E, Amato M, Bussey H. Comparison of an anticoagulation clinic with usual medical care. *Arch Intern Med.* 1998;158: 1641-1647.
30. Fitzmaurice D, Hobbs R, Murray E, et al. Oral anticoagulation management in primary care with the use of computerized decision support and near-patient testing. *Arch Intern Med.* 2000;160:2343-2348.

J Med Syst (2007) 31:49–62
DOI 10.1007/s10916-006-9042-z

ORIGINAL PAPER

Design and Development of a Mental Health Assessment and Intervention System

Ramesh Farzanfar · Allison Stevens · Louis Vachon · Robert Friedman · Steven E. Locke

Received: 24 July 2006 / Accepted: 22 September 2006 / Published online: 19 December 2006

Abstract Mental health disorders are the leading cause of disability and functional impairment in the United States (1 in 5). The negative effect of mental health disorders is felt both in the personal and public lives of the affected individuals, particularly in the workplace where it adversely impacts productivity. Only a small fraction of the affected people in the work force seeks help. The cost to employers and the economy of these untreated individuals is staggering. Some employers have tried to address employees' emotional well-being by establishing Employee Assistance Programs. Yet, even these programs do not sufficiently address existing barriers to the detection and treatment of mental health disorders in the workplace. This paper describes the design of an automated workplace program that uses an Interactive, computer-assisted telephonic system (Interactive Voice Response or IVR) to assess workers for a variety of mental health disorders and subsequently refers untreated and inadequately treated workers to appropriate treatment settings.

R. Farzanfar · A. Stevens · R. Friedman
Medical Information Systems Unit,
Boston University Medical Campus,
Boston, Massachusetts

R. Farzanfar (✉)
MISU, 560 Harrison Avenue, Suite 404,
Boston, Massachusetts 02118
e-mail: rfarzanf@bu.edu

L. Vachon
Department of Psychiatry, Boston University
Medical Campus,
Boston, Massachusetts

S. E. Locke
Center for Clinical Computing and Department of Psychiatry,
Beth Israel Deaconess Medical Center,
Boston, Massachusetts

Keywords System Design · Automated assessment of mental health · IVR systems · Workplace productivity

Introduction

Mental health disorders are the leading cause of disability and among the most prevalent of chronic diseases in the United States and worldwide [1]. More than 54 million Americans (1 in 5) have a mental health disorder in a given year [2]. and approximately 15% of all adults who have a mental health disorder also experience a co-occurring substance (alcohol or drug) abuse disorder which complicates treatment [3].

Data on functional impairment in the population demonstrate that up to 9% of U.S. adults experience significant functional impairment as a result of mental health disorders. Of these individuals, 7% suffer from chronic mental disorders (for at least one year) [6]. There is also evidence that untreated comorbid mental health disorders are associated with poorer outcomes among individuals with physical disorders [7]. Although mental health disorders are treatable, only about 8 million individuals with mental health disorders (16%) seek treatment [3]. The cost of these untreated individuals to employers and the nation's economy is staggering [8, 9]. In most circumstances, however, mental health issues, although treatable, are neglected in the workplace until they cause significant problems such as reduced productivity and/or absenteeism [10, 11].

To respond to this epidemic, some employers have taken on the responsibility of addressing employees' psychological and emotional well-being by establishing Employee Assistance Programs (EAPs) or other activities such as Annual Depression Screening Days. Yet, even these programs have generally not addressed the substantially high barriers

to the detection and treatment of mental disorders in the workplace. Although EAPs are geared to helping employees with personal, family and work concerns, most workers do not take advantage of the services provided by the EAPs or other employer-sponsored resources [12]. Also, many in-house programs are cost-conscious and thus only devote their resources to immediate and obvious problems [13]. Often, other barriers such as the stigma of mental and emotional health problems, and/or the employees' lack of knowledge about EAPs prevent them from seeking help in the work environment. There is a clear need for comprehensive but cost-efficient workplace programs for the detection and treatment of mental health disorders that address all employees and preserve privacy.

This paper describes the design and development of a confidential workplace program that focuses on workplace productivity and symptom improvement of employees who have undiagnosed and/or untreated (or partially treated) mental health disorders. The program uses a computer-assisted telephonic system, based upon Interactive Voice Response (IVR) to assess workers for a mental health disorder that could adversely impact their workplace performance. The program then refers affected workers who are untreated or who are inadequately treated to appropriate treatment settings. The development of this system is supported by a grant funded by the Centers for Disease Control and Prevention (CDC), R01 DP000116. The grant proposal was a response to a well-publicized CDC initiative with the principal goal of assisting employers to address factors that negatively impact employees' productivity. The ultimate objective of this system is to increase productivity by reducing absenteeism and partial functionality (presenteeism) related to untreated or undertreated mental illness in the employee population.

The Telephone-Linked Communications for Detection of Mental Health Disorders in the Workplace (TLC-Detect) assesses each participating employee for a group of relevant mental health disorders by administering a panel of reliable and validated assessment instruments. If an employee screens positive for a mental health disorder, the system provides information, education and referrals to a variety of mental health care resources and treatment options specifically for that disorder. The system also motivates employees to engage in treatment and implements automated periodic follow-ups to monitor mental health symptom severity and treatment maintenance. Finally, the system motivates the employees to remain in treatment.

The TLC-Detect program is being evaluated in a randomized controlled trial to examine its impact on productivity measures (absenteeism and presenteeism), the cost-effectiveness of the intervention and the feasibility of using TLC-Detect in a workplace setting. Below, we will describe the TLC technology on which the mental health assessment system is based.

The TLC technology: An overview

The Telephone-linked communications (TLC) system

TLC is a totally automated telecommunications system that has been used to screen patients and consumers (individuals who do not have a known health condition) with specific health conditions, promote self-care behaviors, help patients and consumers modify their unhealthy behaviors and monitor patients with chronic disease and alerting their providers about clinical problems [14, 15]. The TLC health condition assessment applications have been applied to the assessment of problem drinking behavior [16], chemotherapy related symptoms [17] and depression symptoms [18]. TLC applications that address patient/consumer behavior modification include systems for diet and physical activity [19, 20], smoking cessation [21]. on time mammography, medication adherence in hypertension [22], depression [23], and other chronic diseases as well as maintenance of healthy behavior. The TLC disease management applications include those for angina, COPD [24], children asthma, etc.

During TLC conversations, TLC speaks to the users over the telephone using computer-controlled pre-recorded human speech. The users communicate with TLC by speaking into the telephone receiver or by using the touch tone keypad. TLC conversations last from a few minutes to 30 min depending on the particular TLC program. In addition to asking questions, TLC provides information, advice, and depending on the program, behavioral counseling. The specific questions and the responses given to the user depend upon the behavioral theory or model used in the program design as well as what the user reported during previous conversations, and earlier during the same conversation. TLC combines an interactive voice response (IVR) subsystem for generating speech over the telephone, a speech recognition subsystem for recognizing what the user is saying, a database management subsystem for storing and managing system and user data, and a conversation control subsystem. The design of the TLC behavioral interventions is based on behavioral theory (the Transtheoretical Model [25, 26]), Social Cognitive Theory [27], and on the heuristics of experienced counselors, including telephone counselors.

TLC-Detect's design objectives: General considerations

TLC-Detect is designed to deliver an interactive, confidential and standardized mental health assessment and treatment promotion program to employees over time. The system may be accessed from work or from home or anywhere else depending on the user's preference as it involves only a telephone call. TLC-Detect was deliberately designed to respond to keypad data entry in lieu of speech recognition because the

designers understood the employees' concerns about confidentiality and because pressing the numbers on the keypad, even when people are at work, can be carried out without any disclosure of the subject matter to those who may be nearby.

The principal objectives of the TLC-Detect system are: 1) to screen employees of a participating organization for the presence of a mental or emotional disorder, and 2) to persuade them to seek help. The system was designed to screen both the employees who have never been screened positive for a mental/emotional health disorder as well as those who may have been diagnosed with a mental health disorder (and were possibly treated) but are not currently under treatment. The goal is for the system to assist a distressed employee by evaluating his or her condition for the presence of symptom patterns seen in nearly 20 disorders, using validated and standardized questionnaires as evaluative tools. Once an employee is determined to be positive for a pattern consistent with a specific disorder, the system delivers an education/intervention program that provides multiple opportunities for the person to hear a variety of available options for referral and treatment. This includes providing recommendations to the employees about obtaining professional help or engaging in self-help based on the disorder they may be suffering from and taking into account the level of symptom severity. TLC-Detect then educates and motivates the employees to engage in treatment and/or self-care. Following this, the system will monitor the employees' mental health symptoms and treatment maintenance over time and will motivate them to remain in treatment by calling them periodically to monitor their progress. Contacts between the employees and TLC-Detect occur on a monthly basis after the assessment and subsequent intervention and referral have been offered. The major thrust of the conversation throughout the program is to encourage the employee to seek help. Once the employee has sought and received mental health care, it is the primary hypothesis of this study that the workplace productivity will be improved.

The TLC-Detect system consists of the following components: (1) a Mental Health Assessment Module that screens for symptom patterns consistent with common mental health disorders that are likely to affect workplace productivity, (2) a Treatment Module that educates the user about an identified mental health condition, offers referrals based on the severity of the mental health condition, and assesses the willingness of the employee who has a mental health condition to engage in treatment, and (3) a Follow-up Treatment Module that will monitor and promote the person's treatment maintenance. The overall outline and flow of calls for the duration of the study are demonstrated in Fig. 1.

The content of the screening and symptom assessment dialogues was based on reliable and valid mental health screening instruments. The content of the treatment and follow-up dialogues was based on the recommendations of two experienced mental health care clinicians. Below, we will describe a few additional components of the TLC-Detect System.

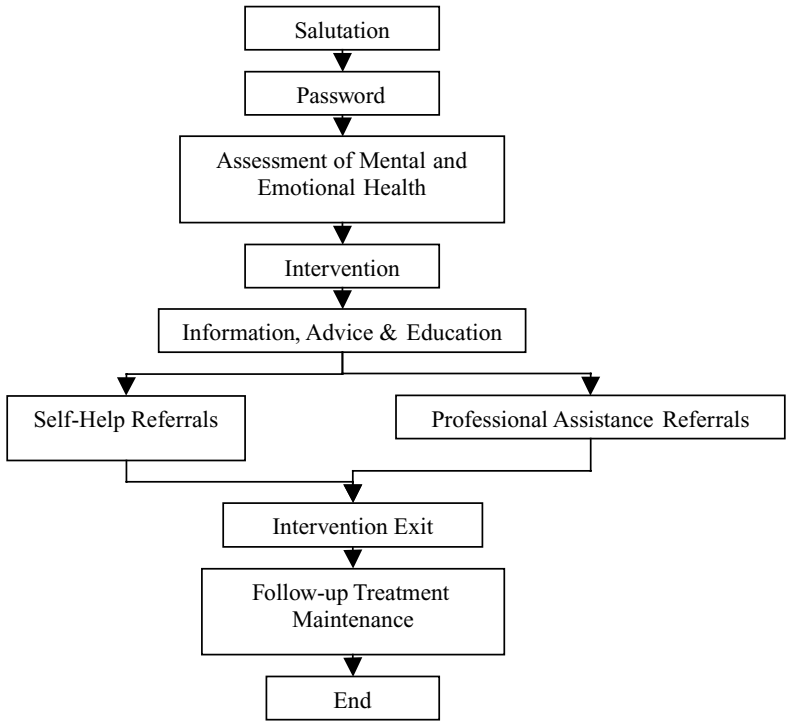


Fig. 1 Flow of the overall structure of TLC-detect's conversation

The TLC-Detect's frame

Like all TLC applications TLC-Detect is comprised of structural components that are designed to ensure that the system is user-friendly. These components consist of: 1) A greeting and salutation segment that initiates all conversations. After greeting, the user is asked to enter a password to ensure confidentiality. 2) A TLC Trainer component that consists of instructions on how to use the system. These instructions are carried out by TLC itself. TLC takes the user step-by-step through the process of using the system with examples of the types of questions that are asked during actual TLC conversations, how to respond to questions, how to repeat a question, how to change a response, etc. 3) A TLC Helpline component which is an automated messaging center that provides assistance to subjects if questions arise in the use of the TLC system. Subjects can independently call the Helpline if they experience difficulties using the TLC system or if they have questions about how to use it. The Helpline allows the user to leave a standard voicemail message that will be retrieved by the project staff. The TLC-Helpline is available to users 24 hours a day, 7 days a week. Messages left on the Helpline will be addressed within 24 hours or the next business day, during weekends and holidays. Upon receiving the message, a staff person will follow-up with the study participant to troubleshoot. If for any reason a study participant chooses to use the Helpline during a conversation with the TLC-Detect, s/he can also choose to hear the instructions in the TLC Trainer again. Furthermore, this function can also be used to end a current call and schedule another one to complete the conversation or leave a message for the study staff. 4) The TLC Call Scheduling: in most TLC applications, this particular component determines follow-up call schedules for all users. However, for TLC-Detect this component controls additional functions which include scheduling calls for users who have more than one disorder and thus need to receive further education and information as well as those who need more than one referral to mental health care resources. This component also controls calls made to users who need to end their conversation with TLC-Detect prematurely. These individuals usually receive a call from TLC-Detect the day after the conversation. If they are not available, TLC can leave a message either on an answering machine or with the person who answers the phone. In either case, strict confidentiality is observed as the message will only contain a few words reminding the user to call TLC. 5) TLC-Free Speech component that is intended to perform the following functions: a) to reduce users' frustration due to the closed ended nature of the interaction, at the end of each conversation, users are asked whether they would like to leave a message for the research staff, b) users are asked to inform the study staff about potential difficulties they encountered during system utilization, and c) during the follow-up conversations users are requested to leave a message to inform the study staff about any additional resources they might have utilized to seek treatment that were not offered by TLC-Detect.

TLC Detect's screening functions

An important function of TLC-Detect is to screen employees for the presence of the following mental health disorders: somatization, major & minor (atypical) depression, suicidal ideation, acute bereavement, post-partum depression, dysthymia, panic disorder, generalized anxiety disorder, social phobia, post-traumatic stress disorder, acute stress disorder, bipolar disorder, alcohol and drug abuse and dependence, whether they have recently been a victim of violence and general stress such as financial, marital, family and work-related problems. Subjects are also screened for the degree of functional impairment they experience and the amount of social support they receive as mitigating factors that could influence the severity of their condition.

TLC-Detect not only evaluates users for symptom severity, it can also identify disorders at subsyndromal levels, i.e., when symptoms do not meet full diagnostic criteria for a particular disorder but may still be problematic. Cut-offs for symptom severity and subsyndromal levels are based on the recommendations made in the literature for each screening instrument. In cases where such recommendations were not available, we based these cutoffs on the recommendations of mental health care clinicians who advised us during the design phase as well as the diagnostic criteria described in the DSM-IV.

The Patient Health Questionnaire (PHQ) has been automated to screen for 9 of the above disorders [28, 29]. The PHQ is the self-administered version of the Primary Care Evaluation of Mental Disorders (PRIME-MD). PRIME-MD was developed as a screening instrument for mental disorders in a primary care setting. PRIME-MD is designed to be administered by a clinician while the PHQ is self-administered and thus more suitable for an IVR system. The PHQ screens for somatization, major and minor (atypical) depression, generalized anxiety disorder, panic disorder, eating and alcohol disorders, functional impairment, stress, and victimization due to violence as well as menstruation complications in women. To make the screening more appropriate for assessment of mental health among employees, we removed a number of modules from the PHQ and added others. For example, questions screening for alcohol disorders were replaced with several instruments that screened for risky alcohol use, alcohol abuse, and alcohol dependence since this distinction is important for treatment. Moreover, we felt that some of these modules would be relevant only to a small minority of the individuals in the target population or were already addressed by other questions in the PHQ (e.g.,

problems with menstruation and menstrual cramps). We also used other validated instruments to screen for important disorders that were not included in the PHQ (e.g., bipolar disorder, social phobia, post-traumatic stress disorder). The PHQ has both binary (yes/no format) and interval scaled questions. We made no modifications to the content or format of the PHQ questions. The PHQ has been evaluated for reliability and validity with acceptable results [28].

Unspecified emotional distress

To enter the study, employees must screen positive for a certain level of emotional and mental health stressors. However, some study participants while eligible for the study might end up screening negative for all disorders assessed by TLC-Detect due to any of the following reasons: 1) a person's actual symptom patterns may not be consistent with any of the disorders included in the TLC-Detect's design, 2) the system may lack the sensitivity and specificity to accurately assess the presence of a particular condition, and 3) some individuals may not disclose their true feelings or experience during the screening interview. As a result, assessment of their mental and emotional health status may be inaccurate. We, therefore, included in the design an "Unspecified Emotional Distress" section to help those employees who responded positively to eligibility questions that required the subject to report the presence of some degree of stress in their lives. All employees who are directed to this section will receive general information, education and referrals for stress management and self-care strategies that would be helpful to them regardless of the specific disorder they might have.

General architecture of the TLC-Detect's assessment module

The first session in this study consists of an extensive interview conducted by the research staff during which eligibility criteria are assessed and the outcome questionnaires are administered to the eligible participants. The interview process is totally automated and is carried out by a Computer Assisted Telephone Interview (CATI) system. The CATI interviews help bypass time-consuming data coding, editing, and entry processes and thereby reducing interviewer error. It also allows for complex question branching and immediate scoring of study questionnaires that is accomplished automatically by the computer. Upon completion of baseline demographic questions, outcome data collection and randomization of participants into experimental and control groups, all participants are connected to the TLC-Detect system by the interviewer. Assessment is carried out during this first contact with TLC-Detect. The system begins by welcoming users and subsequently offering training on how to use the system. Subsequently, TLC-Detect leads the user through the first level of assessment. The initial assessment consists of questions that all users receive. Based on the users' response to these questions, further screening for each disorder may be conducted by immediately taking users through additional and more detailed (specific) questions.

Since the first call is relatively long, users will be offered the option to complete the call in more than one session. In this case, the next call should occur as soon as possible (within a week). Those who might be cut off from TLC-Detect (due to deliberate hang ups or technical problems) are also encouraged to call back to complete the call as soon as possible. Those users who fail to call back will receive a call from TLC-Detect within a day or two.

The system uses the PHQ to screen for somatization [30] depression, panic disorder and generalized anxiety disorder. We designed the system so that it could screen for various forms of depression such as major and minor (atypical) depression, dysthymia, (a less severe, yet more chronic form of depression), postpartum depression and acute bereavement. [This strategy helps in directing users to the appropriate treatment recommendations and prevents the overlapping of intervention and referral recommendations.]

An important component of depression screening is assessment of suicidality. Based on a user's response to the suicidality question (the PHQ question #2i), three additional questions are asked to determine a risk level from low to urgent. Based on the risk level, TLC-Detect 'suggests,' 'recommends' or 'urges' subjects to call a suicide hotline number, their health care provider or if in need of immediate help, to go to the emergency room or call 911. For all but the highest risk level, TLC-Detect offers users the option to end the call so they can contact someone for help. For the highest risk level, even though the call is ended and users are urged to seek help, the study's Principal Investigator and a staff psychiatrist are notified so that they may follow-up with the user within 24 hours.

Additional assessments include screening for social phobia (PHQ and Social Phobia Inventory) [31], post traumatic stress disorder and acute stress disorder (PHQ, Impact of Events Scale [32–34], Acute Stress Disorder Scale) [35–37], as well as bipolar disorder (PRIME-MD and Mood Disorder Questionnaire) [38–40].

The system then proceeds to screen users for alcohol and drug problems. TLC-Detect was designed to address three categories of alcohol and drug problems based on the DSM-IV criteria and the recommendations of experienced substance abuse professionals [41, 42]. The three categories for alcohol use problems are: 1) risky alcohol use (Single Alcohol Screening Question or SASQ and a Quantity-Frequency question), [42, 43] 2) alcohol abuse (Short Index of Problems) [44], and 3) alcohol dependence (Composite International Diagnostic Interview Short Form) [45, 46]. A

positive response to risky alcohol use will automatically refer users to sub-modules that ask questions assessing alcohol abuse and dependence. Similarly, for drug problems, the three categories are: 1) drug use (Center for Substance Abuse Treatment TIP-24). A negative response to this question triggers two additional questions that address prescription drug abuse. A positive response to any of these questions will trigger questions for the following two categories of drug abuse and dependence, 2) drug abuse (DSM-IV for substance abuse), and 3) drug dependence (Section H of the CIDI-SF). It is important to note that users will receive an intervention for only one of these three categories (of substance abuse, alcohol or drugs) even if a user screens positive for more than one category. The intervention provided will target the most severe category into which a subject falls. This means that users who screen positive for dependence are not screened for abuse and all users that are positive for abuse or dependence would have also been positive for the initial screening questions. Therefore, if a user screens positive for risky alcohol use and alcohol dependence, s/he will receive the intervention for alcohol dependence and if a user screens positive for drug use and drug abuse, s/he will receive the drug abuse intervention. In both alcohol and drug dependence, TLC-Detect asks an additional question that determines the degree of functional impairment caused by such dependence (on a scale from 0 to 4, with a score of 4 being the most severe). TLC-Detect also assesses the degree of functional impairment due to substance abuse on a scale from 0 to 4. Based on the users' answers to these questions, users that score higher than a 2 on either the functional impairment scale or admit they use more than one type of drug are moved up to the next severity level for that disorder. The reason for weighing the severity level is the understanding that those who admit they have difficulties with their family, work and social relationships and those who use more than one type of drug tend to have a more severe drug problem [47, 48].

To determine whether the user has ever been a victim of violence, TLC-Detect asks question #13 of the PHQ [have you been hit, slapped, kicked or otherwise physically hurt by someone or has anyone forced you to have an unwanted sexual act?]. Users will be allowed to skip this question if they are unwilling to respond to it. TLC-Detect will keep track of how many users do in fact skip this question.

Functional impairment, social support, stress: Treatment intensity adjuster (TIA) score

In order to provide appropriate intervention to users it is important to determine the degree to which a particular disorder has had a harmful impact on users' family, work and social life. The system designers felt that availability of social support and experiencing daily stressors are significant factors contributing to an individual's ability to cope. As a result, a Treatment Intensity Adjuster (TIA) was created that takes into consideration functional impairment, social support and presence of stress. Functional impairment is determined by the PHQ's question #11 [49]. An initial screening to assess stress is carried out by using the Perceived Stress Scale-4 [28, 50]. We wrote three questions ourselves to determine availability of social support to users. Inadequate social support was identified by whether users have someone to turn to for help, someone to confide in and how often users felt lonely.

The scores from the above factors are added to create the TIA score. Based on the TIA, if a user's response to the "stress," "social support" and "functional impairment" questions is cumulatively greater than 9, then the severity level of the disorder for that particular user would be moved up one level [51]. For example, if a person's score for depression is mild, however, that person's cumulative TIA score is greater than 9, then that person's depression severity level is moved up to "severe." In this manner, TLC-Detect identifies a person's inability to cope with their emotional distress factors and will address this appropriately.

General architecture of TLC-Detect's treatment/intervention module

The TLC-Detect's Treatment/Intervention Module is designed to be seamlessly presented to employees immediately after they have been screened for mental health disorders by the system. The objective of this module is to motivate the employee to enter into and sustain an effective treatment situation for their mental health condition(s). [Specific intervention options tailored to particular disorders are offered to users immediately after they complete the screening.] The Module presents the following topics: 1) a description of the user's specific condition, 2) availability of effective treatment to help the employee deal with the distress they are feeling, 3) an assessment of the employee's willingness to engage in mental health care, and 4) referrals for self-help and/or professional assistance.

Co-morbid disorders

When co-morbid mental disorders, i.e., presence of more than one co-occurring disorder, are identified by TLC-Detect, users will be provided with the intervention for all of their identified disorders. This creates a challenge to the system: to assign priorities for treatment and self-care recommendations for a variety of co-occurring disorders. Since receiving an intervention for more than one disorder would be

Table 1. Rank order of disorder severity

Disorder Rank	Group Rank 1	Group Rank 2	Group Rank 3	Group Rank 4
1	Violence	Bipolar disorder	Bipolar disorder previously diagnosed	Social phobia
2	Postpartum depression	Major depression	Dysthymia	Panic attack
3		PTSD	Minor depression	Somatization
4		Acute stress disorder	Generalized anxiety disorder	
5		Panic disorder	Alcohol risky use	
6		Alcohol dependence	Acute bereavement	
7		Drug abuse	Stress	
8		Alcohol abuse		

burdensome during one telephone call, the system is designed to parse the interventions and logically distribute them over several subsequent sessions with the proviso that users with co-morbidities receive all of the intervention content within the first month of their participation. These "co-morbid" employees are encouraged to speak with TLC-Detect once a week until they have been exposed to all of the prioritized and queued intervention content. If these employees happen to have more disorders than the number of weeks in a month, the system divides the calls evenly depending on the number of the disorders. These calls will be considered, "Continued Intervention" Calls and not "follow-up" calls. In order to give users options and to complete the intervention as soon as possible, during the (first) intervention call and throughout these "continued intervention" calls, users are asked if they would like to receive interventions for additional disorders during the same call, allowing them to determine the pace and intensity of the intervention. It is, thus, possible for a person who suffers from co-morbid disorders to complete assessment and the intervention for all of his/her disorders in one telephone call, contingent upon their motivation, persistence and time availability.

The intervention module

The intervention module consists of two sub-modules: education and referral. All users who screen positive for one or more disorders will receive education for each disorder identified, regardless of severity. The education component always begins with a description of the disorder, its common symptoms and usual treatment methods as well as a few tips and practical advice on self- help approaches. Referrals fall into two different categories: *self-management and professional assistance*. Within each category, there are several components that refer users to the appropriate resources for self-help as well as professional assistance. For example, *self-management* may include advice and information on the benefits of regular exercise and healthy diet, instruction and education for meditation, information about where to obtain self-help workbooks, and information on how to access support groups, when available. *Professional assistance* implies the involvement of a mental health professional in the intervention and may include recommendations on receiving assistance from a health care provider to manage stress, information on group therapy, and other information related to Employee Assistance Programs and behavioral health evaluation.

The education and information that are provided as part of the intervention will be rank ordered based on the significance and severity of the disorder. To operationalize rank-ordering of the disorder severity, we created a Matrix (see Table 1).

Based on the Matrix in Table 1, to users with more than one disorder, TLC-Detect will recommend for which disorder they should listen to an intervention first. The recommendation is based on the group rank a given disorder is assigned into (inter-group ranking). The intervention for a disorder that belongs to the group ranked #1 is recommended before a disorder that falls into the group ranked #3. In addition, if a user has more than one disorder in one group, TLC-Detect's recommendation will be based on the severity level of the disorders within that group (intra-group ranking). For example, if a person has mild bipolar disorder together with severe panic disorder (both in group #2), the person will be provided information and counseling about panic disorder first. If both disorders happen to have the same severity level, then TLC-Detect will choose a disorder based on its intragroup rank. The intragroup rank is determined by the disorder's position on the group list. The highest rank (most severe disorder within the group) is at the top of the group list and the lowest rank (least severe disorder within the group) is the last on the list. For example, if a participant has severe bipolar disorder and severe post-traumatic stress disorder, s/he will hear about bipolar disorder first because it is placed above PTSD on the list. (Disorders with subsyndromal severity are presented last regardless of [inter or intra] group assignment.)

Table 2 shows how provision of referrals to patients with different disorders was operationalized. Based on the

Table 2 Intervention referrals based on disorder and severity

			Self-management					Professional assistance		
Disorder	Instruments	Severity	Exercise	Diet	Meditation	Workbook	Support Group	Stress management	Group Therapy	EAP/BH eval
Somatization	PHQ (Patient Health Questionnaire)	Subsyndromal	4	–	3	–			3	–
		Moderate	4	–	3	–		2	3	–
		Severe	–	–	–	–		3	3	2
Major Depression	PHQ	Subsyndromal	3	4	–	3	4	5	–	–
		Moderate	3	4	–	3	4	5	3	2
		Severe	–	–	–	–	–	–	–	1
Minor (Atypical) Depression	PHQ	N/A	3	4	–	3	4	5	–	4
Acute Bereavement	CIDI-SF	N/A	–	–	–	–	2	–	3	3
Postpartum Depression	CIDI-SF	N/A	–	–	–	–	–	–	–	1
Suicide	PHQ & Locke/Vachon (1)	Low risk	–	–	–	–	–	–	–	3
		Medium risk	–	–	–	–	–	–	–	3
		High risk	–	–	–	–	–	–	–	3
		Urgent risk	–	–	–	–	–	–	–	1
Dysthymia	PRIME-MD (2)	Subsyndromal	3	4	–	3	4	5	3	–
		Moderate	3	4	–	3	4	–	3	2
		Severe	–	–	–	3	4	–	3	2
Panic Disorder	PHQ	Subsyndromal	–	–	–	2	3	4	–	–
		Moderate	–	–	–	3	4	–	3	2
		Severe	–	–	–	3	4	–	3	2
Panic Attack	PHQ	Moderate	–	–	–	–	–	–	–	2
		Severe	–	–	–	3	–	–	–	2
Genralized Anxiety Disorder	PHQ	Subsyndromal	4	–	3	3	4	2	–	–
		Moderate	4	–	3	3	4	2	3	2
		Severe	–	–	–	2	3	3	3	1
Social phobia	SPIN (Social Phobia Inventory)	Subsyndromal	–	–	–	3	–	–	–	–
		Moderate	–	–	–	3	3	2	3	2
		Severe	–	–	–	4	–	–	3	2
Posttraumatic Stress Disorder	IES (Impact of Event Scale)	Mild	–	–	–	3	3	–	–	4
		Moderate	–	–	–	4	4	–	3	2
		Severe	–	–	–	4	4	–	3	2

Table 2 Continued

Disorder	Instruments	Severity	Self-management					Professional assistance		
			Exercise	Diet	Meditation	Workbook	Support Group	Stress management	Group Therapy	EAP/BH eval
Acute Stress Disorder	ASDS (Acute Stress Disorder Scale)	Subsyndromal	–	–	–	3	3	–	–	4
		Moderate	–	–	–	4	4	–	3	2
		Severe	–	–	–	4	4	–	3	2
Bipolar Disorder	MDQ (Mood Disorders Questionnaire)	Subsyndromal	–	–	–	–	–	2	–	4
		Low	–	–	–	2	2	–	–	3
		Moderate	–	–	–	2	–	2	–	2
		Severe	–	–	–	3	–	–	–	1
Bipolar Disorder, previously diagnosed	PRIME-MD	Mild	–	–	–	2	2	–	–	3
		Moderate	–	–	–	3	3	–	–	2
		Severe	–	–	–	–	–	–	–	2
Alcohol Dependence	CIDI-SF (Composite International Diagnostic Interview-Short Form)	Mild	4	–	4	3	2	2	–	–
		Moderate	4	–	–	3	2	–	3	1
		Severe	–	–	–	–	2	–	2	1
Alcohol Risky Use	SASQ (3)	N/A	4	–	4	3	2	2	–	–
Alcohol Abuse	SIP (Short Index of Problems)	Mild	4	–	4	3	2	2	–	–
		Moderate	4	–	–	3	2	–	3	1
		Severe	–	–	–	–	2	–	2	1
Drug Use	CSAT TIP #24 (4)	N/A	4	–	4	3	2	–	3	2
Drug Abuse	CIDI-SF & DSM-IV (5)	Mild	4	–	4	3	2	–	3	2
		Moderate	–	–	–	3	2	–	3	1
		Severe	–	–	–	–	2	–	2	1
Drug Dependence	CIDI-SF	Mild	4	–	4	3	2	–	3	2
		Moderate	–	–	–	3	2	–	3	1
		Severe	–	–	–	–	2	–	2	1
Stress	PHQ & PSS-4 (6)	Mild	–	4	4	3	3	2	–	–
		Moderate	–	4	4	3	3	2	–	–
		Severe	–	4	4	3	2	3	–	–

Table 2 Continued

Disorder	Instruments	Severity	Self-management					Professional assistance		
			Exercise	Diet	Meditation	Workbook	Support Group	Stress management	Group Therapy	EAP/BH eval
Violence	PHQ	N/A	–	–	–	–	–	–	–	2
Unspecified Emotional Distress	Negative for all instruments	Mild	4	4	4	3	3	2	–	4
		Moderate	4	4	4	3	3	2	–	4
		Severe	4	4	4	3	2	3	–	4

Note. Intervention Key:

Critical = 1.

Strongly Recommended = 2.

Recommend = 3.

Suggest = 4 & 5.

1 – Mental Health Care Clinicians' recommendations.

2 – Primary Care Evaluation of Mental Disorders.

3 – Single Alcohol Screening Question & Quantity-Frequency Question.

4 – Center for Substance Abuse Treatment: Treatment Improvement Protocols #24.

5 – Diagnostic and Statistical Manual of Mental Disorders - IV.

6 – Perceived Stress Scale.

[a]Functional Impairment and Social Support are measured in order to weight the severity level. Functional Impairment is measured by the PHQ. Social Support is measured based on 3 questions developed by our mental health care clinicians.

experience of the clinical experts who participated in the intervention design, mental health conditions were deemed deserving of varying intensities of referral urgency. In this Table, disorders and severity levels are identified in two separate columns. The table includes a legend based on which TLC-Detect's recommendations to users are graded from "suggestion" to "critical." For example, to a person who has screened positive for moderate dysthymia, TLC-Detect "strongly recommends" contacting their Employee Assistant Program or seeking consultation with a mental health professional while also "recommending" exercise and self-help work books and "suggesting" a healthy dietary regimen. Or, to a person who is screened positive for severe bipolar disorder, TLC-Detect "recommends" a self-help workbook, and also states that it is "critical" for that person to seek evaluation by a mental health care professional or to consult with the Employee Assistance Program.The numbers in the Table thus denote the specific type of recommendation TLC-Detect provides to users during a given conversation. Furthermore, referrals signified by numbers 1 and 2 are always offered during the first intervention call while those denoted by numbers 3, 4 and 5 are presented during the follow up calls.

The follow-up and maintenance periods

After the initial assessment and intervention are completed, subjects will be followed for a period that may take approximately twelve months. This period is divided into two phases: follow-up and maintenance. During the follow-up period, TLC-Detect will make monthly calls to participants: 1) to inquire whether they have adhered to the recommended referral advice and maintained treatment, 2) to offer the intervention module again if a participant has not adhered to the advice, 3) to explore whether participants have experienced symptom improvement or worsening, or no change, and finally, 4) to provide additional referrals and/or information not provided before. Maintenance calls are shortened versions of the follow-up calls and occur every two months. As in the follow-up calls, during maintenance conversations TLC-Detect also assesses users' symptom severity (better, same, worse).

Based on the design of the follow-up and maintenance modules, users with certain disorders will be followed up more closely than others. In these cases, TLC-Detect asks additional questions specific to the disorder in order to provide targeted intervention if necessary. Here are three examples: 1) patients with substance abuse problems require regular and thorough follow-up evaluations [52]. During their follow-up calls, the system is designed to explore possible barriers to quitting and to offer solutions to address each barrier. 2) Similarly, because mood instability is a common characteristic of individuals with bipolar disorder, they will have their manic symptoms re-assessed during monthly TLC-Detect follow-up calls. This will ensure that an emerging manic or depressive mood swing that occurs during the follow-up period will be detected and to alert the patient and the system's psychiatrist that a manic episode may be occurring. 3) Those individuals who have reported to TLC-Detect that they have experienced a single panic attack will be followed-up with regular calls to make sure that they have not experienced additional attacks. These individuals will be placed in the "panic disorder" category and will be provided with the appropriate intervention. 4) For users who disclose that they have been victims of violence, the system and the study's mental health staff will address the risk of the subject remaining in a dangerous environment.

The follow-up module components

An important component of the follow-up module is TLC-Detect's efforts to obtain external clinical corroboration of the nature of the user's probable mental disorder. TLC-Detect will ask whether the user has sought help from a mental health care professional to learn if the mental health professional agreed with the system's assessment of the disorder. If a user indicates that s/he has consulted with a mental health care professional and that the diagnosis made by that clinician does not concur with TLC-Detect's assessment, then the system will do one of two things: 1) In those instances where the disorder diagnosed by a user's mental health clinician is among those including in the TLC-Detect's disorder repertoire, then the system will call the user during the following week to offer options to provide the corresponding intervention module. As described earlier, TLC-Detect will spread out the follow-up calls throughout the month if users are diagnosed with more than one disorder to reduce the burden of long or complex calls. 2) If, however, lack of concordance between TLC-Detect and the patient's clinician suggests the diagnostic inferences made by the system are incorrect, or if TLC-Detect is not designed to provide information on the clinician-assessed diagnosis, then those users will receive the intervention provided for the "Unspecified Emotional Distress" described earlier.

A second important component of the Follow-up Module is an assessment of the patients' motivation to change their behavior. This is accomplished through a determination of the users' stage of readiness to seek help. This assessment is informed by the Transtheoretical Model of Behavior Change [53] in which a person's readiness to engage in a positive health behavior is categorized into one of five stages (Precontemplation, Contemplation, Preparation, Action and Maintenance).

Readiness for treatment component

The "Readiness for Treatment" component is comprised of a number of questions that are asked once the intervention module has been offered and is intended to determine whether users are ready to seek treatment. Users will be asked whether they anticipate "getting into treatment" for the identified problem(s) and if so when. If they answer negatively or report they have no plans to seek treatment within the next 2 months, they will be assigned to the Precontemplation stage. Those who answer that they are planning to get help within a month will be categorized as in Preparation; otherwise users will be assigned to the Contemplation stage. Because of the enrollment criteria, none of the users will be in the Action or Maintenance stages. During each follow-up call, TLC-Detect will re-administer the "engaging in treatment" staging questions and then will provide a stage-appropriate response. All participants who progress to the action and maintenance stages of behavioral change will be offered the materials in the maintenance module.

The first set of follow-up calls will assess whether the user has followed TLC-Detect's recommendations. Subsequently, the system will provide encouragement, support and reinforcement to those who report that they have sought care and are remaining in treatment. For those who have not sought care or have reported that they do not plan to get help within the next two months, a list of common potential barriers to seeking and remaining in treatment will be explored (e.g., transportation problems, residing too far from the source of care, family problems or other emergencies, being too busy, forgetting, therapeutic pessimism, previous history of a bad experience with a mental health care provider, too depressed to take action, etc.). TLC-Detect will then provide advice and solutions corresponding to each barrier reported.

During follow-up TLC-Detect conversations, those users who have not adhered to referral recommendations will hear the referral information provided again, along with additional education to promote adherence. All users may also hear new referrals and information never heard before, added to the system during regular updates of referral resources. All users will be provided with book titles they can read about their disorder(s). Users not adhering to at least one referral recommendation will be delivered the "Barriers" and "Readiness to Change" educational components during a TLC-Detect follow-up call. This procedure will continue throughout the first six months of system utilization.

The last six months of the project will be devoted to the "Maintenance" calls targeting those who have sought help. In contrast, those who never seek help will continue to remain in the Follow-up Module

The maintenance module

The Maintenance Module is designed to determine whether the user is still engaged in treatment. When users respond that they are no longer in treatment, TLC-Detect will assess the reason for treatment discontinuation. During the enquiry, four response options are offered: 1) difficulty carrying out the recommendations, 2) symptom improvement, 3) lack of interest in the recommendation, and 4) none of the above. Those who respond "none of the above," will be directed to a free speech component where they can record an open-ended message and describe the problem. It should be added that all participants will be invited to leave a message in this free speech section.

During Maintenance calls, all users will be offered an opportunity to listen again to the information and referrals from the Intervention Module corresponding to their specific disorder. In the very last interaction, TLC-Detect will inquire whether users have experienced any new symptoms that the system has not yet explored or whether users have been newly diagnosed with a disorder not yet detected by the system. This feature will inform the system designers about disorders that may be prevalent but were not included in the initial design of TLC-Detect.

Intervals between TLC-detect contacts

The first TLC-Detect follow-up call will occur one month after the initial (screening and intervention) contact call. Thereafter, the employee will receive calls every month for six months, and every other month thereafter (during the final six months of use). Participants will receive brief reminder calls in the middle of each month to reinforce the information they have received during the last TLC conversation. The mid-month calls will be optional and users will have the option not to listen to the information provided.

Conclusion

Undiagnosed or under-treated mental health disorders have negative impact on workers' productivity. Mental health screening of workers is challenging due to the stigma attached to mental disorders. Many employers who recognize the significant impact of mental health disorders on productivity have made attempts to assist workers by deploying Employee Assistance Programs and other approaches designed to address a wide-range of employees' mental and emotional health problems. Most of the time, however, employees are reluctant to use these programs for fear of loss of confidentiality, stigma and most importantly, fear if loss of employment. As the number of individuals with undiagnosed

mental disorders is substantial, early detection and intervention is likely to help reduce subsequent complications for both patients, employers and society. Automated systems may provide the functionality that facilitates mental health screening and intervention in the workplace where employees may be otherwise hesitant to seek or receive help. The development and evaluation of an assessment and treatment referral tool, TLC-Detect, is important because the system addresses a major public health problem; under detection and under-treatment of mental health disorder in the population.

Investigators have explored the possibility that people may be more willing to divulge to a computer personal and intimate details of their lives, with less fear of negative consequences [21]. In fact, research in human-computer interaction has shown that some individuals may be more truthful during disclosure of potentially embarrassing or private personal information to a computer in comparison to disclosures to a human interviewer [54–56]. Evidence that computer-administered interviews encourage self-disclosure has led to the development of important applications, such as computer-based interviews to detect risky health behaviors [16, 56–60]. An automated system such as TLC-Detect that assesses employees using the telephone and provides education, treatment referrals and self-care advice is an efficient, low cost and confidential tool that can improve mental and emotional distress among employees. Ultimately, successful automated behavioral telehealth systems offer the promise of enhanced productivity in the workplace.

Acknowledgments This study was funded by the Centers of Disease Control and Prevention.

References

1. Harnois, G., and Gabriel, P., *Mental health at work: impact, issues and good practices. World Health Organization. International Labour Organization*: Geneva, Switzerland, 2000.
2. Regier, D. A., Darrow, W. E., Rae, D. S., Manderscheid, R. W., and Locke, B. Z., The de facto US mental and addictive disorders service system: Epidemiologic Catchment Area prospective 1-year prevalence rates of disorders and services. *Arch. Gen. Psychiatry*. 50:85–94, 1993.
3. United States Department of Health & Human Services, *Mental Health: A Report of the Surgeon General*. Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health: Rockville, MD. 1999.
4. Kessler, R. C., Berglund, P. A., and Foster, C. L., The social consequences of psychiatric disorders. *Am. J. Psychiatry*. 154:1405-1411, 1997.
5. Kessler, R. C., Mickelson, K. D., Barber, C., and Wang, P., *The association between chronic medical conditions and work impairment, in Caring and Doing for Others: Social Responsibility in the Domains of Family, Work, and Community*, R. AS, Ed. Chicago. Ill: University of Chicago Press, 2001.
6. Shabetai, R., Depression and heart failure. *Psychosom. Med.* 64:13–14, 2002.
7. Kessler, R. C., McGonagale, K., and Zhao S., Lifetime and 12-month prevalence of DSM-III-R Psychiatric disorders in the United States, Results from the National Comorbidity Survey. *Arch. Gen. Psychiatry*. 51:8–19, 1994.
8. Kessler R., and Frank R. G., The impact of psychiatric disorders on work loss days. *Psycholo. Med.* 27:861–873, 1997.
9. Robinson G., Chimento L., Bush S., and Papy J., *Comprehensive mental health insurance benefits: case studies*. 2001, Center for Mental Health Services, Substance Abuse and Mental Health Services Administrations. National Advisory Mental Health Council, Health Care Reform for Americans with Severe Mental Illness: report of the National Advisory Mental Health Council: Rockville, MD.
10. Broadhead WE, Blazer DG, George LK, and Tse CK, Depression, disability days, and days lost from work in a prospective epidemiologic survey. *J. Am. Med. Assoc*. 264:2524–2528, 1990
11. S, Shapiro. Utilization of health and mental health services. *Arch. Gen. Psychiatry*. 41:971–979, 1984
12. Mercer, William. *Mercer/Foster Higgins National Survey of Employer-Sponsored Health Plans*. Chicago, IL, 2001.
13. Kahn, J. P., and Langlieb, A. M., *Mental Health and Productivity in the Workplace*. Jossey-Bass, San Francisco, CA, 2003.
14. Friedman, R., Stollerman, J. E., Mahoney, D. M., and Rozenblyum, L., The virtual visit: using telecommunications technology to take care of patients. *J. Am. Med. Inform. Assoc*. 4:413–425, 1997.
15. Friedman R., The use of computer telephony to provide interactive health information. *Am. Med. Inform. Assoc. Proceedings of the AMIA*. 2003: 539-543.
16. Rubin, A., Marks, L., Migneault, J., Goldstein, E., Ludena, K., and Friedman, R., Automated telephone screening for problem drinking. *Ann. Beha. Med.* 25 Supplement, 2003.
17. Mooney, K., Beck, S., Friedman, R., and Farzanfar, R., A pilot study of telephone-linked care (TLC) for cancer symptom monitoring. *Cancer Pract*. 1:147–157, 2001.
18. Farzanfar, R., Frishkopf, S., Friedman, R., and Ludena, K., Evaluating an automated mental health care system: making meaning of human computer interaction. *Computers in Human Behavior.* 2006: In press.
19. Delichatsios, H., Glanz, K., Tennstedt, S., Smigelski, S., Pinto, B. M., Kelley, H., and Gillman, M. W., Randomized trial of a "talking computer" to improve adults' eating habits. *Am. J. Health Promot.* 15:215–224, 2001.
20. Pinto, B., Marcus, B. H., Kelley, H., Tennstedt, S., and Gillman, M. W., Effects of a computer-based, telephone-counseling system on physical activity. *Am. J. Prev. Med.* 23: 113–120, 2002.
21. Ramelson, H. Z., Friedman, R. H. F., and Ockene, J. K., An automated telephone-based smoking cessation education and counseling system. *Patient Educ. Couns*. 36:131–143, 1999.
22. Friedman, R., Kazis, L. E., Jette, A,. Smith, M. B., Stollerman, J., Torgerson, J., and Carey, K. B., A telecommunication system for monitoring and counseling patients with hypertension: impact on medication adherence and blood pressure control. *Am. J. Hypertens*. 9:285–292, 1996.
23. Farzanfar, R., Locke, S. E., Vachon, L., Charbonneau, A., and Friedman, R., Computer telephony to improve adherence to antidepressants and clinical visits. *Ann. Behav. Med*. Annual Meeting Supplement. 25: S161, 2003.
24. Young, M., Sparrow, D., Gottlieb, D., Selim, A., and Friedman, R. H., A Telephone-Linked Computer System for COPD Care. *Chest.* 119: 1565–1575, 2001.
25. Prochaska, J., Velicer, W. F., DiClemente, C. C., and Fava, J., Measuring processes of change: applications to the cessation of smoking. *J. Consul. Clin. Psychol*. 56:520–528, 1988.

26. Velicer, W. F., DiClemente, C. C., Prochaska, J. O., and Brandenburg, N., Decisional balance measure for assessing and predicting smoking status. *J. Pers. Soc. Psychol.* 48:1279–1289, 1985.
27. Bandura, A., *Social Cognitive Theory*. Englewood Cliff, NJ:1977.
28. Spitzer, R. L., Kroenke, K., and Williams, J. B., Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. *Patient Health Questionnaire. JAMA*. 282:1737–1744, 1999.
29. Williams, L., Brizendine, E. J., Plue, L., Bakas, T., Wanzhu, T., Hendrie, H., and Kroenke, K, Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke*. 36:635–638, 2005.
30. Kroenke, K., Spitzer, R. L., and Williams, J. B., The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom. Med*. 2:258–266, 2002.
31. Connor, K. M., Davidson, J. R. T., Churchill, L. E., Sherwood, A., Foz, E., and Weisler, R. H., Psychometric properties of the Social Phobia Inventory (SPIN). *Br. J. Psychiatry*. 176:379–386, 2000.
32. Horowitz, M., Wilner, N., and Alvarez, W., Impact of Event Scale: A measure of subjective stress. *Psychosom. Med*. 41:209–218, 1979.
33. Sundin, E. C., and Horowitz, M., Impact of Event Scale: Psychometric properties. *Br. J. Psychiatry*. 180:205–209, 2002.
34. Joseph, S., Psychometric evaluation of Horowitz's impact of event scale: a review. *J. Trauma. Stress*.13:101–113, 2000.
35. Bryant, R. A., Moulds, M. L., and Guthrie, R. M., Acute Stress Disorder Scale: a self-report measure of acute stress disorder. *Psychol. Assess*. 12:61–67, 2000.
36. Morrison, A. P., Frame, L., and Larkin, W., Relationships between trauma and psychosis: a review and integration. *Br. J. Clin. Psychol*. 42:331–353, 2003.
37. Breslau, N., Chilcoat, H. D., Kessler, R. C., and Davis, G. C., Previous exposure to trauma and PTSD effects of subsequent trauma: Results from the Detroit Area Survey of Trauma. *Am. J. Psychiatry*. 156:902–907, 1999.
38. Das, A. K., Olfson, M., Gameroff, M. J., Pilowsky, D. J., Blanco, C., Feder, A., Gross, R., Neria, Y., Lantigua, R., Shea, S., and Weissman, M. M., Screening for bipolar disorder in a primary care practice. *JAMA*. 293:956–963, 2005.
39. Robert, M. A., Hirschfeld, J., Williams, B. W., Spitzer, R., Calabrese, J. R., Flynn, L., Keck, P. E., Jr., Lewis, L., McElroy, S. L., Post, R. M., Rapport, D. J., Russell, J. M., Sachs, G. S., and Zajecka, J., Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *Am. J. Psychiatry*. 157:1873–1875, 2000.
40. Miller, C. J., Klugman, J., Berv, D. A., Rosenquist, K. J., and Nassir Ghaemi, S., Sensitivity and specificity of the Mood Disorder Questionnaire for detecting bipolar disorder. *J. Affect. Disord*. 81:167–171, 2004.
41. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR. Fourth Edition, American Psychiatric Association Washington, DC. 2000.
42. Saitz R. Unhealthy Alcohol Use. *N. Engl. J. Med*. 352:596–607, 2005.
43. Canagasaby, A., and Vinson, D. C., Screening for hazardous or harmful drinking using one or two quantity-frequency questions. *Alcohol Alcoholism* 40: 208–213, 2005:
44. Feinn, R., Tennen, H., and Kranzler, H. R., Psychometric properties of the short index of problems as a measure of recent alcohol-related problems. *Alcohol. Clin. Exp. Res*. 27:1436–1441, 2003.
45. Kessler, R. C., Mroczek, D., Ustun, B., and Wittchen, H. U., The World Health Organization Composite International Diagnostic Itnerview Short-Form (CIDI-SF). *Int. J. Methods. Psychiatr. Res*. 7:171–185, 2000.
46. Anderson, G., Carlbring, P., Kaldo, V., and Strom, L., Screening of psychiatric disorders via the internet. A pilot study with tinnitus patients. *Nord. J. Psychiatry*. 58:287–291, 2004.
47. Kreek, M. J., Multiple drug use patterns: recent trends and associated medical consequences. In: Mello, N. H., (Ed.,) *Advances in Substance Abuse, Behavioral and Biological Research*. Vol. IV. London: Jessica Kingsley, 1991.
48. Collins, R. L., Ellickson, P. L., and Bell, R. M., Simultaneous polydrug use among teens: prevalence and predictors. *J. Subst. Abuse*. 10:233–253, 1998.
49. Cohen, S, Kamarck, T, and Mermelstein, R. A Global Measure of Perceived Stress. *J. Health. Soc. Behav.* 24:385–396, 1983.
50. Cohen, S, and Williamson, G. M., Perceived stress in a probability sample of the United States. In Spacapan S., and Oskamp S. (Eds.), *The Social Psychology Of Health*. Newbury Park, CA: Sage, 1988.
51. Abrams, Helen S, Harvard community health plan's mental health redesign project: A managerial and clinical partnership. *Psychiatric Quarterly*, 64:13–31, 1993.
52. McKay J. R., Lynch KG, Shepard DS, Morgenstern J, Forman RF, and Pettinati H. M. Do patient characeristics and initial progress in treatment moderate the effectiveness of telephone-based continuing care for substance use disorders?. *Addiction*, 100(2):216–226, 2005 Feb.
53. Rosenfeld, P, Giacalone, R. A., Knouse, S. B, Doherty, L. M., Vicino, S. M., Kantor, J., and Greaves, J., Impression Management, Candor, and Microcomputer-based organizational surveys: an individual differences approach. *Comput. Hum. Behav*. 7:23–32, 1991.
54. Turner, C., Ku, L., Rogers, S. M., Lindberg, L. D., and Pleck, J. H., Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science*. 280:867–873,1998.
55. Gerbert, B, Bronstone, A., Pantilat, S., McPhee, S., Allerton, M., and Moe, J., When asked, patients tell: disclosure of sensitive health-risk behaviors. *Med. Care*. 37:104–11, 1999.
56. Locke S. E, Kowaloff H. B, Hoff R. G, Safran C, Popovsky M. A, Cotton D. J, Finkelstein D. M, Page P. L, and Slack W. V. Computer-based interview for screeningblood donors for risk of HIV transmission. *J. Am. Med. Assoc*., 268:1301–1305, 1992.
57. Erdman, H., Klein, M. H., and Greist, J. H., The reliability of a computer Interview for drug use/abuse information. *Behav. Res. Methods. Instrum*. 15:66–68, 1983.
58. Lucas, R., Mullins, P. J., Luna, C. B., and McInroy, D. C., Psychiatrists and a computer as interrogators of patients with alcohol-related illnesses: a comparison. *Br. J. Psychiatry*. 131:160–67, 1977.
59. Baer, L., and Greist, J. H., An interactive computer-administered self-assessment and self-help program for behavior therapy. *J. Clinical Psychiatry*. 58:23–28, 1997.
60. Kobak, K., Greist, J. H., Jefferson, J. W., Mundt, J. C., and Katzelnick, D. J., Computerized assessment of depression and anxiety over the telephone using interactive voice response. *MD Computing*. 16:64–68:1999.

Available online at www.sciencedirect.com
ScienceDirect
Computers in Human Behavior
ELSEVIER
Computers in Human Behavior 23 (2007) 1167–1182
www.elsevier.com/locate/comphumbeh

Evaluating an automated mental health care system: making meaning of human–computer interaction

Ramesh Farzanfar *, Sophie Frishkopf, Robert Friedman, Kevin Ludena

Medical Information Systems Unit, Boston University Medical Center, 720 Harrison Avenue, Suite 1102, Boston, MA 02118, USA

Available online 8 December 2004

Abstract

Objectives: To qualitatively evaluate the response of patients with unipolar depression who used a computer telephony system designed to monitor their disease severity and support self-care, principally adherence to medication regimen and clinical office visit attendance.
Methods: Weekly in-depth interviews were conducted with 15 patients who used the computer telephony system for 4 weeks. Users had a diagnosis of unipolar depression and took at least one antidepressant. All interviews were audio-taped and immediately transcribed. The transcripts of the interviews were subsequently coded and analyzed thematically by two qualitative researchers.
Results: The patients spoke about the automated system as if it was a social actor. They did not, however, have an illusion that there was a health professional communicating through the system. Instead, they felt that it was designed to appear human-like. The majority offered suggestions intended to make the system behave and sound *more* like a "human professional" and *less* like a "machine". They believed that the system would be more usable, acceptable and effective if these changes were made.

* Corresponding author. Tel.: +1 617 638 7519; fax: +1 617 638 8858.
E-mail address: rfarzanf@bu.edu (R. Farzanfar).

doi:10.1016/j.chb.2004.11.015

Conclusions: These results do not support the "anthropomorphism" construct which posits that users of computer-mediated systems who attribute human qualities to the system are under an illusion that the system is human.

Keywords: Automated patient management; Qualitative evaluation; Anthropomorphism

1. Introduction

During the past few decades computer-based automated systems have been integrated in many facets of peoples' lives. In health care, computer-based systems designed to provide health information and advice to patients and consumers are on the rise (Slack, 1997). Some of these systems appear similar to a human health professional: interactive and designed to emulate the style of a human counselor (Friedman, 1998). They ask questions, provide tailored and appropriate responses to users, offer practical advice and educate by providing useful and up-to-date health information. Some computer-based systems are deliberately designed to display human qualities, such as empathy, humor and objectivity (Friedman, Stollerman, Mahoney, & Rozenblyum, 1997). The goal is to have users respond as they would to a human health professional.

Studies of users of automated systems, in health and other spheres of human life note that many users actually respond to them in anthropomorphic terms: (Friedman, 1995; Kahn, Friedman, & Hagman, 2002) users assign human characteristics such as emotions, feelings and other qualities to the inanimate computer system (Fogg & Nass, 1997; Moon, 1998; Card & Lucas, 1981; Aroyo, 2000; Koda & Maes, 1995). The present paper presents the results of a formative evaluation, conducted to explore the attitudes of users of a computer-based telephony system designed to function and appear like a human health professional who cares for patients with depression (TLC-Depression). We sought to evaluate whether, patient users established anthropomorphic relationships with the system, and, if so, the meanings of these relationships to them.

2. Methods

2.1. The study setting

The TLC-Depression is designed to be used by ambulatory patients with unipolar depression. It includes modules for monitoring the patients' disease severity and for promoting anti-depressant medication-taking and visit adherence. The study was conducted in our headquarters (Medical Information Systems Unit) at Boston Medical Center (BMC). Many of the participants received their mental health care at the BMC's Psychiatry Clinic.

2.2. A description of the automated health system

2.2.1. The basic system: telephone-linked communications

The automated telephony system we evaluated, used a computer-based telecommunications system technology that has been developed to monitor patients with chronic diseases, help them care for themselves (Friedman et al., 1997; Friedman et al., 1996; Friedman et al., 1998) and assist patients and consumers make important changes in health behavior (Friedman, 1998; Ramelson, Friedman, & Ockene, 1999; Jarvis, Friedman, Heeren, & Cullinane, 1997; Delichatsios et al., 2001). The system called Telephone-Linked Communications (TLC) can be used as a stand-alone service by patients/consumers or it can supplement the care provided by clinicians. During automated telephone conversations, TLC asks users questions to monitor the targeted disease(s) and health behaviors, provides information, education, advice and behavioral counseling to modify or sustain behavior, and reports important findings to responsible caregivers. Patients communicate by either pressing the keys on the telephone keypad or by speaking directly into the telephone receiver.

Either the patient or the system can initiate a conversation. Should the patient fail to call when expected, TLC will call the patient. Upon telephoning, TLC is described. Thereafter, patients are asked to phone in regularly (weekly in most TLC applications). At the beginning of a TLC conversation, the patient is asked to enter a password to ensure security. The system then begins the conversation with a salutation, follows with questions and provides responses to the patient's answers. In addition to questioning, TLC provides education and behavioral reinforcement, such as counseling on how to take medications at prescribed times. TLC conversations were designed to emulate telephone conversations between patients and health professionals, drawing on specific behavioral theories. A typical conversation lasts between 10 and 15 min, depending on the number and complexity of the topics addressed and the user's responses. After each conversation, TLC stores in a database information the user has communicated. TLC relays important information obtained from the conversations to the patient's health care provider via electronic transmission or fax. Reports of an emergency nature are sent immediately.

2.3. The Telephone Linked Communications for Depression

Telephone Linked Communications for Depression (TLC-Depression) is a multi-contact system designed for weekly use over an extended period of time. The system was programmed to use SpeechWorks,™ which uses a continuous speech (''speaker-independent'') voice recognition system which interprets what a person says. The system does not require an individual user to ''train'' it in order to be understood. TLC-Depression is comprised of two modules: (1) Disease monitoring and (2) Self-care. Different modules of the TLC-Depression scripts were recorded by a male voice professional. In the following sections, we describe these modules.

2.3.1. The TLC-Depression disease monitoring module

The objective of the TLC-Depression monitoring module is to provide between office visits timely clinical information to mental health care providers regarding the status of their patients. The information collected is designed to aid the provider in disease management and is obtained by TLC-Depression during regular (weekly) assessment of symptom severity. The content of the module is based on a standardized depression screening questionnaire: the Patient Health Questionnaire (PHQ-9) which is part of the PRIME-MD (Primary Care Evaluation of Mental Disorders) (Kroenke, Spitzer, & Williams, 2001).

The PHQ-9 contains nine symptom questions and provides a quick and simple method to diagnose depression, track depression severity, and monitor a patient's response to therapy. Each TLC-Depression call asks all PHQ-9 questions. One result of the assessment is the determination of suicidality and, if it is present, whether there is suicidal ideation (thoughts only), suicidal planning (thought about how to accomplish it) and/or suicidal intent (intention to carry out a plan). Based on what the patient reports, the TLC-Depression disease monitoring module generates one of three types of reports to the responsible mental health care provider: Important (reporting suicidal ideation), Urgent (reporting suicidal planning) and Emergent (reporting suicidal intent). Important and urgent alerts are sent immediately to the provider's staff who receive these alerts and place them in the mailbox (physical or electronic) of the responsible clinician. Emergency alerts are not only sent immediately, but TLC-Depression automatically and immediately pages the Director of the Psychiatric Emergency Department and telephones her private line to leave a message on voice mail. All of the alerts contain the name, address and telephone number of the patient, the nature of the alert (suicidal ideation, planning or intent), and the clinician's name.

2.3.2. The TLC-Depression self-care module

The TLC self-care module is comprised of different segments that assess and promote: (A) anti-depressant medication-taking; (B) office visit attendance; (C) three additional aspects of self-care: regular exercise, sleep hygiene, and responsible alcohol use.

2.3.2.1. Anti-depressant medication taking. The purpose of this self-care management component is to monitor the patients' use of prescribed anti-depressants and promote complete adherence to the regimen. The system collects medication data from automated monitors on the antidepressant medication bottles on a monthly basis while patients communicate self-reported adherence weekly. The main objective of collecting both self-report and MEMS data was to enhance the rigor of the data. Collecting only self-reported adherence data is considered "systematically" unsound because adherence is not objectively measured and the data may not be accurate (Haynes, Taylor, & Sackett, 1979).

We use special bottle caps (Medication Event Monitoring Systems [MEMS™], Aardex Corp.) to detect "dose events" and from these, calculate medication regimen adherence. Previous work has shown that assessment of adherence by

MEMS caps is both reliable and valid; the method is the gold standard for determining patient medication-taking adherence (Thompson, Peveler, Stephenson, & McKendrick, 2000; Cramer & Rosenheck, 1999; Engstrom, 1991; Cramer, Mattson, Prevey, Scheyer, & Quellette, 1989). MEMS has two parts: a standard plastic vial with threaded opening and a cap for the vial that contains a micro-electronic circuit that registers times when the bottle is opened or closed. Each closing of the cap constitutes a dose event. Medication dose events stored in the MEMS caps are transferred to a secure Web server through a MEMS communicator. The communicator is a device that reads and writes the user's medication adherence data over a "serial" cable connected to a computer's serial port. Each time a MEMS cap is "read" or "written" on, it is placed on the communicator and the data on the cap are transferred. Special software at the Web server analyzes the MEMS data, calculates medication-taking parameters (total percentage of doses taken during a time period, pattern of medication-taking [e.g., some patients may fail to take medication on weekends] and variability of medication-taking time [e.g., some patients may take medication erratically instead of at regular intervals), and stores them in a database at the Web site where they are electronically "retrieved" by the TLC system. Retrieval of data was made possible through the development of software that functions as a bridge that queries into the MEMS data repository. The computed results of the queries are then forwarded to the TLC data repository. Data transfer occurs within minutes of its download from the MEMS communicator.

2.3.2.1.1. The automated conversation. The conversation in the medication adherence module contains the following topics: (1) medication regimen knowledge; (2) medication-taking self-report; (3) MEMS data and feedback; (4) motivation assessment; (5) confidence assessment; (6) identification of barriers; (7) myths and misconceptions about anti-depressants.

(1) The medication regimen knowledge segment begins by TLC-Depression asking the patient to collect his/her antidepressant bottle(s). Next, the system ascertains whether the patient is taking his/her prescribed anti-depressant referring both to generic and common trade name(s). If the correct medication is confirmed, TLC-Depression determines whether the patient's understanding of the unit dose is the same as that prescribed. This question is followed by one that asks about the number of units (tablets, capsules, etc.) the patient takes at one-time (per dose) and the number of times this dosing is repeated daily. In instances when there is a discrepancy between what TLC-Depression has recorded as the prescription and the patient's report, the patient's current prescription is reviewed with the patient, his/her pharmacy and/or physician's office. The objective is to determine whether the discrepancy is due to the patient's lack of understanding of the regimen or to a physician-prescribed change in the regimen that was not communicated to TLC-Depression. Subsequently, discrepancies are corrected and the patient is notified about any misunderstandings he/she might have.

(2) Medication-taking self-report: TLC-Depression asks patients to get their TLC medication calendar. The calendar is given to the patients during their baseline visit to record the doses of their anti-depressants taken daily, facilitating reporting of them to TLC for the previous week (last seven days). TLC-Depression then asks about prescription refills, pocketed doses (all pills that have been taken out of the bottle and are either pocketed or left in a pill tray, etc.) and missed doses, if any. Once the medication-taking regimen has been established, the system compares the proportion of prescribed doses taken (medication adherence) for the previous week with the adherence percentage calculated during previous calls for previous weekly time periods. Based on the comparison, TLC-Depression gives the patient feedback on whether his/her level of adherence has changed and, if so whether it is better or worse and by what degree.

(3) MEMS data and feedback: TLC-Depression uses MEMS caps adherence data downloaded on a monthly basis to give feedback to patients regarding adherence level (proportion of prescribed doses taken), drug holidays and variability in the time of day in medication-taking. Data on medication-taking patterns and time variability are only collected by MEMS caps and not through self report. In the beginning of this section, however, TLC-Depression communicates that the MEMS caps are not perfect and sometimes may not be accurate. For example, it begins by saying "the bottle cap will give most people an accurate report of when they have taken or missed their medication. In some cases, however, it can be less accurate. You told me that you sometimes take pills out of your bottle to take them later (pocketing doses). So, the information gathered by the bottle cap may not be the whole picture of when you take your medication, but it should be pretty close."

(4) Assessment of the patients' motivation to take his/her anti-depressant: First, TLC-Depression asks them to rate the importance of taking the medication from 0 to 10. Next, it responds to the patient's entry and follows with comparison to the rating during the previous week. During the first 3 weeks of conversations, all highly motivated patients (ratings of 8–10) skip the rest of the conversation scripted for the section and will proceed to the next. For patients who rate their motivation for taking the medication below eight, TLC-Depression presents a list of common reasons (three at a time) why people sometimes do not want to take their anti-depressants and asks them to choose those that are relevant.

(5) Assessment of patient's confidence that she/he can follow his/her prescription: The conversations in this section use techniques from Self-Efficacy Theory (Bandura, 1977) and motivational interviewing to boost patients' confidence (Miller & Rollnick, 2002). TLC-Depression assesses the patients' confidence levels by asking them to rate their confidence in following their anti-depressant prescription regimen. If a patient's response reveals low confidence, TLC will say: "Does your clinician know that you think you cannot successfully follow your prescription? Be sure to bring this up during your next appointment, or

phone him or her to talk about it. In your TLC calendar notes write your thoughts about your lack of confidence. Share them with your clinician. It will help your clinician find ways to make it easier for you.''

(6) Identification of barriers to taking anti-depressants as prescribed: TLC-Depression identifies the barriers a patient might face that prevent full adherence to his/her antidepressant regimen. In this section patients are presented with a list of three topics at a time to choose from. Topics include ''forgetting,'' ''stress,'' ''busy,'' ''away from home,'' ''side-effects,'' ''late refill,'' ''being sick,'' among others. The focus is on practical approaches to solving medication-taking difficulties and presenting an array of strategies for various situations. For example: ''People most often explain why they do not take their medication by saying something like, 'I do not know; I just forget to take it'. Specifically, people are most likely to forget or put off taking pills when they are very busy or under a lot of stress. Say 'forget' if you would like to hear about ways to prevent yourself from forgetting your medication. For handling busy times that interfere with medication, say 'busy'. For information on dealing with stressful times, say 'stress'. If none of them interest you today, say 'None'''.

(7) Examining myths and misconceptions about antidepressants: This section deals with 12 common misconceptions, misunderstandings and mistaken beliefs that patients may have about their anti-depressant or about depression in general. Every two weeks TLC-Depression presents the patients with two myths and ascertains whether they believe in either. The list includes statements such as ''it is OK to stop taking antidepressant while you are sick with flu or cold,'' or ''depression is just a normal part of the aging process.''

2.3.2.2. Office visit adherence promotion. This section is comprised of two parts: (1) assessment of psychiatry clinic attendance and counseling to improve attendance, if required; (2) education to encourage patients to get the most benefit out of their visits. Assessment is conducted by asking questions about the reasons for missed visits followed by advice on how to remedy the problem. Information and advice include such instructions as writing down thoughts or questions before a visit, not saving the biggest concerns for the end of the visit, etc.

2.3.2.3. Additional aspects of self-care. This section was designed to provide patients with information and education on three critical areas of self-help for depression: (1) promoting physical activity for those patients whose amount of physical activity may not be sufficient based on standards set by Centers for Disease Control and Prevention (CDC) and the American College of Sport Medicine (ACSM); (2) sleep hygiene to help patients whose sleep has been affected by depression; (3) detection of alcohol abuse.

The physical activity promotion component informs and educates patients on the benefits of getting regular exercise by offering support and encouragement for patients to begin an exercise program, or increase the amount of physical activity they engage in. The second area concerns sleep. TLC-Depression asks questions to determine to what extent the patient's depression has affected sleep.

The system then, provides advice and information to the patients for the avoidance of situations that may contribute to sleeplessness. The alcohol segment contains an automated version of the AUDIT (Alcohol Use Disorder Identification Test) which is a standardized, valid and reliable instrument for screening drinking behavior (Babor, de la Fuente, Saunders, & Grant, 1992). The TLC-Depression system activates the AUDIT questions if a patient chooses to learn whether s/he has a drinking problem. For those who score high for problem drinking, TLC-Depression provides advice and offers the telephone number for the Massachusetts alcohol hotline which is a 24-hour hotline offering around the clock access to counseling, emergency support, and information and referral services. TLC-Depression also urges the patient to talk to his/her doctor or a health professional regarding his/her alcohol consumption.

2.4. Design of the formative evaluation study

2.4.1. The target population

Our principal objective was to understand the experience of users of TLC-Depression and identify deficiencies in design that could be corrected. We conducted the qualitative evaluation over two months with 15 patients (9 women, 6 men; 2 Hispanics, 5 Blacks and 8 Whites; the age range was 20–60) diagnosed with depression and on one or more antidepressants. Recruitment was carried out by posting fliers at Boston Medical Center's (BMC) Psychiatry Clinic and other hospital clinics and advertising in a Boston daily newspaper. The fliers and advertisement addressed patients diagnosed with depression and invited them to use and evaluate an automated telephone system "designed to help patients with unipolar depression with adherence to their antidepressant medication regimen." We informed potential participants that they would receive a $100 stipend for participation. We listed a telephone contact number for interested individuals to call and leave a message.

One hundred and thirty individuals left messages. Subsequently, we called potential participants to explain the study, assess eligibility, and obtain verbal informed consent for those who were eligible. Eligibility criteria were: (1) $\geqslant$18 years old; (2) diagnosis of unipolar depression; (3) taking at least one antidepressant; (4) fluency in spoken English. All those who were eligible and gave verbal informed consent were invited for an initial study visit at our research headquarters on a first-called, first-selected basis. All participants signed an informed consent document and a HIPAA authorization form during their study visit. The visit took place within a week of the telephone interview.

Participants were given a system password, shown how to use TLC-Depression, and provided with a user's guide that contained important information on how to use the system, answers to "frequently asked questions," and a "calendar" to enter the time of day when took medication and the date and times of their Psychiatry Clinic appointments. Following the initial visit, subjects had 3 weekly study visits for in-depth interviews about their experiences using TLC-Depression, likes and dislikes, and suggestions about the system and how it could be improved. They had complete freedom to express opinions. Interviews lasted between 30 and 60 minutes

All interviews were audio-taped and immediately transcribed. Thematic analysis of the transcripts began immediately after each interview.

Since the qualitative study preceded a randomized clinical trial that evaluated the impact of TLC-Depression, we pre-tested a clinical trial protocol and instruments during the qualitative study. During the initial visit the participants filled out draft study questionnaires and were given MEMS caps to place on their anti-depressant bottles to monitor/track medication-taking behavior. All participants were instructed to call TLC-Depression four times/week for one month (16 calls in total) in order to attain more contact times over a short study period than would have been the case with the weekly calls planned for the clinical trial.

Since we conducted in-depth interviews over time (a total of three interviews over 3 weeks), we were also able to track, delineate and clarify experiences over time. The first and second authors analyzed and coded the interview transcripts independently in thematic categories. Discrepancies in coding were identified, discussed and resolved.

3. Results

In evaluating TLC-Depression, the participants drew heavily from their experiences with depression, ambivalence about taking antidepressants and, simultaneously, awareness of the fact that adherence to a medication regimen was an important aspect of recovery. They shared with us their personal stories of suffering, struggle and triumph over depression. They identified important issues in the design of TLC-Depression. A number of themes and general issues emerged from analysis of the interviews.

3.1. Humanizing the machine

3.1.1. Sounding more natural

We learned that the system was evoking emotional and behavioral responses from our participants that typically are expressed in interaction with health professionals. For example, their responses to TLC-Depression feedback about medication-taking suggested that feelings and behaviors were affected by a desire to be seen as a "good patient." Despite knowledge that assessment of medication-taking habits was being carried out by a machine, most felt an obligation to improve medication taking behavior. In the words of one female participant: "When *he* [emphasis added] is disappointed because you are doing bad you feel sad." Another woman, was critical of the system because of its long and tedious informational/educational sections, but also said: "It was probably the first time in my life that I got my refill before I needed it. I am one of those people that runs out every month and won't have pills for three days. I have been on it for years, and I still do it. But, the first or second time I called, that [reminder for the patient to refill their antidepressant prescription on time] was one of the options. It is not like he told me anything I didn't know. *But*, *I guess once you've heard his little speech*, *you don't want to disappoint him* [emphasis added]."

The majority of participants referred to the TLC-Depression system as "*he*. " [As described earlier, a male professional recorded the TLC-Depression script.] One individual referred to the system as "it" during his first qualitative evaluation interview session [after using the system one or two times], but then referred to the system as "*he* " consistently thereafter during subsequent interviews. At one point during the third interview he referred to the system as "the voice of my old friend," while on another occasion he called it, "the machine-man!" One participant said that she thought the 'voice' should have a name. When we asked what the name should be, she said: "Dr. X!"

A number of participants wanted us to make TLC-Depression sound more "human-like." This opinion, however, was expressed in two different ways. Some participants wanted TLC-Depression to sound less "robotic," "artificial" and more "conversational." According to one person TLC-Depression talked too slowly as if "*he* " was trying to make sure that the user understood him and as a result sounded "like an American trying to talk to a Frenchman!" Another said that the system sounded "flat," elaborating, "I kind of want to zone out when I am listening to those, 'less than half the days, more than half the days'" [referring to the close-ended multiple choice options to questions in the symptom severity module]. Others, on the other hand, felt that TLC-Depression's delivery was "over-exuberant," "too dramatic," and "loaded with false cheer." For TLC-Depression to sound more "human" it should be less "dramatic". The phrase "that's GREAT" (offered sometimes to give praise) was one example of the over-exuberance and drama that several of our participants objected to. One person requested that the word "great" be said in a more subtle way. Another mentioned that "people who are depressed are so judgmental. When I was not taking medication, the slightest thing... like tone of voice because *they are talking to a machine* – [emphasis added]. You take everything so personal because you are not really connecting to the world."

Another way for TLC-Depression to sound more "human-like" was to become more engaging. One common suggestion was for the system to know the name of the user, and repeat the name several times during the conversation. In fact, the system was designed to call users by their first name. As we began our study, it took a few days to integrate the users' actual names into the system, thus participants were not addressed personally during their initial uses. Over a month's participation, they were exposed to two different versions of TLC-Depression: one did not address them by name and one did. After using the system for a few days, and as users began to hear their names, we started hearing positive comments. One woman said: "I like it when he says my name. I didn't like it when he didn't say my name." Several participants were pleasantly surprised when they realized that TLC remembered and referred to the content of previous conversations. To them, this not only made the system more "intelligent" and human-like, but it also facilitated greater connection. We were applauded for having such a feature in our system.

At the same time, some participants felt that TLC-Depression sounded "condescending." One man said that the voice spoke to him as if he "were a five year old." This opinion was also expressed by a woman who said "it feels condescending. [praising me by saying] Good job!' You are not a cheerleader!"

3.1.2. Sounding like a health professional

Many participants wanted to improve the TLC-Depression system so it sounded more like a "professional" human agent. For example, one woman said "most shrinks tend to be quieter and talk more slowly. They seem to be able to get to the point where they are not extreme. A noncommittal tone is what seems to need to be in there." Another participant contrasted the TLC voice with the voice of his own psychiatrist: "He is got this tone even in his casual conversation that is very reassuring."

3.1.3. Sounding more sympathetic

Most users said TLC-Depression did not express compassion and empathy adequately, which undercut its "humanness." One woman said that she didn't like the phrase, "I am sorry"; it was more of a cliché than a real expression of sympathy. She added "that is the first thing that brings them [patients] to depression – feeling sorry for themselves." Instead, she said that she preferred "I sympathize with you," adding "I am sorry is like... well... if you feel sorry, what are you gonna do for me?" There seems to be the expectation that the system would not only empathize with those suffering with depression, but provide concrete help as well, a reasonable expectation for a counseling system.

TLC-Depression's lack of sympathy was especially evident to participants when suicidality was discussed. Feedback to users who reported suicidal 'ideation,' 'planning' and 'intent' was worded very carefully. Patients were told to call their clinicians immediately and/or go to the nearest emergency room depending upon urgency of the suicidality. Some felt that this response was not sufficiently "sensitive" or "human." Instead, TLC-Depression needed to demonstrate concern and sympathy for the patient by conveying an appreciation of the anguish and distress a person experienced. A more sensitive tone; one that indicated compassion and concern was desired. Some participants addressed the *manner* in which TLC-Depression responded to patients. In the words of one, "somehow... [laughing] it was really kind of... – do not be so *cheery* about the fact I am about to jump out of a 30th floor window!" To this participant, TLC-Depression communicated a cavalier attitude in the "suicidality" which was perceived as "uncaring": "you are just kind of bullshitting me. I am not going to take you very seriously."

3.1.4. It is not easy to be a machine: thus, bring in a human being

Some participants strongly communicated that having a "machine" respond to their expressions of suicidal thoughts was insufficient or inappropriate, even if the machine sounded like a human being. They advised us that after we provided our "sympathetic" and "caring" feedback to users in response to the suicidality questions, we should direct users to a suicide hot line – an opportunity for users to talk to a *live* person – in addition to suggesting contact with the clinician or going to an emergency room. This response shows that despite making our best efforts to have TLC-Depression sound more human-like, our evaluators wisely knew that it was not possible to design a system that could fully address a suicidal person's needs.

One woman, an English teacher, offered to rewrite TLC-Depression's responses in the suicidality section. After a few weeks, she sent us an email containing the rewritten sections. The communication revealed the difficulties she had experienced trying to rewrite passages to make a "machine" sound "caring," "compassionate," and "effective" enough to convincingly address a person's suicidality state. Realizing the Herculean task she had undertaken, she wrote: "It really is not easy being a machine, to mis-quote Kermit the Frog!" [Kermit the Frog's original remark was: "It is not easy being green!"]

3.1.5. Actions taken

Participants feedback prompted a number of modifications in the content and communication style of TLC-Depression. The most important of which were the following: (1) a comprehensive revision of the script that reduced the length of the conversation to ease burden on users; (2) re-writing the script that addressed TLC-Depression's feedback to those with suicidal ideas and/or plans, to make responses more empathetic and helpful; (3) TLC-Depression's voice was changed from a man's voice to a woman's; (4) modification of words and phrases participants had criticized.

4. Discussion

The vast and sophisticated literature on computer-mediated interactions (Person et al., 2000; Bates, 1994; Nass et al.) encompasses work on "mind," (Searle, 1983) "consciousness," (Cotterill, 1998) "subjectivity," (Manovich, 1995) "intentional stance," (Dennet, 1996) and the use of automated technologies as humanized proxies (Ambrose, 2001). The literature both describes and explains the interactive process between human agents and computers, including how people attribute human qualities and even human agency to machines (Marakas, 2000; Friedman, 1997). The process can involve anthropomorphism, that is, the projection of human characteristic onto the non-human computer-mediated programs.

Reeves and Nass (1996) suggest that human users tend to attribute social characteristics to computers in a "mindless" way while knowing and acknowledging that computers are only machines. The authors believe that many assumptions that have been made about human–computer interaction, particularly about anthropomorphism, are not valid. They challenge the notion that the users are psychologically or socially dysfunctional (Turkle, 1984; Zuboff, 1988) or need to project human qualities onto a computer because humans program the computers and thus there is a strong human element in the program (Dennett, 1998).

In presenting their empirical explorations of interaction between people and computers, Nass, Steuer, Henriksen, and Dryer (1994) use the term, "ethopoeia" to describe their interpretation of the phenomenon. Ethopoeia, they write, "involves a direct response to an entity as human while knowing that the entity does not warrant human treatment or attribution (Nass et al., 1994). The behavior occurs due to a response to "contextual" cues that direct attention to certain features in a phenom-

enon while diminishing attention to others. In their view, we attribute human traits to computers when computers provide cues that all social actors use to communicate. The features used by computers to mimic the human characteristics include use of spoken language and human voice, interactivity, use of human forms and gestures, in screen-based systems, and generally "filling of roles traditionally filled by humans." (Nass & Moon, 2000).

Our evaluation results corroborate Nass's interpretation of human–computer interaction. Our users utilized *metaphors* to describe and give meaning to their experiences using TLC-Depression. Metaphors helped conceptualize perceptions by comparing them with more familiar experiences such as interacting with mental health professionals. The designers of TLC-Depression understood this potential and designed the system to maximally exploit it. The design thus included all the features that, according to Nass and Moon (2000) enhance attribution of social agency. For example, TLC-Depression uses a "human voice" to "converse" with the user. The voice takes on the style and uses language expected of a therapist. The system is "interactive." It "refers to itself" as "I" while "referring to the user" by his/her name. The system "knows" personal and medical information about the user such as his/her gender, age, past history of depression, what medications s/he takes, the user's clinician, etc. The system "tracks" the patient's condition and self-care over time; it has memory. TLC-Depression provides information, advice and counseling. In some ways, it takes on attitudes of a "helping professional." It is designed to respond emotionally, both in tone and content. It expresses empathy, delight and disappointment. The system is designed to communicate politely, never giving an impression of rudeness or hostility. In other words, *the system is designed to behave and sound like a professional.*

It is not surprising, therefore, that our study participants spoke about the system as if it was a social and professional actor. When asked to suggest ways to improve it, the majority of evaluators *tried to further "humanize"* and *"professionalize"* the system. The majority of participants believed that it would be a better experience for patients and more effective if the system sounded and behaved even more like a mental health professional. For example, several participants disapproved of TLC-Depression's "voice" because, they said, the "voice" sounded "mechanical" and not sufficiently "natural." Such comments made us realize that, especially in the absence of visual cues, the tone of "voice" and its vocal delivery of the script were critical components of a system that was designed to create an experience where, as one participant aptly said, "disbelief is suspended." At no time, however, did our participants ever forget or even have doubts about the "inanimate" nature of the automated system they were using.

Systems like TLC-Depression are deliberately designed to mimic a human agent. Their effectiveness probably depends on the extent to which they can behave like a human professional. However, if a user responds to cues as the system designers intended, it does not mean that the person is mistaking the machine for a human being or, in Freudian terms *having the illusion* that there must be a mind or soul within the object. Early in the twentieth century, Freud first used the term "illusion," in his book on religion, (Freud, 1957) to describe a type of "wishful thinking" whereby

human agents "project" their "desires, wishes, phobias," in short, the content of their subjectivity onto objects or onto the world as a whole. We did not observe such "illusion" among users of TLC-Depression.

What we did observe is clear: the users of TLC-Depression, aware of the elemental attributes of the object they were interacting with, simply understood the system the way anyone living in a post-industrial society "understands" a television, a "pet rock," an ATM machine or a computer system. The understanding is elicited by the object itself: the object is a useful and instrumental means to a particular cognitive and psychological stance – an outcome. The cognitive stance precludes any enduring consideration of the object as actually human.

Suggestions for improvement of TLC-Depression were made by users who were conscious of the fundamental difference between the object they were using and a human being. They did not unconsciously create an illusion of the system as human. To our participants, the system was supposed to function like a health professional, that is, sound and speak as if it understood the anguish and suffering of a person with depression. It was supposed to respond as a mental health care provider would. Our results suggest that the construct "anthropomorphism" is overdrawn, and does not adequately portray a nuanced picture of complexities in the human–computer interface.

Acknowledgement

This research was funded by a grant from National Institute of Mental Health.

References

Aroyo, A. (2000). Beyond butlers: intelligent agents as mentors. *Journal of Education Computing and Research, 22*, 373–382.

Ambrose, R. (2001). Humanoids designed to do work. *Proceedings of the IEEE-RAS International Conference on Humanoid Robots, 173–180*.

Babor, T. F., de la Fuente, J. R., Saunders, J., & Grant, M. (1992). *AUDIT: The alcohol use disorders identification test: Guidelines for use in primary health care*. Geneva, Switzerland: World Health Organization.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs: NJ, Prentice-Hall.

Bates, J. (1994). The role of emotions in believable agents. *Communication of the ACM, 37*, 122–125.

Card, W. O., & Lucas, R. W. (1981). Computer interrogation in medical practice. *International Journal of Man-Machine Studies, 14*, 49–57.

Cramer, J. A., Mattson, R. H., Prevey, M. L., Scheyer, R., & Quellette, V. (1989). How often is medication taken as prescribed? A novel assessment technique. *Journal of American Medical Association, 261*, 3273–3277.

Cramer, J. A., & Rosenheck, R. (1999). Enhancing medication compliance for people with serious mental illness. *Journal of Nervous and Mental Disease, 187*, 53–55.

Cotterill, R. (1998). *Enchanted looms: Conscious networks in brains and computers*. Cambridge, UK: Cambridge University Press.

Delichatsios, H. K., Friedman, R. H., Glanz, K., Tennstedt, S., Smigelski, C., Pinto, B. M., et al. (2001). Randomized trial of a "talking computer" to improve adults' eating habits. *American Journal of Health Promotion, 15*, 215–224.

Dennet, D. (1996). *Kinds of minds*. New York: Basic Books.
Dennett, D. (1998). *Brainchildren: Essays on designing minds*. Cambridge: The MIT Press.
Engstrom, F. (1991). Clinical correlates of antidepressant compliance. In J. A. Cramer & B. Spilker (Eds.), *Patient compliance in medical practice and clinical trials*. New York: Raven Press.
Freud, Z. (1957). *The future of an illusion*. New York: Norton.
Friedman, B. (1995). "It's the computer's fault" – Reasoning about computers as moral agents April 7–11. *CHI' 95*.
Friedman, B. (1997). *Human values and the design of computer technology*. New York: Cambridge University Press.
Friedman, R. H. (1998). Automated telephone conversations to assess health behavior and deliver behavioral interventions. *Journal of Medical Systems, 22*, 95–102.
Friedman, R. H., Kazis, L. E., Jette, A., Smith, M. B., Stollerman, J., Torgerson, J., et al. (1996). A telecommunication system for monitoring and counseling patients with hypertension: impact on medication adherence and blood pressure control. *American Journal of Hypertension, 9*, 285–292.
Friedman, R. H., Stollerman, J. E., Mahoney, D. M., & Rozenblyum, L. (1997). The virtual visit: using telecommunications technology to take care of patients. *Journal of American Medical Informatics Association, 4*, 413–425.
Friedman, R. H., Stollerman, J., Rozenblyum, L., Belfer, A. S., Mahoney, D., Steinback, S. (1998). A telecommunications system to manage patients with chronic disease. *Presentatios Proceedings: American Medical Informatics Association Annual Symposium*, 1330–1334.
Fogg, B. J., & Nass, C. (1997). Silicon sycophants: the effects of computers that flatter. *International Journal of Human–Computer Studies, 46*, 551–561.
Haynes, R. B., Taylor, D. W., & Sackett, D. L. (1979). *Compliance in health care*. Baltimore: Johns Hopkins University Press.
Jarvis, K. L., Friedman, R. H., Heeren, T., & Cullinane, P. M. (1997). Older women and physical activity: using the telephone to walk. *Women's Health Issues, 7*, 24–29.
Kahn, P. H., Friedman, B., & Hagman, J. (2002). "I care about him as a pal": Conceptions of robotic pets in online aib discussion forums April 20–25. *CHI'2002*.
Koda, T., & Maes, P. (1995). *Agents with faces: The effects of personification of agents*. Cambridge, MA: MIT Media Lab.
Kroenke, K., Spitzer, R. H., & Williams, J. B. W. (2001). The PHQ-9 validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*, 1497–1525.
Moon, Y. (1998). Are computers scapegoats? Attributions of responsibility in human–computer interaction. *International Journal of Human–Computer Studies, 49*, 79–94.
Miller, W., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change*. New York: The Guilford Press.
Manovich, Lev (1995). From the externalization of the psyche to the implantation of technology 90–100. In Rstzer Florian (Ed.), *Mind revolution interface brain/computer*. Munich: Akademie zum Dritten Jahrtausend.
Marakas, G. M. (2000). A theoretical model of differential social attributions toward computing technology: when the metaphor becomes the model. *International Journal of Human–Computer Studies, 52*, 719–750.
Nass, C., Steuer, J., Henriksen, L., & Dryer, C. (1994). Machines, social attributions, and ethopoeia: performance assessments of computers subsequent to "self" or "other" evaluations. *International Journal of Human–Computer Studies, 40*, 543–559.
Nass, C., & Moon, Y. (2000). Machines and mindlessness: social responses to computers. *Journal of Social Issues, 56*, 81–103.
Nass, C., Steuer, J., Tauber, E. R. Computers are social actors. Available from www:http:/hcistanford.edu/cs147/readings/casa/.
Person, P., Laaksolahti, J., & Lonnqvist, P. (2000). *Anthropomorphism – A multi-layered phenomenon*. Kista, Sweden: American Association for Artificial Intelligence.
Ramelson, H. Z., Friedman, R. H., & Ockene, J. K. (1999). An automated telephone-based smoking cessation education and counseling system. *Patient Education and Counseling, 36*, 131–143.

Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.

Slack, W. V. (1997). *Cybermedicine: How computing empowers doctors and patients for better health care*. San Francisco: Jossey-Bass Publishers.

Searle, J. (1983). *Intentionality: An essay in the philosophy of mind*. New York: Cambridge University Press.

Thompson, C., Peveler, R., Stephenson, D., & McKendrick, J. (2000). Compliance with antidepressant medication in the treatment of major depressive disorder in primary care: a randomized comparison of fluxetine and a tricyclic antidepressant. *American Journal of Psychiatry, 157*, 338–343.

Turkle, S. (1984). *The second self: Computers and the human spirit*. New York: Simon & Schuster.

Zuboff, S. (1988). *In the age of the smart machine the future of work and power*. New York: Basic Books.

Epidemiologic Perspectives & Innovations

Methodology

Translational methods in biostatistics: linear mixed effect regression models of alcohol consumption and HIV disease progression over time

Mariel M Finucane[1], Jeffrey H Samet[2,3] and Nicholas J Horton*[1]

Address: [1]Department of Mathematics and Statistics, Smith College, Northampton, MA, 01063, USA, [2]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston MA, 02118, USA and [3]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, 02118, USA

Email: Mariel M Finucane - mfinucan@hsph.harvard.edu; Jeffrey H Samet - jsamet@bu.edu; Nicholas J Horton* - nhorton@email.smith.edu

* Corresponding author

Published: 19 September 2007

Epidemiologic Perspectives & Innovations 2007, **4**:8 doi:10.1186/1742-5573-4-8

Received: 4 May 2006
Accepted: 19 September 2007

This article is available from: http://www.epi-perspectives.com/content/4/1/8

Abstract

Longitudinal studies are helpful in understanding how subtle associations between factors of interest change over time. Our goal is to apply statistical methods which are appropriate for analyzing longitudinal data to a repeated measures epidemiological study as a tutorial in the appropriate use and interpretation of random effects models. To motivate their use, we study the association of alcohol consumption on markers of HIV disease progression in an observational cohort. To make valid inferences, the association among measurements correlated within a subject must be taken into account.

We describe a linear mixed effects regression framework that accounts for the clustering of longitudinal data and that can be fit using standard statistical software. We apply the linear mixed effects model to a previously published dataset of HIV infected individuals with a history of alcohol problems who are receiving HAART (n = 197). The researchers were interested in determining the effect of alcohol use on HIV disease progression over time. Fitting a linear mixed effects multiple regression model with a random intercept and random slope for each subject accounts for the association of observations within subjects and yields parameters interpretable as in ordinary multiple regression. A significant interaction between alcohol use and adherence to HAART is found: subjects who use alcohol and are not fully adherent to their HIV medications had higher log RNA (ribonucleic acid) viral load levels than fully adherent non-drinkers, fully adherent alcohol users, and non-drinkers who were not fully adherent.

Longitudinal studies are increasingly common in epidemiological research. Software routines that account for correlation between repeated measures using linear mixed effects methods are now generally available and straightforward to utilize. These models allow the relaxation of assumptions needed for approaches such as repeated measures ANOVA, and should be routinely incorporated into the analysis of cohort studies.

Background

The National Institute on Alcohol Abuse and Alcoholism estimates that more than 13 million Americans suffer from alcohol dependence or abuse [1]. HIV infection has major health consequences, with estimates of 940,000 infected Americans [2]. These two health concerns are related, and alcohol problems have been reported to be more prevalent in HIV-infected patients. Among 665 patients who were establishing primary care for HIV infection, half were determined to have an alcohol problem based on the CAGE questionnaire or clinical assessment [3]. Before the advent of highly active antiretroviral therapy (HAART), however no association between alcohol use and HIV disease progression was found [4]. Samet et al. [5] hypothesized that in the age of HAART, alcohol use, because of its potential interaction with a variety of HIV clinical issues including medication adherence, might accelerate HIV disease progression. They found that, among a cohort of HIV-infected individuals with a history of alcohol problems (the HIV-Alcohol Longitudinal Cohort, or HIV-ALC), those individuals receiving HAART and consuming alcohol had significantly higher viral RNA (ribonucleic acid) levels at baseline.

In this paper, we provide a tutorial on linear mixed effect models to study repeated measures in this dataset including follow-up data collected on subjects in the HIV-ALC cohort [6]. Longitudinal cohort studies have the advantage of providing detailed information about how a given set of variables changes over time in an individual patient and of facilitating the study of the factors that influence this change. By collecting repeated measurements, we gain the ability to distinguish between the degree of variation across time for one person (within-individual change), and the variation among people (between-individual change). However, longitudinal studies present some statistical complexities, since the customary assumption that all observations are independent usually does not hold.

In addition to the usual assumptions of regression methods, models for a single outcome assume that all observations of a particular variable are independent of one another: knowing the value of one observation of a variable provides no information about the others, after controlling for known covariates. This assumption does not hold true in longitudinal studies, however, as multiple observations of a variable on a particular person are likely positively correlated (i.e. the errors may reflect a systematic trend within each individual).

One approach to this problem involves excluding all followup data from the analysis and using only the baseline data from the cohort – in this single-time-point subset of the original dataset, the assumption of independence of observations is plausible. However, this method utilizes only part of the available data, and is highly inefficient and inadvisable. Unless the correlation is quite high between baseline and follow-up data, such an inefficient approach will lead to less precise estimates, and does not allow for assessment of time-varying exposures and outcomes.

Another approach is to assume that repeated measurements on an individual are independent despite the fact that they are likely correlated. This may introduce bias into the estimates of variability of the models' parameters, and is not recommended (see [7] for a case study of the perils of this mis-modeling).

A more principled approach, which we will illustrate in this paper, involves modeling the within-individual relatedness (clustering) of measurements in order to make use of all the data and simultaneously obtain unbiased estimates of parameter variability. Although models that take this clustering into consideration are more complicated, they are also more powerful since they facilitate the study of change over time. These issues have received a great deal of attention in the statistical literature in recent years, and the books by Diggle and colleagues [8] and Fitzmaurice, Laird and Ware [9] provide excellent overviews of the field.

A classic method used to account for repeated measurements in linear models is repeated measures analysis of variance (RM-ANOVA). This model was developed for settings with discrete covariates, complete data, and common measurement occasions for all subjects [9]. This approach has some disadvantages in practice, however, since in many longitudinal studies observations may be unbalanced and/or incomplete and assumptions regarding equal covariance between all observations may not be tenable [10]. As we will illustrate, other approaches, such as the linear mixed effects model that we describe, are more attractive in this setting.

In this paper, we describe the linear mixed effects (LME) or random effects/random coefficients model of Laird & Ware [11], a versatile model that accounts for clustering. Other approaches to estimation in this setting are discussed by Fitzmaurice et al. [9]. The LME approach provides a flexible yet parsimonious way of modeling the association among repeated measurements. These within-subject associations are often of secondary interest in longitudinal studies and the parameters that describe them in the LME model are thus termed *nuisance parameters*. The *substantive parameters* are those that describe the relationship of primary interest between study variables. The LME approach estimates the nuisance parameters and substantive parameters simultaneously, yielding consistent esti-

mates of the substantive parameters if the model for the covariance and the mean are appropriately specified.

Samet et al. [5] conducted a cross-sectional analysis of the baseline data from a cohort of HIV-infected individuals with a history of alcohol problems (HIV-ALC). In a multiple linear regression model that controlled for a number of potential confounding variables, they found that among subjects who were on HIV medications, those subjects who used any alcohol (moderate or at risk use) had significantly higher mean viral log RNA levels (p = .006) than subjects who reported no drinking (fully abstinent) during the previous 30 days; this association was attenuated (p = 0.04) when adherence to HIV medications was included as a predictor in the model. Further analysis of data from this cohort has been reported [6].

In this paper, we will fit LME models to conduct a secondary analysis, using longitudinal methods to further explore the issues they considered in their baseline analysis. The goal of our analysis will be to explicate linear mixed effects models in the context of understanding the association between alcohol consumption and the progression of HIV/AIDS. Using repeated measures data will allow us to take full advantage of all information available in this cohort study and to assess how the associations between alcohol use and HIV RNA levels changed over time. Use of the LME is preferable to other approaches such as classical RM-ANOVA, because it allows loosening of assumptions that may not be tenable. While these methods are particularly well-suited to the analysis we consider, they are also applicable to many other types of longitudinal epidemiology studies.

Analysis

Methods

We perform a secondary analysis of the HIV-Alcohol Longitudinal Cohort (HIV-ALC), a follow-up study of HIV-infected patients with past or current history of alcohol problems. The primary purpose of this longitudinal cohort was to examine HIV progression of these subjects, and prior results have been published previously [12-14]. Participants were recruited between July 1997 and July 2001. All participants resided in the Greater Boston area and were recruited through the following sources: Boston Medical Center (BMC) Diagnostic Evaluation Unit (56%), posted fliers (17%), BMC Primary Care Clinic (13%), respite facility for homeless persons (5%), methadone clinic (4%), subject referrals (4%), and Beth Israel Deaconess Medical Center (BIDMC) (2%). Persons recruited outside BMC or BIDMC were pre-screened by telephone, and potentially eligible individuals were invited to complete the screening process in person. The Institutional Review Boards (IRB) of BMC, BIDMC and Smith College approved this study.

Patients who were HIV-infected and had a history of alcohol problems were identified by explicit eligibility criteria: confirmed HIV infection and a history of alcohol problems. Patients not receiving care at BMC or BIDMC were asked to document their HIV diagnosis by providing either HIV testing documentation or their HIV prescription medications. Clinical assessment (in 10% of participants) or 2 or more positive responses to the CAGE questionnaire (in 90%) were used to identify participants with the criterion 'history of alcohol problems.' The CAGE questionnaire [15] is a short, validated questionnaire with good reliability in identifying problem drinkers. Diagnostic interviews for alcohol problems in a sub-sample of CAGE-positive subjects (n = 141) revealed a lifetime history of alcohol dependence (80%) or abuse (15%) [3]. Additional entry criteria included the following: evidence of unimpaired cognitive function as determined by a score of 21 or more on the Mini Mental State Examination [16]; no plans to leave the Boston area during the subsequent 2 years; and fluency in English or Spanish. For the Spanish interview instrument, standardized scales in Spanish were used when available; the remainder of the questionnaire was translated from English into Spanish, back-translated and revised.

There were 444 eligible subjects screened at these various sites, 350 (79%) provided informed consent and agreed to participate in the study. After providing informed consent, subjects were scheduled to be interviewed every 6 months for a maximum of 7 visits. Since no follow-up occurred after July 2001, subjects enrolled late in the study had only a baseline assessment and a small number of follow-up visits. Laboratory values of HIV RNA measured within 3 months of each visit were obtained from medical records whenever available. If not measured during routine clinical care, blood samples were drawn during the visit by nursing staff and tested for HIV RNA level. Participants were compensated US$20 or an equivalent gift certificate to a local grocery store. In this paper we analyze the subset of this cohort made up of all participants receiving highly active antiretroviral therapy (HAART) at baseline (n = 197).

Outcome variable

The base 10 log of one plus the viral load of HIV RNA (log(RNA+1)) was used as a primary outcome of HIV disease progression in the HIV-ALC study. Measurement of HIV RNA was performed using branched-chain DNA techniques [17]. The lower threshold for detection at the time of the study was 50 copies/ml – values < 50 were analyzed as 0. We note that this approach is somewhat ad-hoc, and more principled approaches have been developed [18].

Measures of alcohol consumption

In the HIV-ALC study, alcohol use in the 30 days before each interview was used as a measure of the usual pattern of use. To encourage accurate reporting of alcohol consumption, breath alcohol level was also measured before the interview [19]. Alcohol consumption was calculated using alcohol quantity and frequency questions as well as the Addiction Severity Index, an assessment instrument with well-documented reliability and validity in this population [20]. Alcohol use was initially classified as 'abstinent', 'moderate', or 'at-risk', based on the National Institute on Alcohol Abuse and Alcoholism (NIAAA) recommendations, which define at risk drinking as more than 14 drinks per week (or more than four in one day) for men, and more than seven drinks per week (or more than three in one day) for women [1]. Any alcohol consumption below these levels was considered moderate use in this study. Following the approach of Samet et al [5], a dichotomous indicator of any consumption (yes/no) was used in this analysis. Alcohol use was measured at each study visit and thus varied over time.

Measures of adherence to HIV medication

Adherence to HAART was self-reported using the AIDS Clinical Trials Group instrument. Subjects reported the names of their antiretroviral medications as well as the number of doses and the total number of pills prescribed and taken daily [21]. The 3-day self-reported number of pills missed was computed for each HIV medication. Adherence was defined as a proportion of prescribed doses taken (0–1).

Other factors

Basic demographic information included each participant's age in years, race/ethnicity (4 groups: black, white, Latino, other), and gender. Homelessness was defined as spending at least one night in a shelter or on the street in the 6 months prior to the interview. The number of doses of therapy each subject received per day was also recorded. A group of 151 subjects in the cohort participated in a randomized controlled trial of a HAART adherence intervention [22]. Involvement in the ADHERE trial was included as a three-category variable in our analysis (intervention/control/not enrolled). Another three-category variable was included that described a subject's primary HIV risk factor: injection drug use, men having sex with men, or heterosexual sex. Finally, to assess whether there were any important cohort effects due to date of entry in the study, we categorized subjects according to the year in which they entered the study (1997–2001).

In this secondary analysis we wanted to replicate the results of Samet et al [5] relating to adherence and alcohol consumption, while utilizing the additional information regarding follow-up observations. In addition, we extended the previous analysis to assess whether subjects' adherence to medications was acting as an effect modifier for the association of alcohol consumption with HIV progression, i.e. whether the effect that alcohol had on RNA levels differed across degrees of HAART adherence.

Statistical methods using linear mixed effects regression models

We fit linear mixed effects (LME) regression models ([9,11]) for log(RNA+1) levels over time. To help ground the discussion of these models in the context of our example, we display (in Figure 1) the observed log(RNA+1) levels, adherence percentages, and alcohol abstinence values (yes/no) for a sample of nine subjects. LME models account for clustering of longitudinal data points (for example, note that subject 4011's log(RNA+1) values are consistently higher than subject 4189's) and thus provide valid estimates of the regression parameters of interest and their standard error. LME methods are also attractive in this setting because, unlike classical repeated measures ANOVA [22], they can loosen assumptions regarding the form of associations within subjects and incorporate imbalance in longitudinal data (note that subject 4180 is only observed at baseline, whereas subject 4129 has six follow-up visits). Furthermore, LME models distinguish within-subject from between-subject sources of variation, and also describe how individual and population mean response trajectories change over time. At the same time, LME models are particularly useful because their covariance structures can often be described in a flexible and parsimonious fashion.

The underlying premise of LME methods is that an outcome of interest is determined by some factors that affect all subjects in the same way and by other factors that affect individuals in different ways. This premise is reflected in the LME model by dividing the mean model regression parameters into two distinct groups: fixed effects (or population effects) and random effects (or subject-specific effects). The fixed effect parameters are shared by the entire study population. The other parameters, the random effects, are allowed to vary randomly from one individual to another. These random effects are attractive because one participant's RNA levels might consistently be higher than the mean while another's might be lower due to unmeasured factors such as genetic make-up, immunologic factors, HIV mutations conferring resistance, environment, education, personal habits, etc. These differences would be reflected in the random-effects portion of the model, thus allowing each individual to have his/her own subject-specific mean response trajectory over time. These random effects parameters reflect the natural heterogeneity of the population and thus account for within-individual clustering of data points.

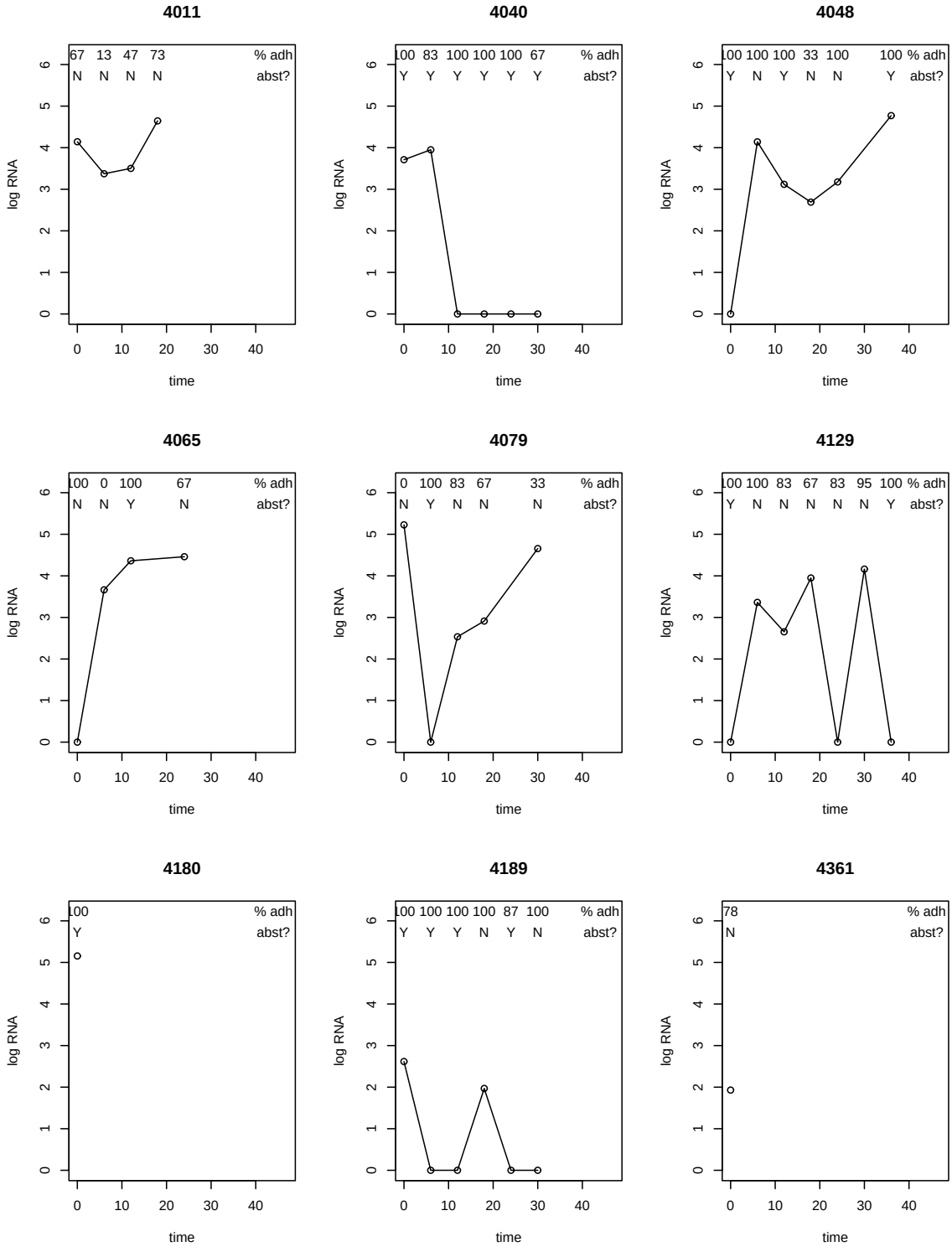


Figure 1
Observed log(RNA+1), adherence and abstinence status over time for 9 subjects.

In the LME framework, each participant's outcome trajectory is modeled as a combination of the population characteristics that are assumed to be shared by all individuals (fixed effects), and that individual's unique subject specific effects (random effects). The mean response trajectory in the population is obtained using a weighted average of the random effects, which are almost always a subset of the fixed effects. We will now introduce two important special cases of the LME model using the notation of Fitzmaurice et al. [9].

Random intercept model

The most straightforward case of a linear mixed-effects model is one in which each subject has only one random effect – a randomly determined intercept (or individual level). This model assumes that controlling for a subject's level (intercept) sufficiently accounts for the association between repeated measurements.

To illustrate this approach, consider a study with only two time-points and two levels of drinking. In this study, subject *i*'s predicted log(RNA+1) level at time-point *j* would be given by:

$$E\,[RNA_{ij}|t_i, drk_{ij}, b_i] = \beta_0 + \beta_1 t_{i1} + \beta_2 t_{i2} + \beta_3 drk_{ij} + b_i$$

◦ β_0 is the population's average intercept.

◦ t_1 and t_2 are dummy variables for time; β_1 and β_2 are their associated fixed effect regression parameters.

◦ $drk_{ij} = 1$ if person *i* was using alcohol at time *j*, $drk_{ij} = 0$ if person *i* was abstaining at time *j*; β_3 describes the population effect of alcohol use on log(RNA+1) levels.

◦ b_i is person *i*'s random intercept. In particular, b_i represents the deviation of the i^{th} individual's intercept from the population's intercept β_0.

By averaging over the distribution of the subject-specific effects b_i, we obtain the mean response profile in the population characterized by the fixed effect regression parameters of interest:

$$\overline{RNA_j} = \beta_0 + \beta_1 t_{i1} + \beta_2 t_{i2} + \beta_3 drk_{ij}.$$

In Figure 2, we use an illustration due to Fitzmaurice et al. [9] to demonstrate how this model could be applied in a simple example. In this example, as shown on the left side of the Figure, person A's measurements are consistently higher than person B. Thus, person A would have a positive random intercept term ($b_A > 0$) whereas person B's random intercept would be negative ($b_B < 0$). The individuals' observed responses are allowed to vary randomly above and below their conditional mean trajectories

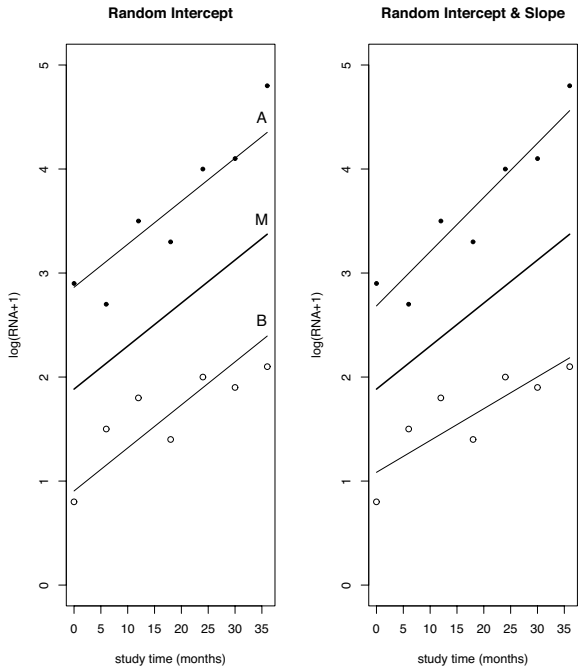


Figure 2
Hypothetical observed and predicted lines for two subjects from random intercept and random slope model.

because of the inclusion of the error terms e_{ij}. (In this example, e_{A1} is positive whereas e_{A2} is negative.) By averaging over these random effects, we obtain the marginal mean response trajectory, M, the predicted outcome trajectory for an 'average' subject in the population, described using the fixed effects parameters.

The usefulness of the random intercept model however, is limited by the fact that this model constrains the correlation between repeated measurements to be the same no matter how close or far apart in time the measurements are taken. For example, in a study with six time-points, the random intercept model makes the restrictive assumption that the correlation between subject *i*'s first and second measurements is the same as the correlation between subject *i*'s first and sixth measurements: $corr\ (Y_{i1}, Y_{i2}) = corr\ (Y_{i1}, Y_{i6})$. This assumption of time-invariant correlation is likely to be unrealistic for repeated observation of HIV RNA (e.g. we would expect two consecutive measurements to be more tightly correlated than two measurements taken far apart in time). We will now introduce a more flexible model that allows correlations between repeated measures to change over time.

Random intercept and slope model

Another set of random effects covariance structures, which makes less restrictive assumptions about the associations between measurements, arises when additional parameters (besides the intercept) are considered random and subject-specific. The random intercept and slope model allows both the intercept and slope to vary randomly among subjects. The model given above would thus change slightly:

$$E\,[RNA_{ij}|t_i,\, drk_{ij},\, b_i] = \beta_0 + \beta_1 t_{i1} + \beta_2 t_{i2} + \beta_3 drk_{ij} + b_{0i} + b_{1i} t_{ij}$$

◦ b_{0i} is subject i's random intercept.

◦ b_{1i} is subject i's random slope.

The random intercept and slope model can best be understood by considering another simple two-person experiment, depicted on the right side of Figure 2, again due to Fitzmaurice et al. [9]. In this example, person A's intercept and slope are greater than the mean intercept and slope respectively ($b_{0A} > 0$, $b_{1A} > 0$) because his or her measurements are on average higher than the average in the group and are also increasing at a faster rate. By the same token, the intercept and slope of person B are less than the population averages for intercept and slope ($b_{0B} < 0$, $b_{1B} < 0$). So, the mean trajectory has intercept β_0 and slope β_1; person A has intercept $\beta_0 + b_{0A} > \beta_0$ and slope $\beta_1 + b_{1A} > \beta_1$; person B has intercept $\beta_0 + b_{0B} < \beta_0$ and slope $\beta_1 + b_{1B} < \beta_1$. As was the case in the random intercept model, the inclusion of the error terms e_{ij} allows the observed measurements to deviate randomly from the subject-specific trajectories.

Importantly, the covariance structure of this model is less restrictive than the random intercept model. In particular, the random intercept and slope model allows correlations between measurements to change with time – the correlation between Y_{ij} and Y_{ik} is modeled as a function of the times of measurement. To illustrate the increase in flexibility that comes from including a random slope, we return to the example of the HIV-ALC dataset, noting that in the random-intercept model, the correlation between any two measurements on the same person, regardless of how far apart they were in time, was constrained to be *corr* $(Y_{ij}, Y_{ik}) = 0.40$. The addition of a random slope, however, allowed this quantity to vary according to the times of measurement: in the random-intercept-and-slope model, *corr* $(Y_{i0}, Y_{i6}) = 0.47$ and *corr* $(Y_{i0}, Y_{i36}) = 0.16$. This agrees with intuition – one would expect measurements of viral RNA made 6 months apart in time to be more tightly correlated than measurements made three years apart. In addressing the HIV data analysis with repeated measurements taken months apart, this is an attractive feature, and we adopted this approach.

Estimation of the regression parameters of interest as well as the variance-covariance matrix of the random effects (assuming multivariate normality) proceeds simultaneously. Two options for maximization include a standard likelihood or a restricted likelihood (REML), where the former is biased in small samples. An extensive discussion of estimation can be found in [11].

It is often the case that primary scientific interest lies in the interpretation of the parameters that describe the mean and subject-specific trajectories. An appealing feature of LME models is that after the within-subject association has been accounted for (using either the random intercept model, the random intercept and slope model, or a more complex model), the nuisance parameters that describe this covariance structure can typically be ignored and focus can be given to interpreting the substantive parameters. Settings where the variance parameters are of interest in their own right (such as studies of observer variation) can be accommodated by LME models as well.

While we have focused on models for responses that are approximately Gaussian, extensions to other types of outcomes (e.g. counts or dichotomous variables) have been undertaken using the mixed effects framework. The text by Fitzmaurice, Laird and Ware [9] provides an accessible introduction to random effects for the generalized linear model.

We now turn to the specification of the log(RNA+1) regression model for the fixed effects parameters. The inclusion of follow-up data in our analysis enabled us to study the effect of time on HIV RNA levels as well as the interaction between time and alcohol use, our covariate of primary interest. Entry into the study was not linked to any treatment or clinical event, and we therefore had no specific hypotheses about relationships among time, time-varying measures of alcohol consumption and adherence, and changes in log(RNA+1) viral loads.

We included a relatively rich set of covariates into our linear mixed effects model, incorporating: time (df = 6), age, race/ethnicity (df = 3), gender, any report of homelessness, number of doses of HAART prescribed per day, adherence to HAART, involvement in the ADHERE study (df = 2) [22], year of entry into the HIV-ALC study, and primary HIV risk factor. We justified the use of this 'inclusive approach' [23] in two ways. First, any of these variables could potentially be a confounder of the relationship of interest. Second, in a study in which only 44% of intended observations were made, the inclusion of what Collins et al. [23] call 'auxiliary variables' is suggested. While the focus of this paper is not on missing data, we note that LME methodology yields consistent estimates when missing data are *missing at random* in the sense of

Little and Rubin [24]. Informally, this means that missingness is ignorable when it is related only to observed quantities. By incorporation of additional information already collected in the study, assumptions regarding the ignorability of missing data become more plausible (yet we note that the validity of the assumption of ignorability remains inherently untestable without additional data regarding missing observations).

Pairwise interactions between time, alcohol consumption and adherence were included in the model, and retained if their p-values were less than 0.10. These interactions were considered because there was substantive interest in these factors. Because of the complications in interpreting multiple degree of freedom interactions, only those achieving a modest degree of statistical significance were included in the final model.

In the cross-sectional analysis conducted by Samet et al. [5], controlling for adherence to HAART yielded attenuated results (p = 0.04 as opposed to p = .006). To further explore whether adherence might be an effect modifier of the drinking/RNA association (i.e. whether the effect of drinking on RNA was modified depending on the values of adherence), we decided to test for the significance of the interaction effect between alcohol use and adherence in our model. This effect was moderately significant (p = .02) and the interaction was retained. R version 2.4.1 and SAS version 9.1 were used for estimation. The Appendix provides the syntax needed to fit the LME model in three general purpose statistical packages: R, SAS and Stata.

Results

Table 1 describes the analytic sample of the HIV-ALC cohort on HAART. Key characteristics include the following: 58% report injection drug use as their primary HIV risk factor; 22% are homeless; 40% currently using alcohol, 18% are female, and the average age is 40 years.

Table 2 displays the results of the multiple longitudinal LME regression model of log(RNA+1) levels. Alcohol use was a significant predictor of log(RNA+1) levels with abstainers having lower levels of RNA on average. There was a significant interaction between alcohol consumption and adherence (p = 0.02) and the interaction was retained. Subjects who used alcohol and were less adherent to their medications had significantly higher log(RNA+1) levels than non-drinkers who had better adherence to their HAART regimens, alcohol users who had better adherence to their HAART regimens, and non-

Table 1: Characteristics of the HIV-ALC Cohort on HAART at baseline (n = 197)

	Percent	Count
Primary HIV risk factor		
Men sex with men	21%	42
Injection drug use	58%	115
Heterosexual sex	20%	40
Race/ethnicity		
Black	41%	80
White	37%	73
Latino	22%	43
Other	1%	1
Uses alcohol	40%	79
Female	18%	36
Homeless	22%	43
Enrollment year		
1997	10%	19
1998	33%	65
1999	37%	72
2000	16%	32
2001	5%	9
ADHERE enrollment		
Not enrolled	49%	96
Control	26%	52
Intervention	25%	49
	Mean (SD)	min, max
Doses of HAART/day	5.0 (1.6)	2, 10
3 day HAART adherence	0.9 (0.2)	0, 1
Age	40.8 (7.4)	19.5,66.2
Log_{10}(RNA+1)	2.0 (1.9)	0, 5.7

Note: due to rounding some values may not sum to 100%

Table 2: Summary of LME Model of Log_{10}(RNA+1) (n = 618 observations derived from 197 subjects)

	Est (SE)	p-value	Multiple *df* p-value
Intercept	1.9 (.98)	.06	
Time			.37 (*df* = 6)
Time0	.58 (.48)	.23	
Time6	.69 (.48)	.15	
Time12	.84 (.48)	.08	
Time18	.88 (.50)	.08	
Time24	.58 (.49)	.24	
Time30	.18 (.49)	.72	
Time36	0	.	
Drink	3.6 (.97)	.0003	
Adherence	-.39 (.58)	.50	
Time*Drink			.01 (*df* = 6)
Time0*Drink	-1.7 (.81)	.04	
Time6*Drink	-2.2 (.81)	.006	
Time12*Drink	-2.4 (.82)	.003	
Time18*Drink	-1.9 (.82)	.02	
Time24*Drink	-1.5 (.84)	.07	
Time30*Drink	-.93 (.85)	.27	
Time36*Drink	0	.	
Drink*Adherence	-1.6 (.67)	.02	
Age	-.02 (.01)	.12	
Female	.21 (.27)	.45	
Homeless	.04 (.19)	.85	
Doses/day	-.02 (.05)	.65	
Enrollment year	.22 (.11)	.04	
Race/ethnicity			.22 (*df* = 3)
Black	.21 (.23)	.37	
Latino	.04 (.28)	.88	
Other	1.8 (.93)	.05	
White	0	.	
ADHERE assignment			.27 (*df* = 2)
Non ADHERE	-.39 (.24)	.11	
ADHERE treatment	-.16 (.25)	.17	
ADHERE control	0	.	
Primary HIV risk factor			.25 (*df* = 2)
Men sex with men	.46 (.33)	.17	
Injection drug use	.44 (.27)	.11	
Heterosexual sex	0	.	

drinkers who were less adherent to their HAART regimens. To help interpret the statistically significant interaction between alcohol consumption and adherence, Figure 3 shows the predicted log(1 + RNA) trajectories over time for four hypothetical subjects, one who doesn't drink and is 100% adherent, one who doesn't drink and is 0% adherent, one who uses alcohol and is 100% adherent, and one who uses alcohol and is 0% adherent (each hypothetical subject represents an 'average' subject with respect to baseline covariates). The time by adherence interaction effect was dropped as there was little evidence that it added to the model (p = 0.65).

To assess the value of incorporating follow-up data, we also fit a cross-sectional model utilizing only data from the baseline timepoint (n = 197, Table 3). The significance of the alcohol consumption by adherence interaction was attenuated (p = 0.053), in part due to the reduced sample size.

We also assessed the importance of accounting for clustering, by fitting a model for all time points that inappropriately ignored the correlation. This incorrect model yielded a spuriously statistically significant interaction effect for alcohol consumption and adherence (p = 0.0006).

For any model, it is important to verify assumptions made in estimation. For the LME random intercept and slope model, in addition to assumptions of standard multiple regression models, estimation proceeds assuming that the distribution of the random intercepts and slopes is approximately bivariate normal. Figure 4 displays a histo-

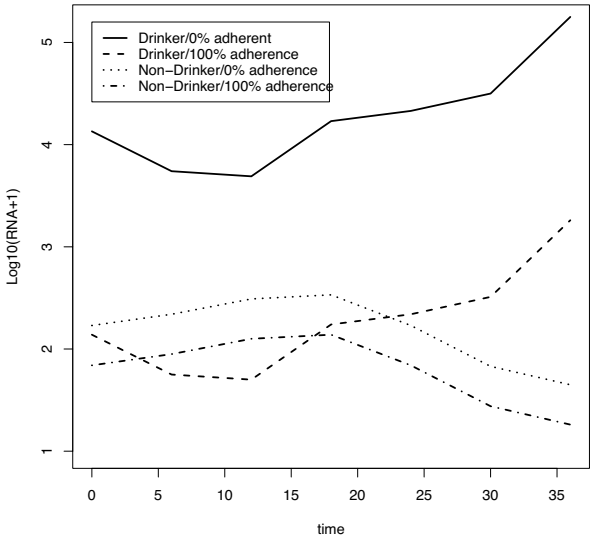


Figure 3
log(RNA+1) predicted values over time from random intercept and slope model by adherence and abstinence status.

gram (with normal [mean = 0, variance = 1.5] density overlaid) of the random intercepts, while Figure 5 displays the histogram of the random slope parameters (with normal [0,0.0019] overlaid). Neither histogram presents strong evidence against the normality assumption.

Figure 6 displays the scatterplot of random slopes and intercepts, which appears to be a cloud of points consistent with a bivariate normal density, albeit with a number of points on the line given by slope = -.02*intercept. The 9 subjects labeled in Figure 1 are indicated to help interpret this scatterplot. For subjects with only one observation (e.g. 4180, 4361) the predicted slope is essentially *borrowed* from other values within the sample. Many of the subjects in the HIV-ALC cohort were observed at only 1 timepoint. The negative correlation (-0.53) of the cloud of points indicates that there is an inverse association between intercepts and slopes: subjects with low log(RNA+1) values at baseline are likely to see increases in log(RNA+1) values over time (i.e. have more potential for increase), while subjects with high log(RNA+1) values at baseline are likely to decrease over time. Subjects whose log(RNA+1) values increased over time had positive random slopes (e.g. 4065, 4048) while subjects with a decrease had negative slopes (e.g. 4189 and 4140).

Conclusion

Our primary goal was to motivate and illustrate the use of linear mixed effect regression models for longitudinal epidemiologic data in an alcohol research setting. Using the LME model to account for clustering within subject enabled us to make full use of the available data. Whereas Samet et al. [5] conducted a cross-sectional analysis of the baseline data from the HIV-ALC cohort, we analyzed data collected at 7 different timepoints. By utilizing this additional information, we were able to detect the subtle interaction effect between adherence and alcohol

Table 3: Summary of Linear Regression Model of Log_{10}(RNA+1) at baseline (n = 197 observations)

	Est (SE)	p-value	Multiple *df* p-value
Intercept	1.0 (1.3)	.43	
Drink	2.6 (1.2)	.03	
Adherence	-.37 (1.0)	.71	
Drink*Adherence	-2.4 (1.2)	.05	
Age	0.0 (.02)	.98	
Female	.58 (.37)	.12	
Homeless	.44 (.32)	.18	
Doses/day	-.07 (.09)	.43	
Enrollment year	.32 (.14)	.02	
Race/ethnicity			.94 (*df* = 3)
Black	.06 (.32)	.84	
Latino	.06 (.36)	.88	
Other	1.2 (1.8)	.53	
White	0	.	
ADHERE assignment			.36 (*df* = 2)
Non ADHERE	-.12 (.33)	.73	
ADHERE treatment	.35 (.37)	.35	
ADHERE control	0	.	
Primary HIV risk factor			.13 (*df* = 2)
Men sex with men	.83 (.7)	.08	
Injection drug use	.69 (.37)	.06	
Heterosexual sex	0	.	

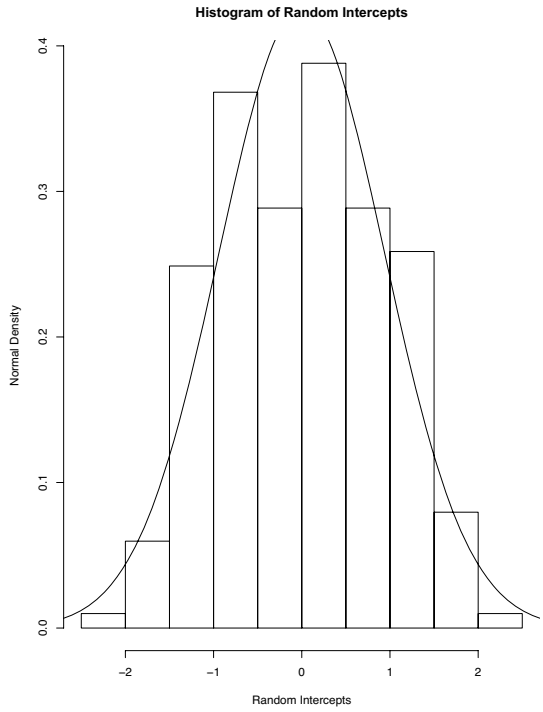


Figure 4
Histogram of random intercepts from random slope model (plus normal density).

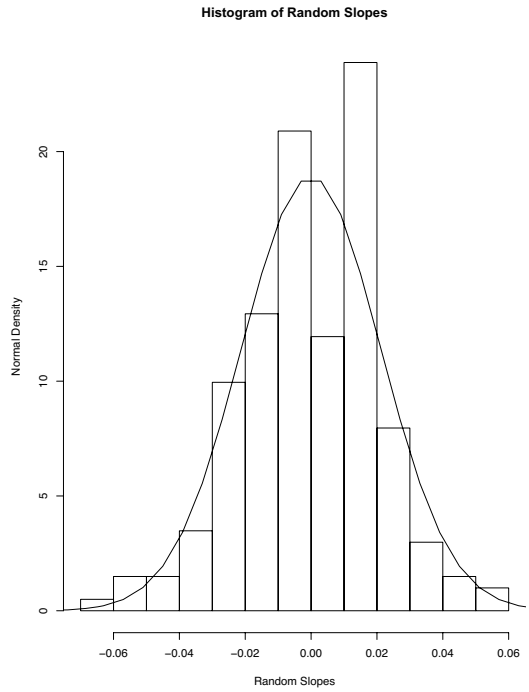


Figure 5
Histogram of random slopes from random slope model (plus normal density).

consumption that was moderately significant in our longitudinal analysis (p = .018), but only borderline significant (p = .053) in a cross-sectional baseline analysis that included all other covariates except time. While the substantive conclusions regarding this interaction are similar in both models, the potential efficiency gain of including all available observations should not be dismissed. The results of our analysis suggest that for this sample of subjects with a history of alcohol problems either adhering to one's HAART regimen, or abstaining from alcohol was significantly associated with relatively lower levels of log(RNA+1) viral load. Limitations of this investigation include self-report of alcohol consumption and adherence, missing data, and relatively modest sample size. We also note the possibility that effects of readiness to change as a predictor of subsequent drinking behavior may be mediated/moderated by factors such as "self-efficacy," which the current analyses do not directly address. Further exploration of the interacting effects of drinking and adherence thus seems merited.

The linear mixed effects models that we have described provide a flexible structure for modeling the covariance among repeated observations, thus yielding valid estimates of regression parameter variances. We concur with the advice of Fitzmaurice and colleagues who stated that given modern computing capabilities, which support a wider class of models for longitudinal data, "there is little reason to analyze longitudinal data under the inherent limitations and constraints imposed by the repeated measures ANOVA model" [9].

After the association in longitudinal studies has been accounted for, focus can shift to interpreting the substantive parameters that describe the relationships of scientific interest. The interpretation of parameters from a multiple regression model is of crucial importance, and a similar process of interpretation is needed for the LME model.

Extensions to non-normally distributed outcomes (e.g. binary or count outcomes), while not discussed in detail here, are tractable, as routines to fit both linear and nonlinear models exist in general purpose statistical software (including but not limited to R, SAS, S-plus, SPSS, and Stata). While there are a number of additional complica-

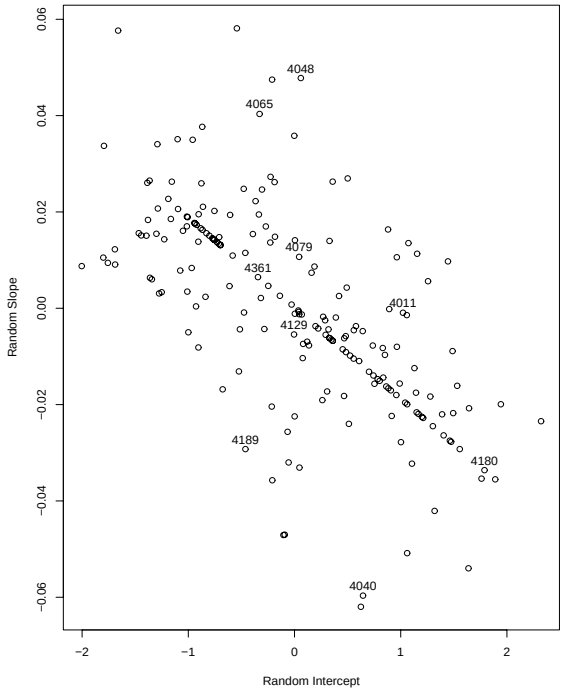


Figure 6
Scatterplot of random intercepts and random slopes (9 subjects displayed in Figure 1 are indicated by their identification numbers).

tions in fitting non-linear models, in terms of computational requirements and convergence, the general framework is analogous to the linear setting that we describe. These models are applicable to a wide range of outcomes arising in alcohol studies, and should be utilized routinely.

Incomplete observations arise in most longitudinal studies. It is rarely, if ever, the case that every planned measurement can be successfully obtained; some subset of these intended measurements are often missing. LME methods incorporate incomplete data under the assumption that missingness is at random (MAR, not related to unobserved quantities). Although estimates made under MAR have been shown to be relatively robust to small deviations from this assumption [23] this is not always true, and it is important to consider whether this (untestable) assumption is tenable. An extensive literature exists regarding the use of non-ignorable non-response models to assess sensitivity to the MAR assumption [25].

As with any model, verification of other assumptions (residual analysis, examination of influential points, etc.) is critically important. We focused attention on assumptions of normality of the random effects parameters, but other model-checking is always indicated. The multiple-bias methods of Greenland [26] provide a general framework for consideration of non-sampling errors that could affect results in substantial ways.

In this report we have provided a brief introduction and application of LME models, but have only touched on many important issues and have neglected other crucial aspects. More comprehensive descriptions of these methods exist (e.g. [8,9]) and are appropriate next steps for analysts considering use of these models. In addition, other approaches to the analysis of longitudinal or clustered data have been proposed. The population averaged generalized estimating equation (GEE) approach of Liang and Zeger [27] is another feasible approach, particularly for non-normally distributed outcomes. Extensions of the random effects framework using Bayesian estimation [28] have been utilized to address additional complexities and loosen assumptions (i.e. use of a t distribution rather than a normal distribution for the random effects). Finally, latent variable models [29] provide an attractive framework with a similar flavor, with implementations available (e.g. Stata and Mplus). These models extend the random effects model to fit multilevel factor and item response models, latent class models, and multilevel structural equation models.

Competing interests
The author(s) declare that they have no competing interests.

Authors' contributions
MF and NH designed and carried out the secondary analysis. MF helped to draft the manuscript. JS conceived of the original study, participated in its design and coordination. All authors contributed to the creation of earlier drafts, and all read and approved the final manuscript.

Appendix
Code to implement linear mixed effects models in general purpose statistical software.

```
R

1 library(nlme)

2 lmefit < – lme(logrna ~as.factor(time) +
drkhaz2 + pct3d +

3      as.factor(time)*drkhaz2 +
drkhaz2*pct3d + age +
```

```
4       as.factor(race) + female + homeless
+ dose_day +

5       as.factor(adhere3) + as.fac-
tor(hivrsk) + cohort2,

6       data = ds, random = ~time2 | patid)
```

In line 1, we load the "nlme" (non-linear mixed effects) library. In line 2, we use the command "lme" to fit an LME model with response variable "logrna", with output object "lmefit". Lines 2–5 specify the model's fixed-effects covariates, with categorical variables designated using "as.factor". In line 6, we specify the analysis dataset, and indicate that subject's slopes over "time2" (a continuous version of time) are to be random. To indicate the subject clustering, we specify the subject ID number variable, as the grouping factor, using the code "| patid".

SAS

```
7   proc mixed data = ds;

8       class patid time hivrsk race
adhere3;

9       model logrna = time drkhaz2 pct3d
time*drkhaz2

10          drkhaz2*pct3d age race female
homeless dose_day

11          adhere3 hivrsk cohort2/s;

12       random int time2/type = un subject
= patid s g;

13 run;
```

In line 7, we call SAS PROC MIXED, applying it to a dataset called "ds". In line 8, we specify the categorical variables. In lines 9–11, we use the "model" statement to specify our response variable and fixed effects, with "/s requesting the regression solution. In line 12, we specify a random intercept and slope model using a continuous time variable ("time2") and that the data is clustered by the subject ID variable "patid", with an unstructured working covariance matrix for the random effects parameters.

Stata

```
14 xi: xtmixed logrna i.time drkhaz2 pct3d
i.drkhaz2*i.time

15      i.drkhaz2*pct3d age i.race female
homeless

16      dose_day i.adhere3 i.hivrsk
cohort2 || patid: time2,

17      covariance(unstructured)
```

The "xi" command allows the dynamic creation of categorical variables as well as interactions for the "xtmixed" command. The clustering is indicated by the "|| patid:" command, with "time2" given as the continuous measure of time and unstructured working covariance matrix for the random effects parameters.

Acknowledgements

This research was supported by National Institutes on Alcohol and Alcoholism grants RO1-AA013216, RO1-AA11785, RO1-AA10870, K24-AA015674 and the Smith College Committee on Faculty Development. We thank the reviewers as well as Seville Meli, Richard Saitz and Kristin Tyler for helpful comments on an earlier draft.

References

1. **Physicians' guide to helping patients with alcohol problems** [http://www.niaaa.nih.gov/Publications/EducationTrainingMaterials/guide.htm]
2. **AIDS Epidemic Update – United Nations Programme on HIV/AIDS** [http://www.unaids.org]
3. Samet JH, Phillips SJ, Horton NJ, Traphagen ET, Freedberg KA: **Detecting Alcohol Problems in HIV-Infected Patients: Use of the CAGE Questionnaire.** *AIDS Research and Human Retroviruses* 2004, **20:**151-155.
4. Dingle GA, Oei TP: **Is alcohol a cofactor of HIV and AIDS? Evidence from immunological and behavioral studies.** *Psychological Bulletin* 1997, **122:**56-71.
5. Samet JH, Horton NJ, Traphagen ET, Lyon SM, Freedberg KA: **Alcohol Consumption and HIV Disease Progression: Are They Related?** *Alcoholism: Clinical and Experimental Research* 2003, **27:**862-867.
6. Samet JH, Cheng DM, Libman H, Nunes D, Meli S, Tibbetts N, Saitz R: **Alcohol consumption may affect HIV disease progression.** *Poster presentation, XV International AIDS Conference, Bangkok Thailand* 2004.
7. Cannon MJ, Warner L, Taddei JA, Kleinbaum DG: **What can go wrong when you assume that correlated data are independent: an illustration from the evaluation of a childhood health intervention in Brazil.** *Statistics in Medicine* 2001, **20:**1461-1467.
8. Diggle PJ, Heagerty P, Liang KY, Zeger SL: *Analysis of Longitudinal Data* Oxford University Press; 2002.
9. Fitzmaurice GM, Laird NM, Ware JH: *Applied longitudinal analysis* John Wiley & Sons, New York; 2004.
10. McCulloch CE: **Repeated measures ANOVA, RIP?** *Chance* 2005, **18:**29-33.
11. Laird NM, Ware JH: **Random-effects models for longitudinal data.** *Biometrics* 1982, **38:**963-974.
12. Ehrenstein V, Horton NJ, Samet JH: **Inconsistent condom use among HIV-infected patients with alcohol problems.** *Drug and Alcohol Dependence* 2004, **73:**159-166.
13. Samet JH, Horton NJ, Meli S, Freedberg KA, Palepu A: **Alcohol Consumption and Antiretroviral adherence among HIV-infected persons with alcohol problems.** *Alcoholism: Clinical and Experimental Research* 2004, **28:**572-577.
14. Palepu A, Horton NJ, Tibbetts N, Dukes K, Meli S, Samet JH: **Substance abuse treatment and emergency department utilization among a cohort of HIV-infected persons with alcohol problems.** *Journal of Substance Abuse Treatment* 2003, **25:**37-42.
15. Ewing JA: **Detecting alcoholism: the CAGE questionnaire.** *Journal of the American Medical Association* 1984, **252:**1905-1907.

16. Folstein MF, Folstein SE, McHugh PR: **Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician.** *Journal of Psychiatric Research* 1975, **12:**189-198.
17. Pachl C, Todd JA, Kern DG, Shereidan PJ, Fong SJ, Stempien M, Hoo B, Besemer D, Yeghiazarian T, Irvine B: **Rapid and precise quantification of HIV-1 RNA in plasma using a branched DNA signal amplification assay.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1995, **8(5):**446-454.
18. James Hughes P: **Mixed Effects Models with Censored Data with Application to HIV RNA Levels.** *Biometrics* 1999, **55:**625-629.
19. Gibb KA, Bing EG, Fleishman JA, London AS, Caetano R, Burnam MA, Longshore D, Morton SC, Orlando M, Shapiro M: **Accuracy and usefulness of a breath analyzer.** *Annals of Emergency Medicine* 1984, **13:**516-520.
20. McLennan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, Pettinati H, Argeriou M: **The fifth edition of the Addiction Severity Index.** *Journal of Substance Abuse Treatment* 1992, **9:**199-213.
21. Chesney MA, Ickovics JR, Chambers DB, Gifford AL, Neidig J, Zwickl B, Wu AW: **Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: the AACTG Adherence instruments.** *AIDS Care* 2000, **12:**255-266.
22. Samet JH, Horton NJ, Meli S, Dukes K, Tripp T, Sullivan L, Freedberg KA: **A randomized controlled trial to enhance antiretroviral therapy adherence in patients with a history of alcohol problems.** *Antiretroviral Therapy* 2005, **10:**83-93.
23. Collins LM, Schafer JL, Kam C: **A comparison of inclusive and restrictive strategies in modern missing data procedures.** *Psychological Methods* 2001, **6:**330-351.
24. Little RJA, Rubin DB: *Statistical analysis with missing data* 2nd edition. John Wiley & Sons; 2002.
25. Diggle P, Kenward MG: **Informative drop-out in longitudinal data analysis.** *Applied Statistics* 1994, **43:**49-73.
26. Greenland S: **Multiple bias modeling for analysis of observational data (with discussion).** *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2005, **168:**267-306.
27. Liang K-Y, Zeger SL: **Longitudinal data analysis using generalized linear models.** *Biometrika* 1986, **73:**13-22.
28. Best NG, Spiegelhalter DJ, Thomas A, Brayne CEG: **Bayesian analysis of realistically complex models.** *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 1996, **159:**323-342.
29. Skrondal A, Rabe-Hesketh S: *Generalized latent variable modeling: multilevel, longitudinal and structural equation models* Boca Raton: Chapman and Hall/CRC; 2004.

The association between hepatitis C infection and prevalent cardiovascular disease among HIV-infected individuals

Matthew S. Freiberg[a,b], Debbie M. Cheng[c,d], Kevin L. Kraemer[a], Richard Saitz[c,f], Lewis H. Kuller[b] and Jeffrey H. Samet[c,e]

Objective: To examine the association between hepatitis C and prevalent cardiovascular disease (CVD) among HIV-infected individuals.

Design: A cross-sectional analysis of data from the HIV–Longitudinal Interrelationships of Viruses and Ethanol (HIV–LIVE) cohort, a prospective cohort of HIV-infected individuals with current or past alcohol problems.

Methods: We analysed health questionnaire and laboratory data from 395 HIV-infected individuals (50.1% co-infected with hepatitis C) using logistic regression to estimate the odds ratio (OR) for the prevalence of CVD among those co-infected with hepatitis C and HIV compared with those infected with HIV alone.

Results: The prevalence of CVD was higher among those co-infected with hepatitis C compared with those with HIV alone (11.1 versus 2.5%, respectively). After adjusting for age, the OR for the prevalence of CVD was significantly higher among those with hepatitis C co-infection (adjusted OR 4.65, 95% confidence interval 1.70–12.71). The relationship between hepatitis C and CVD persisted when adjusting for age and other sociodemographic characteristics, substance use, and cardiovascular risk factors in separate regression models.

Conclusion: Co-infection with hepatitis C among a cohort of HIV-infected individuals was associated with a higher age-adjusted odds for the prevalence of CVD. These data suggest that hepatitis C infection may be associated with an increased risk of CVD among those co-infected with HIV.

AIDS 2007, **21**:193–197

Keywords: Cardiovascular disease, cardiovascular risk factors, hepatitis C, HIV, myocardial infarction, substance abuse

From the [a]Division of General Internal Medicine, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA, the [b]Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA, the [c]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA, the [d]Departments of Biostatistics, and [e]Social and Behavioral Sciences, and the [f]Youth Alcohol Prevention Center and Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA.

Correspondence to Matthew S. Freiberg, MD, MSc, Departments of Medicine and Epidemiology, University of Pittsburgh, 230 McKee Place, Suite 600, Pittsburgh, PA 15213, USA.

E-mail: freibergms@upmc.edu

Received: 25 July 2006; revised: 6 September 2006; accepted: 4 October 2006.

ISSN 0269-9370 © 2007 Lippincott Williams & Wilkins

Introduction

Whether certain chronic infections are a risk factor for cardiovascular disease (CVD) is uncertain. Studies suggest that there is a link between some infectious pathogens and atherosclerosis [1,2]. The mechanism for the development of atherosclerosis may be colonization of the vasculature by the pathogen [3] or stimulation of the inflammatory cascade [4]. Some studies have reported a higher prevalence of cardiovascular risk factors [5,6], carotid atherosclerosis [7–9] and coronary heart disease [10] among individuals with hepatitis C infection. In contrast, other studies found no such association [11–13], but were limited by the number of hepatitis C-infected participants [11] and the lack of hepatitis C viral RNA [11–13] and substance use data. The absence of information on injection drug, cocaine and alcohol use could confound the relationship between hepatitis C and CVD [11,12].

With the advent of antiretroviral therapy and prolonged survival, coronary heart disease is now common among individuals with HIV infection [14]. Of those infected with HIV, studies have reported co-infection with hepatitis C ranging from 16.1% [15] up to 30% [16]. Compared with individuals infected only with HIV or hepatitis C, those who are co-infected may be at a higher risk of atherosclerosis and CVD. Therefore, we investigated the association between hepatitis C and CVD in a cohort of individuals with HIV infection and substance abuse.

Methods

We analysed baseline data collected from 395 HIV-infected participants in the HIV–Longitudinal Interrelationships of Viruses and Ethanol (HIV–LIVE) study, a prospective cohort of individuals with HIV infection and current or past alcohol problems. During August 2001 to July 2003, 400 HIV–LIVE participants were enrolled from four different sources: (i) a previous cohort of HIV-infected participants with alcohol problems; (ii) Boston Medical Center's Diagnostic Evaluation Unit, an intake clinic for HIV-infected patients; (iii) Beth Israel Deaconess Medical Center primary care and specialty clinics; and (iv) other local healthcare sites or shelters in the Boston area. Inclusion criteria included: (i) an HIV antibody test positive by enzyme-linked immunosorbent assay and confirmed by Western blot; (ii) two or more affirmative responses to the CAGE alcohol screening questionnaire [17] or by physician-investigator diagnosis of alcoholism; (iii) ability to speak English or Spanish; and (iv) at least one contact person who was likely to know the subject's whereabouts. Participants were excluded if the 30-item Folstein Mini-Mental State Examination score [18] was less than 21 or a trained interviewer assessed that the patient was incapable of comprehending the informed consent or of answering interview questions. For the present cross-sectional analysis, five participants were excluded because of missing hepatitis C RNA data ($n=4$) or a history of CVD data ($n=1$). Additional study design details have been published elsewhere [19]. The Institutional Review Boards of the Boston Medical Center, Beth Israel Deaconess Medical Center, and the University of Pittsburgh approved the study.

The primary outcome variable was self-reported CVD. CVD was defined as a 'yes' response to one of the three following questions: 'Has a doctor ever told you that you had...' (1) peripheral vascular disease (hardening of the arteries in your neck or legs, atherosclerosis); (2) a stroke, cerebrovascular accident, blood clot or bleeding in the brain, or transient ischemic attack; or (3) a heart attack (myocardial infarction). Our secondary outcome was myocardial infarction. Hepatitis C infection, defined as detectable hepatitis C viral RNA (either from serum collected at the time of enrollment or from participants' medical records), was the main independent variable. Additional co-variates included the following: age; race (white versus non-white); sex; current CD4 cell count (>200 cells/μl); obesity (body mass index $\geq 30\,kg/m^2$); current adherent antiretroviral therapy use defined as 100% adherence to all antiretroviral medications over the preceding 3 days before enrollment in the study; and self-reported health conditions defined as a 'yes' response to 'Has a doctor ever told you that you had...' diabetes or high blood sugar or sugar, hypertension or high blood pressure, high cholesterol, renal disease (poor kidney function or blood tests showing high creatinine), anemia (low red blood cell count, hemoglobin), or cirrhosis of the liver. For substance use variables, we defined current smoking as a 'yes' response to 'Do you currently smoke cigarettes every day or on some days?'; current at-risk alcohol consumption as drinking more than 14 standard drinks for men (more than seven for women) per week or more than four drinks on one occasion for men (more than three drinks for women) in the past 30 days; cocaine use as self-reported use of 'cocaine, crack or free base'; and injection as a 'yes' response to 'In your lifetime, have you ever injected drugs?'. Homelessness was defined as any night on the street or in a shelter in the past 6 months. We defined lipodystrophy as a 'yes' response to 'Since you began taking HIV medications, have you noticed any of the following changes: more fat in the back of your neck or increase in your waist, 'butt' or chest out of proportion to your arms or legs?' [20].

Statistical analyses

We obtained descriptive statistics for all variables and assessed the relationship between hepatitis C and CVD and other co-variates using *t*-tests for continuous variables and chi-square analysis or Fisher's exact tests for categorical

variables. We used logistic regression models to estimate the odds ratio (OR) for the prevalence of CVD and myocardial infarction (separate analyses for each outcome) using hepatitis C infection as the main independent variable while adjusting for age. Additional analyses adjusted for age and then each of the covariates separately. The limited number of self-reported CVD events precluded models that adjusted for all covariates simultaneously. When hepatitis C-positive and hepatitis C-negative participants were missing data on body mass index ($n=10$ and $n=9$, respectively), CD4 cell count ($n=9$ and $n=13$, respectively), HIV viral RNA ($n=16$ and $n=20$, respectively), and a history of lipodystrophy ($n=6$ and $n=1$, respectively) or when hepatitis C-negative participants were missing data on a history of injection drug use ($n=1$) and current at-risk alcohol consumption ($n=1$), we excluded those participants from the analyses adjusted for those factors. As previous studies reported an association between hepatitis C and diabetes, we performed secondary analyses restricting the sample to those who reported no history of diabetes.

Results

Compared with the 197 HIV-infected participants who were not infected with hepatitis C, those with hepatitis C co-infection ($n=198$) were significantly older (43.9 versus 40.9 years), had a higher prevalence of ever having had diabetes (10.1 versus 4.1%), cirrhosis (10.6 versus 2.5%), myocardial infarctions (6.5 versus 0.5%), and CVD (11.1 versus 2.5%), and a lower prevalence of hypercholesterolemia (19.7 versus 33.0%, $P<0.05$ for all; Table 1). The unadjusted OR for the prevalence of CVD and myocardial infarction were significantly higher among those with hepatitis C co-infection compared with those who were infected only with HIV [OR 4.80; 95% confidence interval (CI) 1.78–12.95 and OR 13.77; 95% CI 1.78–106.33, respectively]. After adjusting for age, the association between hepatitis C and both CVD and myocardial infarction persisted (OR 4.65; 95% CI 1.70–12.71 and OR 12.86; 95% CI 1.65–100.33, respectively; Table 2). Furthermore, after each covariate was added separately to the age-adjusted model, the

Table 1. Characteristics of 395 individuals with HIV and alcohol problems with and without hepatitis C co-infection.

Subject characteristics	Undetectable hepatitis C viral RNA ($N=197$)	Detectable hepatitis C viral RNA ($N=198$)
Age (median in years)	40.9	43.9*
Female, *n* (%)	46 (23.3)	54 (27.3)
Race, *n* (%)		
White	63 (32.0)	68 (34.0)
Non-white	134 (68.0)	130 (66.0)
Prevalent CVD, *n* (%)	5 (2.5)	22 (11.1)*
Ever had a heart attack (myocardial infarction), *n* (%)	1 (0.5)	13 (6.5)*
Ever had peripheral vascular disease (hardening of the arteries in your neck or legs; atherosclerosis), *n* (%)	0 (0.0)	4 (2.0)
Ever had a stroke, cerebrovascular accident, blood clot or bleeding in the brain, or transient ischemic attack), *n* (%)	4 (2.0)	8(4.0)
Ever had hypertension, *n* (%)	47 (23.9)	55 (27.9)
Ever had high cholesterol, *n* (%)	65 (33.0)	39 (19.7)*
Ever had diabetes, *n* (%)	8 (4.1)	20 (10.1)*
Current smoker, *n* (%)	145 (73.6)	157 (79.3)
Current at-risk alcohol consumption, *n* (%)[a]	66 (33.7)	58 (29.3)
Ever cocaine use, *n* (%)	84 (42.6)	91 (46.0)
Ever injection drug use, *n* (%)[b]	40 (20.4)	176 (88.9)*
BMI $\geq 30\,kg/m^2$, *n* (%)[c]	39 (20.9)	37 (19.6)
Current CD4 cell count >200 cells/μl, *n* (%)[d]	155 (82.5)	144 (77.8)
HIV viral RNA undetectable, *n* (%)[e]	59 (32.6)	52 (29.2)
Current adherent antiretroviral medication, *n* (%)[f]	86 (43.6)	99 (50.0)
Ever had lipodystrophy, *n* (%)[g]	49 (25.7)	56 (28.4)
Ever had cirrhosis, *n* (%)	5 (2.5)	21 (10.6)*
Ever anemic, *n* (%)	47 (23.9)	51 (25.8)
Ever had kidney disease, *n* (%)	7 (3.6)	17 (8.6)
Ever homeless, *n* (%)	40 (20.3)	56 (28.3)

BMI, Body mass index; CVD, cardiovascular disease.
[a]$N=196$ hepatitis C-negative participants and $n=198$ hepatitis C-positive participants who had current alcohol consumption assessed.
[b]$N=196$ hepatitis C-negative participants and $n=198$ hepatitis C-positive participants who had history of injection drug use assessed.
[c]$N=187$ hepatitis C-negative participants and $n=189$ hepatitis C-positive participants who had BMI measurement.
[d]$N=188$ hepatitis C-negative participants and $n=185$ hepatitis C-positive participants who had CD4 cell count measurement.
[e]$N=181$ hepatitis C-negative participants and $n=178$ hepatitis C-positive participants who had HIV-RNA measurement.
[f]$N=185$ participants (46.8%) were on antiretroviral therapy and were 100% adherent in the past 3 days; 60 participants (15.2%) were on antiretroviral therapy and were not 100% adherent; 150 (38%) were not on antiretroviral therapy.
[g]$N=191$ hepatitis C-negative participants and $n=197$ hepatitis C-positive participants who answered the lipodystrophy question.
*$P<0.05$.

Table 2. Age-adjusted associations between hepatitis C infection and cardiovascular disease among 395 individuals with HIV infection and alcohol problems.

Adjustment variable (all are adjusted for age)	Adjusted OR (95% CI) for association between hepatitis C and CVD[a]
Age only	4.65 (1.70,12.71)
Female	4.66 (1.70,12.74)
Race (white versus non-white)	4.68 (1.70, 12.83)
Ever had hypertension	4.69 (1.70, 12.89)
Ever had high cholesterol	5.44 (1.96, 15.11)
Ever had diabetes	4.52 (1.65, 12.40)
Current smoker	4.86 (1.77, 13.36)
Current at-risk alcohol consumption[a]	4.65 (1.70,12.72)
Ever cocaine use	4.66 (1.70, 12.74)
Ever injection drug use[b]	5.53 (1.99,15.33)
BMI $\geq 30\,kg/m^2$[c]	4.00 (1.44, 11.14)
Current CD4 cell count >200 cells/μl[d]	4.51 (1.64, 12.39)
HIV viral RNA (detectable versus undetectable)[e]	4.61 (1.67,12.71)
Current adherent antiretroviral medication[f]	4.65 (1.70, 12.71)
Ever had lipodystrophy[g]	4.56 (1.66, 12.51)
Ever had cirrhosis	4.91 (1.80, 13.43)
Ever anemic	4.66 (1.70, 12.79)
Ever had kidney disease	4.80 (1.76, 13.13)
Ever homeless	4.67 (1.70, 12.81)

BMI, Body mass index; CVD, cardiovascular disease; CI, confidence interval; OR, odds ratio.
[a]$N=196$ hepatitis C-negative participants and $n=198$ hepatitis C-positive participants who had current alcohol consumption assessed.
[b]$N=196$ hepatitis C-negative participants and $n=198$ hepatitis C-positive participants who had history of injection drug use assessed.
[c]$N=187$ hepatitis C-negative participants and $n=189$ hepatitis C-positive participants who had BMI measurement.
[d]$N=188$ hepatitis C-negative participants and $n=185$ hepatitis C-positive participants who had CD4 cell count measurement.
[e]$N=181$ hepatitis C-negative participants and $n=178$ hepatitis C-positive participants who had HIV-RNA measurement.
[f]$N=185$ participants (46.8%) were on antiretroviral therapy and were 100% adherent in the past 3 days; 60 participants (15.2%) were on antiretroviral therapy and were not 100% adherent; 150 (38%) were not on antiretroviral therapy.
[g]$N=191$ hepatitis C-negative participants and $n=197$ hepatitis C-positive participants who answered the lipodystrophy question.
Referent group is HIV-infected hepatitis C RNA-negative participants.

relationship between hepatitis C co-infection and CVD remained unchanged (Table 2). In participants without diabetes, the association between hepatitis C and CVD was similar (OR 4.11; 95% CI 1.47–11.49) to analyses that included participants with diabetes.

Discussion

In the HIV–LIVE cohort, those co-infected with hepatitis C and HIV had a higher unadjusted and age-adjusted prevalence of CVD than those who were infected with HIV alone. When individual confounders were added separately to the age-adjusted models, the relationship between hepatitis C and CVD remained unchanged. Analyses excluding participants with diabetes yielded similar results.

Previous studies describing the relationship between hepatitis C and CVD among HIV-infected individuals are sparse and their results have been inconsistent. In the HIV Epidemiologic Research Study, hepatitis C antibody-positive women with HIV infection had higher rates of hospitalizations for CVD conditions compared with those only infected with HIV [21]. In contrast, among HIV/hepatitis C co-infected participants of the HIV Outpatient Study, the age-adjusted prevalence of CVD was not significantly higher compared with those infected with HIV alone [22]. The association between HIV and hepatitis C co-infection and CVD, however, may have varied between the studies because the definition of CVD used was not the same.

Among individuals infected with hepatitis C, but not HIV, our findings are consistent with previous studies reporting an increased prevalence of carotid atherosclerosis [7–9] and prevalent CVD [10]. Whereas some studies reported no significant association between hepatitis C and atherosclerosis, those studies were limited by sparse data on substance abuse habits, particularly injection drug use [13], and no data describing the risks of co-infection with HIV and hepatitis C, or analyses of hepatitis C viral RNA data [11–13]. Data on hepatitis C viral RNA is valuable because it can reduce the misclassification that might occur as a result of the approximately 20% of hepatitis C antibody-positive individuals who spontaneously clear the virus [23]. Given the association between hepatitis C infection and diabetes, it is possible that the effect of hepatitis C on cardiovascular risk might be mediated solely through diabetes [5]. However, our analyses, which adjusted for a history of diabetes and also restricted the sample to individuals without diabetes, suggest that diabetes is not the mechanism responsible for this association among HIV-infected individuals.

Our study has several limitations that merit comment. First, the outcome variable and several covariates were obtained by self-report, which could have resulted in some non-differential misclassification. However, our self-reported variables that were traditional CVD risk factors (e.g. hypertension and cholesterol) were associated with an increased prevalence of self-reported CVD. Second, as our definition of lipodystrophy included fat accumulation in the buttock region, there is the possibility that some non-differential misclassification occurred because lipodystrophy among HIV-infected individuals typically results in fat loss, rather than accumulation in the buttock region. Third, it is possible that differential misclassification might have occurred among those infected with hepatitis C because they might have had more contact with the healthcare system and consequently had more opportunities to be told of other health conditions (e.g. CVD). However, this potential misclassification was probably minimized by the fact that all the participants had several important health problems (e.g. HIV and alcohol) requiring frequent use of the

healthcare system. Furthermore, when hepatitis C might have resulted in additional clinic visits (e.g. hepatology clinic), this patient–physician encounter would probably not have focused on CVD issues. Fourth, as there were a small number of self-reported cardiovascular events, our ability to adjust for confounders in the multivariable models was limited. Finally, as all the participants in this cohort had alcohol problems, the findings may not be generalizable to the HIV-infected population without alcohol problems.

Nevertheless, this study suggests that among HIV-infected individuals, co-infection with hepatitis C may be independently associated with an increased risk of CVD. This finding, if confirmed in other studies, has important implications for the prevention and treatment of CVD for the substantial number of individuals around the world who are co-infected with HIV and hepatitis C.

Acknowledgements

The authors gratefully acknowledge the contributions of Howard Libman, MD, from Beth Israel Deaconess Medical Center and David Nunes, MD, from Boston Medical Center for their contributions to this study.

Sponsorship: Support for this study came from the following National Institute of Alcohol Abuse and Alcoholism grants: R01-AA13216 (Clinical Impact of HCV and Alcohol in HIV-Infected Persons); R01-AA11785 (Medication Adherence in Alcohol Abusing HIV Patients); RO1-AA10870 (Enhanced Linkage of Alcohol Abusers to Primary Care); K23 AA015914 (Alcohol and Coronary Heart Disease in People with HIV); and K24 AA015674 (Impact of Alcohol use on HIV Infection – in USA and Russia). This research was conducted partly in the General Clinical Research Center at Boston University School of Medicine, USPHS grant M01 RR00533, and the Clinical Research Center at Beth Israel Deaconess Medical Center, USPHS grant MO1 RR01032.

References

1. Siscovick DS, Schwartz SM, Corey L, Grayston JT, Ashley R, Wang SP, *et al.* ***Chlamydia pneumoniae*, herpes simplex virus type 1, and cytomegalovirus and incident myocardial infarction and coronary heart disease death in older adults: the Cardiovascular Health Study.** *Circulation* 2000; **102**:2335–2340.
2. Shah PK. **Link between infection and atherosclerosis: who are the culprits: viruses, bacteria, both, or neither?** *Circulation* 2001; **103**:5–6.
3. Chiu B, Viira E, Tucker W, Fong IW. ***Chlamydia pneumoniae*, cytomegalovirus, and herpes simplex virus in atherosclerosis of the carotid artery.** *Circulation* 1997; **96**:2144–2148.
4. Tillmann HL, Manns MP, Rudolph KL. **Merging models of hepatitis C virus pathogenesis.** *Semin Liver Dis* 2005; **25**:84–92.
5. Mehta SH, Brancati FL, Sulkowski MS, Strathdee SA, Szklo M, Thomas DL. **Prevalence of type 2 diabetes mellitus among persons with hepatitis C virus infection in the United States.** *Ann Intern Med* 2000; **33**:592–599.
6. 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**:401–403.
7. 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**:133–135.
8. Fukui M, Kitagawa Y, Nakamura N, Yoshikawa T. **Hepatitis C virus and atherosclerosis in patients with type 2 diabetes.** *JAMA* 2003; **289**:1245–1246.
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. Vassalle C, Masini S, Bianchi F, Zucchelli GC. **Evidence for association between hepatitis C virus seropositivity and coronary artery disease.** *Heart* 2004; **90**:565–566.
11. 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**:211–213.
12. 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.
13. Butt AA, Evans R, Skanderson M, Shakil AO. **Comorbid medical and psychiatric conditions and substance abuse in HCV infected persons on dialysis.** *J Hepatol* 2006; **44**:864–868.
14. Friis-Moller N, Sabin CA, Weber R, d'Arminio Monforte A, El-Sadr WM, Reiss P, *et al.* **Combination antiretroviral therapy and the risk of myocardial infarction.** *N Engl J Med* 2003; **349**:1993–2003.
15. Sherman KE, Rouster SD, Chung RT, Rajicic N. **Hepatitis C virus prevalence among patients infected with human immunodeficiency virus: a cross-sectional analysis of the US adult AIDS Clinical Trials Group.** *Clin Infect Dis* 2002; **34**:831–837.
16. Anderson KB, Guest JL, Rimland D. **Hepatitis C virus coinfection increases mortality in HIV-infected patients in the highly active antiretroviral therapy era: data from the HIV Atlanta VA Cohort Study.** *Clin Infect Dis* 2004; **39**:1507–1513.
17. 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**:774–777.
18. Folstein MF, Folstein SE, McHugh PR. **"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.** *J Psychiatr Res* 1975; **12**:189–198.
19. Libman H, Saitz R, Nunes D, Cheng DM, Richardson JM, Vidaver J, *et al.* **Hepatitis C infection is associated with depressive symptoms in HIV-infected adults with alcohol problems.** *Am J Gastroenterol* 2006; **101**:1804–1810.
20. Carr A, Emery S, Law M, Puls R, Lundgren JD, Powderly WG, *et al.* **An objective case definition of lipodystrophy in HIV-infected adults: a case-control study.** *Lancet* 2003; **361**:726–735.
21. Gardner LI, Klein RS, Szczech LA, Phelps RM, Tashima K, Rompalo AM, *et al.* **Rates and risk factors for condition-specific hospitalizations in HIV-infected and uninfected women.** *J Acquir Immune Defic Syndr* 2003; **34**:320–330.
22. Tedaldi EM, Baker RK, Moorman AC, Alzola CF, Furhrer J, McCabe RE, *et al.* **Influence of coinfection with hepatitis C virus on morbidity and mortality due to human immunodeficiency virus infection in the era of highly active antiretroviral therapy [see Comment].** *Clin Infect Dis* 2003; **36**:363–367.
23. Liang TJ, Rehermann B, Seeff LB, Hoofnagle JH. **Pathogenesis, natural history, treatment, and prevention of hepatitis C.** *Ann Intern Med* 2000; **132**:296–305.

is done. For instance, in Kuhl and colleagues' study, one in eight women had findings that demanded biopsy. These results would not be expected in the general population. Despite the high recall rates, 52% of biopsy specimens in Kuhl's study showed breast cancer, a figure that is also accepted in screening mammography. In MRI screening of high-risk patients,[12] MRI also doubled the recall rate but the rate of detected lesions per biopsy did not change—one in three biopsies was positive for cancer.

These findings can only lead to the conclusion that MRI outperforms mammography in tumour detection and diagnosis. MRI should thus no longer be regarded as an adjunct to mammography but as a distinct method to detect breast cancer in its earliest stage. A large multicentre breast-screening trial with MRI in the general population is essential.

*Carla Boetes, Ritse M Mann
Radboud University Nijmegen Medical Centre, Department of Radiology, 6500 HB Nijmegen, Netherlands
c.boetes@rad.umcn.nl

We declare that we have no conflict of interest.

1 Kuhl CK, Schrading S, Bieling HB, et al. MRI for diagnosis of pure ductal carcinoma-in-situ: a prospective observational study. *Lancet* 2007; **370:** 485–92.
2 Welch HG, Black WC. Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find? *Ann Intern Med* 1997; **127:** 1023–28.
3 Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: a review. *Breast Cancer Res Treat* 2006; **97:** 135–44.
4 Leonard GD, Swain SM. Ductal carcinoma in situ, complexities and challenges. *J Natl Cancer Inst* 2004; **96:** 906–20.
5 Burstein HJ, Polyak K, Wong JS, Lester SC, Kaelin CM. Ductal carcinoma in situ of the breast. *N Engl J Med* 2004; **350:** 1430–41.
6 Ikeda DM, Andersson I. Ductal carcinoma in situ: atypical mammographic appearances. *Radiology* 1989; **172:** 661–66.
7 Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 2005; **353:** 1773–83.
8 Cao Y, Paner GP, Kahn LB, Rajan PB. Noninvasive carcinoma of the breast: angiogenesis and cell proliferation. *Arch Pathol Lab Med* 2004; **128:** 893–96.
9 Berg WA, Gutierrez L, NessAiver MS, et al. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology* 2004; **233:** 830–49.
10 Lehman CD, Gatsonis C, Kuhl CK, et al. MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. *N Engl J Med* 2007; **356:** 1295–303.
11 Pediconi F, Catalano C, Roselli A, et al. Contrast-enhanced MR mammography for evaluation of the contralateral breast in patients with diagnosed unilateral breast cancer or high-risk lesions. *Radiology* 2007; **243:** 670–80.
12 Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007; **57:** 75–89.

Stroke prevention in elderly patients with atrial fibrillation

See **Articles** page 493

Warfarin is highly effective in the prevention of stroke in atrial fibrillation, with a 64% risk reduction compared with 22% for aspirin.[1] However, there are concerns about whether this benefit can be extrapolated to older age-groups because few patients older than 75 years were enrolled in early clinical trials. Older patients are at the highest risk of stroke, but old age has been identified as an independent risk factor for warfarin-associated haemorrhage.

In today's *Lancet*, Jonathan Mant and colleagues report the results of the Birmingham Atrial Fibrillation Treatment of the Aged (BAFTA) study, in which 973 patients, age 75 years and older and with atrial fibrillation, were randomised to warfarin (target international normalised ratio 2–3) or aspirin 75 mg.[2] The primary endpoint was fatal or disabling strokes (ischaemic and haemorrhagic) or clinically significant arterial embolism. Warfarin was superior to aspirin in the prevention of stroke (1·8% *vs* 3·8% per year) and was no more hazardous than aspirin in terms of major haemorrhage (1·9% *vs* 2·0% per year).

To what extent can the BAFTA results be applied to a real-world population of elderly patients with atrial fibrillation? To answer this question, we must consider the possible influence that selection bias may have had in BAFTA. Once enrolled in BAFTA, warfarin-eligible patients had a 50% chance of receiving aspirin, which is less effective than warfarin at preventing stroke. Because it would be unethical to deny warfarin to patients for whom it is clearly indicated, participation in BAFTA was restricted to patients for whom there was

	BAFTA[2]	NRAF[4]	SPORTIF III[5]	SPORTIF V[6]	AFI[7]	ACTIVE-W[8]
Heart failure (%)	20	56	34	40	22	31
Hypertension (%)	53	56	72	81	46	82
Mean age (years)	82	81	70	72	69	70
Diabetes (%)	14	23	22	25	15	21
Previous stroke (%)	13	25	24	18	17	15
New to warfarin (%)	60	NA	27	15	100	22

NA=not applicable (observational study, no warfarin treatment).

Table: **Stroke risk factors**

clinical uncertainty about which of the two treatments should be used. From a physician's perspective, this eligibility requirement probably encouraged the recruitment of patients at a lower risk of stroke. For example, recently revised guidelines recommend either aspirin or warfarin (depending on the individual physician's and patient's preference) for patients with a CHADS2 score of 1 (CHADS2 scores patients with atrial fibrillation to determine their risk of stroke. For a patient older than 75 years, this would mean he or she had no other stroke risk factors, such as hypertension, diabetes, history of stroke, or congestive heart failure).[3]

Therefore, compared with other study populations, participants in BAFTA had a lower prevalence of risk factors for stroke (table). It is noteworthy that just over a fifth of the 4639 patients identified were enrolled, and 1570 patients were excluded because warfarin was the only appropriate treatment for these patients. The low prevalence of risk factors for stroke might, in addition to good control of blood pressure and the large proportion of patients already taking a vitamin K antagonist at study entry, explain the lower than anticipated rate of thrombotic events in BAFTA. Nevertheless, the fact that BAFTA showed that warfarin is more effective than aspirin, even in this relatively low-risk group of patients, adds to other evidence[8] that, in patients with atrial fibrillation, anticoagulation protects patients against stroke more effectively than antiplatelet therapy.

We agree with Mant and colleagues that the lack of difference in major haemorrhage between the two groups is surprising. Additionally, the rates of haemorrhage in BAFTA were significantly lower than rates in another study of elderly patients treated with warfarin.[9] How might we explain these observations? First, just over four-fifths of the patients in BAFTA were taking warfarin or aspirin before enrolment, which means that BAFTA selected a group of individuals who had already survived exposure to antithrombotic therapy. Second, we presume that combination therapy with an antiplatelet and a vitamin K antagonist (eg, clopidogrel and warfarin), a known independent risk factor for warfarin-associated haemorrhage, was not permitted. Third, the number of potential patients initially invited for an electroencephalogram is not clear. Possibly, the individuals who responded to the invitations to participate represent an overall healthier population (ie, less likely to be recently discharged from hospital). Knowledge of the prevalence of risk factors for haemorrhage (including history of haemorrhage, renal disease, and anaemia) would allow a more informed interpretation of the bleeding rates.

Despite these considerations, BAFTA adds important new information for the care of elderly patients with atrial fibrillation. Mant and colleagues enrolled an unprecedented number of patients in an age group that has been largely under-represented in randomised trials. BAFTA firmly establishes the superior efficacy of warfarin as a stroke-prevention strategy in elderly patients with atrial fibrillation. However, in the future, our greatest challenge will be to identify those patients (elderly or not) who are truly at the highest risk of major bleeding, particularly intracranial haemorrhage. For everyone else, no matter the age group, the benefits of well-managed warfarin substantially outweigh its risks.

David Garcia, Elaine Hylek
Department of Internal Medicine, University of New Mexico, Albuquerque, NM 87131, USA (DG); and Department of Internal Medicine, Boston University, Boston, MA, USA (EH)
davgarcia@salud.unm.edu

DG and EH have received research funding from Bristol-Myers Squibb and acted as consultants for Bristol-Myers Squibb. EH has also received research funding from AstraZeneca.

1 Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med* 2007; **146:** 857–67.
2 Mant J, Hobbs FDR, Fletcher K, on behalf of the BAFTA investigators and the 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:** 493–503.
3 Fuster V, Ryden LE, Cannom DS, et al. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation—executive summary. *Circulation* 2006; **114:** 700–52.
4 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.
5 Executive Steering Committee on behalf of the SPORTIF III Investigators. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): randomised controlled trial. *Lancet* 2003; **362:** 1691–98.
6 Executive Steering Committee for the SPORTIF V Investigators. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation. *JAMA* 2005; **293:** 690–98.
7 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:** 1449–57.
8 The ACTIVE Writing Group on behalf of the ACTIVE Investigators. Clopidogrel plus aspirin versus oral anticoagulation for atrial fibrillation in the atrial fibrillation clopidogrel trial with Irbesartan for prevention of Vascular Events (ACTIVE W). *Lancet* 2006; **367:** 1903–12.
9 Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. *Circulation* 2007; **115:** 2689–96.

A risk model predicted major bleeding in older patients with atrial fibrillation receiving warfarin therapy

Shireman TI, Mahnken JD, Howard PA, *et al.* Development of a contemporary bleeding risk model for elderly warfarin recipients. *Chest* 2006;**130**:1390–6.

Clinical impact ratings GP/FP/Primary care ★★★★★★ IM/Ambulatory care ★★★★★★ Geriatrics ★★★★★★ Haematology ★★★★★★

Q In older patients with atrial fibrillation (AF) who are receiving warfarin therapy at hospital discharge, does a contemporary bleeding risk model predict major bleeding?

METHODS

Design: 2 cohort studies, 1 for derivation and 1 for validation, from the National Registry of Atrial Fibrillation.

Setting: USA.

Patients: 26 345 patients ≥65 years of age (88% >70 y, 43% >80 y, 53% women) (19 875 for derivation; 6470 for validation) with AF who were receiving warfarin therapy at hospital discharge. Exclusion criteria included discharge against medical advice, transfer to another acute care hospital, enrolment in managed care, or death during hospital admission.

Description of prediction guide: the risk score (range 0–4.17) categorised patients into low (score ≤1.07), moderate (score >1.07 to <2.19), or high risk (score ≥2.19) groups. Multivariate analysis of risk factors found 8 independent clinical variables that predicted major bleeding events. The risk score was a summation of the log likelihood of 8 clinical variables: age (≥70 y = 0.49), sex (women = 0.32), remote bleeding event = 0.58, recent bleeding event = 0.62, alcohol or drug abuse = 0.71, diabetes = 0.27, anaemia = 0.86, and receipt of antiplatelet drugs = 0.32.

Outcomes: hospital admission for major acute bleeding within 90 days of index hospital discharge.

MAIN RESULTS

8 independent clinical predictors of major bleeding were identified in the derivation cohort (table). In the validation cohort, major bleeding events developed in 35 (0.9%) low risk patients {likelihood ratio [LR] 0.6}*, 48 (2.0%) moderate risk patients {LR 1.4}*, and 12 (5.4%) high risk patients {LR 3.8}* ($p<0.001$). The area under the receiver operating characteristic (AUROC) curve was 0.632. This model had similar predictive characteristics compared with 2 previous bleeding risks models: Outpatient Bleeding Risk Index (AUROC curve 0.613) and Kuijer *et al*'s model (AUROC curve 0.503).

For correspondence: Dr T Shireman, University of Kansas Medical Center, Kansas City, KS, USA. tshireman@kumc.edu

Source of funding: American Heart Association.

Multivariate predictors of major bleeding from the derivation cohort of older patients with atrial fibrillation receiving warfarin therapy at hospital discharge*

Predictors	Hazard ratio (95% CI)
Age ≥70 y	1.6 (1.1 to 2.5)
Women	1.4 (1.1 to 1.7)
Remote bleeding	1.8 (1.4 to 2.4)
Recent bleeding	1.8 (1.4 to 2.4)
Alcohol or drug abuse	2.0 (1.1 to 3.8)
Diabetes	1.3 (1.0 to 1.7)
Anaemia	2.4 (1.8 to 3.2)
Receipt of antiplatelet drugs	1.4 (1.1 to 1.8)

*CI defined in glossary.

CONCLUSION

In older patients with atrial fibrillation who were receiving warfarin therapy at hospital discharge, a bleeding risk model predicted major bleeding.

*LR defined in glossary and calculated from data in article.

Abstract and commentary also appear in ACP Journal Club.

The risk of stroke in patients with AF can be greatly reduced with warfarin. However, many patients who would benefit from warfarin therapy are not receiving anticoagulation therapy. A model that accurately predicts an individual patient's risk of major bleeding would allow for weighing the risks and benefits of warfarin therapy with greater confidence. Such a model would also identify reversible risk factors that providers can act on to minimise bleeding risk (eg, avoidance of antiplatelet drugs).

Prevalence of AF and risk of warfarin associated bleeding both increase with advancing age. Previously published bleeding risk models included few patients >80 years of age.[1-2] The model derived and validated by Shireman *et al* adds to previous work in this area because >10 000 patients ≥80 years of age were included in the cohort that was studied.

The proposed model has drawbacks. Firstly, the equation is complex, and clinicians will not find it easy to use without an aid such as a calculator. Secondly, only 222 (3.4%) of the 6470 patients in the validation cohort were classified as high risk of bleeding by the present model. Even if the risk model was more user friendly, it is unclear whether clinicians would find such a model, in which >96% of patients are deemed to have ≤2% 90 day risk of major bleeding, helpful in making everyday decisions. Thirdly, 38 089 of 76 177 patients with AF were discharged without receiving warfarin. Many of them were not prescribed warfarin probably because physicians felt the risk of bleeding was excessive. If a substantial number of patients at risk of bleeding was thus excluded from both derivation and validation cohorts, can we generalise these results to an unselected population of AF patients? Finally, international normalised ratio (INR) values were not considered in the model. This limitation, shared by previously published clinical prediction guides, is important because bleeding risk is independently associated both with supratherapeutic INR values and with INR variability.[3]

The study by Shireman *et al* provides a step forward in assessing the bleeding risk of warfarin treated patients. However, the limitations of this study highlight the need for further work in this important area.

David Garcia, MD
Albuquerque, New Mexico, USA
Elaine Hylek, MD, MPH
Boston University
Boston, Massachusetts, USA

1 Beyth RJ, Quinn LM, Landefeld CS. *Am J Med* 1998;**105**:91–99.
2 Kuijer PM, Hutten BA, Prins MH, *et al. Arch Intern Med* 1999;**159**:457–460.
3 Levine MN, Raskob G, Beyth RJ, *et al. Chest* 2004;**126**:287S–310S.

CLINICAL UPDATE

Update in Addiction Medicine for the Generalist

Adam J. Gordon, MD, MPH[1,2,3], Lynn E. Sullivan, MD[4], Daniel P. Alford, MD, MPH[5], Julia H. Arnsten, MD, MPH[6], Marc N. Gourevitch, MD, MPH[7], Stefan G. Kertesz, MD, MSc[8], Hillary V. Kunins, MD, MPH, MS[6], Joseph O. Merrill, MD[9], Jeffrey H. Samet, MD, MA, MPH[5], and David A. Fiellin, MD[4]

[1]Mental Illness Research, Education, and Clinical, Center, University of Pittsburgh, Pittsburgh, PA, USA; [2]Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, Mailcode 151-C, University Drive C, Pittsburgh, PA 15240, USA; [3]Center for Research on Health Care, University of Pittsburgh, Pittsburgh, PA, USA; [4]Yale University School of Medicine, New Haven, CT, USA; [5]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [6]Albert Einstein College of Medicine and Montefiore Medical Center, New York, NY, USA; [7]New York University School of Medicine, New York, NY, USA; [8]University of Alabama at Birmingham School of Medicine, and the Deep South Center on Effectiveness, Birmingham Veterans Affairs Medical Center, Birmingham, AL, USA; [9]University of Washington School of Medicine, Seattle, WA, USA.

KEY WORDS: primary health care; substance-related disorders; review literature.
DOI: 10.1007/s11606-007-0133-4
 2007;22:1190–1194

Generalist physicians can play a critical role in identifying and treating patients with addictions to alcohol, nicotine, and/or other drugs of abuse. In the United States, nicotine dependence and unhealthy alcohol use are the first (18.1%) and third (3.5%) leading causes, respectively, of preventable deaths.[1] Primary care physicians have not traditionally treated substance use despite the harmful effects that addiction can cause in their patients. The objective of this paper is to present recent evidence on recognizing and treating addiction disorders that is relevant for generalist physicians. We conducted an electronic database (PubMed) search to systematically identify recent (January 1, 2003, to June 1, 2006), human subject, English language, peer-reviewed, research articles or publications that impact generalist care for patients with addiction disorders. The search strategy and consensus deliberations were used to identify important articles in the categories of screening strategies for patients with alcohol problems and use of specific pharmacotherapies for patients with alcohol, nicotine, and opioid dependence.

Alcohol Disorders

Anton RF, O'Malley SS, Ciraulo DA, Couper D, Donovan DM, Gastfriend DR et al. Combined pharmacotherapies and behavioral interventions for alcohol dependence JAMA 2006;295(17):2003–2017

Received July 31, 2006
Revised October 9, 2006
Accepted January 16, 2007
Published online May 10, 2007

Alcohol dependence affects approximately 8 million persons in the United States, contributing to substantial morbidity and mortality.[2–4] To date, however, most evidence regarding primary care interventions for unhealthy alcohol use has been in patients with hazardous or "at-risk" drinkers, not alcohol-dependent individuals.[5]

The Combined Pharmacotherapies and Behavioral Interventions (COMBINE) study was designed to test interventions for alcohol-dependent persons, and this trial included medications and an intervention explicitly designed for nonspecialty settings, like primary care.[6,7] In this 9-arm controlled trial, alcohol-dependent persons (n=1,383) were assigned to combinations of the following: (a) a 9-visit primary care counseling intervention offered by medical professionals (typically nurses or physicians) termed Medical Management (MM), 8 arms; (b) daily oral naltrexone (an opioid receptor antagonist), 4 arms; (c) daily oral acamprosate (a putative glutamate modulator), 4 arms; and (d) up to 20 formal counseling sessions from alcohol treatment specialists termed the Combined Behavioral Intervention (CBI), 5 arms.[8,9] Participants were required to be abstinent for 4 to 21 days prior to entry, and candidates were excluded if they had drug abuse or psychiatric disorders requiring medication. Overall follow-up averaged 94% at the end of active treatment (16 weeks) and 82% at 1-year post-treatment (68 weeks), with no significant differences by trial arm.

Investigators assessed multiple trial outcomes, including percent days abstinent and return to heavy drinking. An overall "good clinical outcome" was defined as abstinence or moderate drinking—maximum of 11 (women) or 14 (men) drinks per week—with no more than 2 days on which more than 3 drinks (women) or 4 drinks (men) were consumed during the last 8 weeks of the 16-week trial period.

The percentages of good clinical outcomes at the end of treatment were 58% for MM/placebo, 74% for naltrexone/MM, 71% for CBI/MM, and 74% for naltrexone/CBI/MM. The odds ratios for good outcomes were significantly increased for the combination therapy groups compared with MM/placebo alone (all p<0.01). Relative to persons who received placebo

with medical management, the number needed to treat (NNT) to assure 1 additional good clinical outcome was 7 for CBI/MM, 6 for naltrexone/MM, and 7 for naltrexone/CBI/MM. One year after treatment, trial arm differences still favored the combination therapy groups, but results were no longer significant. The percentage of days abstinent (59–69%) remained higher than at baseline (24–25%).

Interpreting the trial arm comparisons was hindered by the complex study design. Broadly, this study showed that naltrexone and/or CBI in combination with MM were helpful treatments for alcohol dependent patients compared to MM alone. The applicability of COMBINE's findings in contemporary primary care practice remains unclear. While MM was explicitly designed to be offered by clinicians who are not addiction medicine specialists, the time required for the intervention (9 sessions, ranging from 20–45 minutes in length) and the currently available training materials (a 141-page manual) would seem difficult to incorporate in primary care practice.[6]

A separate finding in COMBINE highlights a more general challenge to the study of addiction treatment interventions. Consistent with recent trials on pharmaceutical interventions for alcohol problems, all trial arms achieved alcohol reductions exceeding the differences between trial arms.[10–12] Improvements independent of treatment assignment reflect the dominant impact of the contextual and motivational factors that propel patients to seek treatment and/or reduce drinking in the absence of treatment.[12] Primary care physicians may learn of such clinically important factors in the course of care. How and whether physicians should seek to assist patient recovery processes in light of these factors may prove difficult to study in the context of randomized controlled trials.

National Institute on Alcohol Abuse and Alcoholism. Helping patients who drink too much: a clinician's guide. U.S. Department of Health & Human Services, Rockville, MD. 2005.

For patients who are not alcohol-dependent, evidence supports screening and counseling to reduce misuse in primary care settings, yet this recommendation is sporadically followed in everyday practice.[13,14] For this reason, the National Institute on Alcohol Abuse and Alcoholism (NIAAA) recently issued revised guidelines designed to simplify integration of screening and intervention into primary care.[15] The guide contains evidence and consensus-based recommendations and its chief innovation is a simplified alcohol use screening algorithm. Persons who respond positively to a prescreening question (e.g., "Do you sometimes drink alcoholic beverages?") can be asked the following single question: "How many times in the past year have you had (>5 for men, >4 for women) drinks in a day?" This single-question approach, positive when any such episodes are reported in the prior 12 months, has reasonable sensitivity and specificity for detecting alcohol misuse.[16] Several investigators have examined single-item screening strategies.[17–21] For example, Taj found that a single question had a positive predictive value of 74% [95% confidence interval (CI), 66% to 83%], negative predictive value of 88% (95% CI, 80% to 94%), sensitivity of 62%, and specificity of 93% in detecting problem drinking.[21] Further research is warranted regarding the specific screening question from the Clinician's Guide, although the single question it recommends is similar to those previously studied. When responses to the screening questions are negative, only brief advice regarding safe use and annual rescreening are warranted. When positive, questions to distinguish at-risk use from abuse (interfering with the patient's life) or dependence (associated with tolerance, withdrawal, or preoccupation) are indicated.

The guide outlines and provides scripts for brief advice and assistance that clinicians can tailor to the patient's degree of alcohol use and motivation to change. Additional resources review the alcohol contents of standard drinks, more formal screening instruments, and referral options, as well as an abbreviated guide to pharmacologic treatment of alcohol dependence.

Nicotine Dependence

Critchley J, Capewell S. Smoking cessation for the secondary prevention of coronary heart disease. Cochrane Database Syst Rev 2004;(1):CD003041.

Despite a reduction in prevalence, smoking accounts for over 435,000 deaths annually.[22,23] Smoking is a well-known risk factor for developing coronary artery disease (CAD). However, the magnitude of the impact of smoking cessation in reducing overall and cardiovascular mortality in patients with CAD is not well known.

To estimate the magnitude of risk reduction when a patient with CAD stops smoking, investigators conducted a systematic review and meta-analysis.[24] The authors searched for studies in MEDLINE, EMBASE, Science Citation Index, Cochrane Controlled Trial Register, CINAHL, PsychLit, and Dissertation Abstracts through April 2003. In addition, searches were conducted of relevant conference proceedings, the UK National Research Register, references from retrieved articles, and experts in the field, and contact was made with the investigators from 61 cohort studies of cardiovascular risk. Finally, the authors attempted to identify any randomized clinical trials that collected but had not reported appropriate data.

Because widely accepted standards for assessing quality in observational research are lacking, the authors included those articles that had a minimum of 2-years follow-up (considered adequate time for smoking cessation to reduce risk) and employed 2 independent reviewers who assessed other aspects of quality (e.g., control of confounding, sample size, selection biases, operational definitions of smoking, smoking cessation, and index events).[25,26]

The 20 prospective cohort studies that were identified followed 12,603 patients with CAD (myocardial infarction, stable or unstable angina) for at least 2 years, conducted at least 2 assessments of smoking status, and provided cardiovascular outcomes and all-cause mortality. Eight studies included more than 500 subjects, predominantly males. While most studies had operational definitions of index cardiac events, few had biologic confirmation of smoking status. Six studies were assessed to have good control of confounding, whereas 9 had poor control. Losses to follow up were reported to be relatively small, although few studies were able to capture losses that occurred between the index cardiac event and study enrolment.

The authors conducted a metaregression to help adjust for study heterogeneity; predictor variables included control of confounding variables, minimization of selection biases, and operational definitions for smoking status, smoking cessation,

and outcome events. Among those who quit smoking, the relative risk (RR) of mortality was reduced 36% compared with those who continued to smoke [1,044/5,659 vs 1,884/6,944; crude RR 0.64, 95% CI 0.58 to 0.71]. While these results were dominated by 3 large studies that included patients with bypass Surgery, the results did not differ significantly when limited to trials of patients who were enrolled following myocardial infarction. In addition, the risk of nonfatal myocardial infarctions was reduced by 32% for those who quit smoking (263/2,467 vs 516/3,662; crude RR 0.68, 95% CI 0.57 to 0.82). The authors conducted a sensitivity analysis of those studies that had an initial sample size of at least 500 smokers at baseline, at least 85% follow-up, and adequate or good control of confounding. The main findings were similar in these 6 studies. The risk reduction associated with quitting smoking was similar across index cardiac events, age, sex, country, and time period.

These findings highlight the benefit of smoking cessation for patients with CAD. The 36% risk reduction in all-cause mortality is similar to benefits seen with treating hypercholesterolemia and hypertension.[27–30] While these findings were not demonstrated in randomized trials, the authors attempted to reduce bias in this meta-analysis by creating rigorous eligibility criteria for the studies, and conducting a sensitivity analysis using the highest quality studies. However, there are still potential limitations to the generalizability of the meta-analysis. First, patients who cease smoking following a cardiac event may differ from those who do not. Unmeasured confounders could account for the mortality benefit seen in patients who quit smoking. Secondly, because most studies relied on subject self-report, it is possible that studies were not able to accurately classify the subject's smoking status at follow-up. Nonetheless, the high prevalence of nicotine dependence among patients with CAD, the impact of smoking cessation on mortality, and the effectiveness of smoking cessation interventions all support clinician efforts to aggressively treat nicotine dependence in patients with known CAD. Physicians should address smoking cessation using guideline-consistent strategies in all patients who smoke and consider the added benefit in those patients with established CAD, especially those who appear motivated following a significant cardiac event.[31]

Gonzales D, Rennard SI, Nides M, Oncken C, Azoulay S, Billing CB et al. Varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs sustained-release bupropion and placebo for smoking cessation: a randomized controlled trial. JAMA 2006;296(1):47–55.

Jorenby DE, Hays JT, Rigotti NA, Azoulay S, Watsky EJ, Williams KE et al. Efficacy of varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs placebo or sustained-release bupropion for smoking cessation: a randomized controlled trial. JAMA 2006;296(1):56–63.

Buproprion and nicotine replacement therapies are the mainstay of pharmacotherapy for smoking cessation. Previous research has shown that current smoking cessation treatments are not widely disseminated to the general population of smokers.[32] A new agent, varenicline, is an $\alpha 4\beta 2$ nicotinic acetylcholine receptor partial agonist that stimulates the release of dopamine to reduce craving and withdrawal, while simultaneously blocking nicotine's reinforcing effects. Two industry-funded, randomized, double-blinded trials evaluated varenicline against bupropion SR and placebo in adults who were motivated to stop smoking.[33,34] Randomization was stratified by center. In the first trial, the 1,025 subjects who met inclusion criteria (age 18–75, smoked >10 cigarettes per day, had <3 months of smoking abstinence in the past year, and were motivated to quit) received varenicline, 0.5 mg/day for days 1–3, 0.5 mg twice per day for days 4–7, then 1 mg twice per day through week 12 (n=352), or SR bupropion 50 mg/day for days 1–3, then 150 mg twice per day through week 12 (n=329), or placebo (n=344). The second trial[26] had a similar design and enrolled 1,027 subjects.

The primary endpoint for both trials was continuous abstinence for the final 4 weeks of treatment, defined as the proportion of participants who reported no smoking or use of any nicotine-containing products, confirmed by exhaled carbon monoxide. Secondary endpoints included continuous abstinence rates from week 9 through week 24 and through week 52. The 52-week study completion rates in the first study were 60.5% for varenicline, 56% for bupropion SR, and 54% for placebo, and 70% for varenicline, 65% for bupropion SR, and 60% for placebo in the second study.

In the first trial, the investigators found that subjects receiving varenicline (44.0%) were most likely to achieve the primary abstinence endpoint compared with placebo subjects (17.0%) (p<0.001; NNT, 3.7) and SR buproprion subjects (29.5%, p<0.001, NNT, 6.9). Subjects receiving varenicline (29.5%) were also more likely to achieve secondary abstinence endpoints compared with placebo subjects (10.5%, p<0.001, NNT, 5.2) and SR bupropion subjects (20.7%, p=0.007, NNT, 11.4) at 24-week follow-up. However, varenicline was superior only to placebo (21.9 vs 8.4%, p<0.001, NNT, 7.4) for the primary endpoint at 52 weeks and not to SR buproprion (16.1%, p=0.06). In the second study, varenicline was superior to SR bupropion at 52-week follow-up (30.5 vs 23.4%, p=0.05, NNT, 14.1). Minor adverse events, including nausea and abnormal dreams, were more common in the varenicline group in both trials, but did not cause patients to discontinue therapy.

Because these were the first published trials of varenicline's efficacy for quitting smoking, important questions remain unanswered about how effective varenicline will be in routine clinical practice. These 2 studies were also not designed primarily to test varenicline's efficacy for maintenance of abstinence, although other studies have addressed this question, using higher varenicline doses.[35] Finally, it is currently unknown whether nicotine replacement therapy will be effective when used with varenicline, whether postmarketing surveillance will reveal rare but serious complications, whether there will be important drug interactions with varenicline, or whether insurance plans will cover varenicline.

Opioid Dependence

Turner BJ, Laine C, Lin YT, Lynch K. Barriers and facilitators to primary care or human immunodeficiency virus clinics providing methadone or buprenorphine for the management of opioid dependence. Arch Intern Med 2005;165(15):1769–1776.

Dependence upon heroin and prescription opioid analgesics is an important public health problem.[2,36–38] Historically, effective treatment for opioid dependence has been limited to methadone provided in specialty substance abuse programs. The Drug Addiction Treatment Act of 2000 and the approval of buprenorphine in 2002 have, for the first time, allowed physicians to prescribe medication for the treatment of opioid dependence in primary care settings.

This study examined factors associated with interest in delivering buprenorphine and methadone treatment by primary care and HIV specialty clinics.[39] In a survey of 261 clinics that serve Medicaid patients, 41% of which provided HIV care, medical directors were questioned about knowledge and attitudes towards opioid dependence treatment, and potential facilitators and barriers to providing this treatment. Analyses compared willingness to provide buprenorphine versus methadone and examined the associations of clinic characteristics and attitudes with interest in offering these treatments. Clinics were more interested in providing buprenorphine than methadone (60 vs 33% $p<0.001$), partly because of the perceived lower abuse potential and stigma. Nearly half of respondents had at least 1 negative opinion about patients with opioid dependence. Interest in providing buprenorphine was greater in clinics offering HIV services, treating more patients with chronic pain, or those affiliated with methadone programs. Fifty-seven percent of medical directors reported a high level of concern about adequate reimbursement for treating opioid dependency. Greater than 60% cited the availability of Continuing Medical Education credits and telephone access to addiction medicine expertise as important in enhancing their interest in providing this care.

This study highlights important barriers to adopting buprenorphine treatment in primary care and HIV specialty practices. Information concerning methadone is less relevant given the continued regulatory barriers to providing this treatment outside methadone clinics. The stigma of opioid dependence and its treatment remains a major obstacle to disseminating medication-assisted treatment in medical clinics, although perhaps less so for buprenorphine. Access to substance abuse treatment continues to present major challenges, especially for patients relying on public funding mechanisms that separate medical and specialty addiction treatment services. HIV clinics appear to be appropriate sites for developing buprenorphine programs (http://www.bhives.org). Access to physicians with experience treating opioid dependence (http://www.pcssmentor.org) may be an important factor in the willingness of physicians and clinics to provide office-based treatment.

Mainstreaming Addiction Care Into Practice

Institute of Medicine. Improving the quality of healthcare for mental and substance-use conditions: the quality chasm series. The National Academies Press, Washington, D.C. 2005.

The move to address patients' alcohol and drug use by generalist physicians should be advanced by the current attention to improving the quality of medical care. A recently published Institute of Medicine (IOM) report developed an agenda for change of the health care system's approach to the treatment and prevention of these conditions.[40] The IOM's underlying theme is that only by addressing substance use and mental health problems can patients achieve optimal benefit.

The IOM report advocates coordinated care among primary care, mental health, and substance-use treatment providers. Specifically, coordination models can be straightforward (i.e., formal agreements among mental, substance-use, and primary health care providers) to incrementally more complex arrangements (i.e., case management among systems, co-location of services, to clinically integrated practices).[40] Evidence suggests that the more complex arrangements have stronger evidence for yielding improved patient outcomes. Other recommendations addressed to all clinicians including primary care physicians include the need to screen all mental health patients for alcohol and drug use problems given the high comorbidity of these conditions. The report also stressed maintaining a patient-centered approach to the care of individuals with alcohol and drug problems. That means providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide clinical decisions.[41] The increasing attention to the impact of substance use disorders on patients' medical outcomes should drive efforts to engage generalists in identifying and treating these disorders.

Acknowledgements: *All authors are members of the Society of General Internal Medicine's Substance Abuse Task Force to which we are deeply indebted for its support. The authors also wish to thank Kara Mays for her contributions of coordinating the review and editing this paper.*

Potential Financial Conflicts of Interest: *Dr. Gourevitch reported receiving a grant from Cephalon, regarding use of depot naltrexone (Vivitrol) for treatment of alcohol dependence. The other authors reported no conflicts of interest.*

Corresponding Author: *Adam J. Gordon, MD, MPH, Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, Mailcode 151-C, University Drive C, Pittsburgh, PA 15240, USA (e-mail: adam.gordon@va.gov).*

REFERENCES

1. **Mokdad AH, Marks JS, Stroup DF, Gerberding JL.** Actual causes of death in the United States, 2000. JAMA. 2004;291(10):1238–45.
2. Substance Abuse and Mental Health Services Administration Office of Applied Studies. Results from the 2004 Survey on Drug Use and Health: National Findings. Rockville, MD: Substance Abuse and Mental Health Services Administration Office of Applied Studies; 2005. DHHS publication no. SMA 05-4062, NSDUH series H-28.
3. **Grant BF, Dawson DA, Stinson FS, et al.** The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. Drug Alcohol Depend. 2004;74(3):223–34.
4. **Saitz R.** Clinical practice. Unhealthy alcohol use. N Engl J Med. 2005;352(6):596–607.
5. **Whitlock EP, Polen MR, Green CA, Orleans T, Klein J.** Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2004;140(7):557–68.
6. **Pettinati HM, Weiss RD, Miller WR, Donovan D, Denise B, Rounsaville BJ.** COMBINE Monograph Series, Volume 2. Medical Management Treatment Manual: A Clinical Research Guide for Medically Trained Clinicians Providing Pharmacotherapy as Part of the Treatment for Alcohol Dependence. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2004. DHHS publication no. (NIH) 04-5289.

7. **Anton RF, O'Malley SS, Ciraulo DA, et al.** Combined pharmacotherapies and behavioral interventions for alcohol dependence. JAMA. 2006;295(17):2003–17.
8. **Pettinati HM, Weiss RD, Dundon W, et al.** A structured approach to medical management: a psychosocial intervention to support pharmacotherapy in the treatment of alcohol dependence. J Stud Alcohol Suppl. 2005;(15):170–8.
9. **Miller WR.** Combined behavioral intervention manual. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2004. NIH publication no. 04-5288.
10. **Garbutt JC, Kranzler HR, O'Malley SS, et al.** Efficacy and tolerability of long-acting injectable naltrexone for alcohol dependence: a randomized controlled trial. JAMA. 2005;293(13):1617–25.
11. **Miller WR, Walters ST, Bennett ME.** How effective is alcoholism treatment in the United States? J Stud Alcohol. 2001;62(2):211–20.
12. **Tucker JA, Roth DL.** Extending the evidence hierarchy to enhance evidence-based practice for substance use disorders. Addiction. 2006;101(7):918–32.
13. Agency for Healthcare Research and Quality. The Guide to Clinical Preventive Services, 2005. Recommendations of the U.S. Preventive Services Task Force. Available at http://www.preventiveservices.ahrq.gov. Accessed December 1, 2006.
14. **McGlynn EA, Asch SM, Adams J, et al.** The quality of health care delivered to adults in the United States. N Engl J Med. 2003;348(26):2635–45.
15. National Institute on Alcohol Abuse and Alcoholism. Helping Patients Who Drink Too Much: A Clinician's Guide. Rockville, MD: U.S. Department of Health & Human Services; 2005. NIH publication no. 05-3769.
16. **Fleming MF.** In search of the Holy Grail for the detection of hazardous drinking. J Fam Pract. 2001;50(4):321–2.
17. **Gordon AJ, Maisto SA, McNeil M, et al.** Three questions can detect hazardous drinkers. J Fam Pract. 2001;50(4):313–20.
18. **Canagasaby A, Vinson DC.** Screening for hazardous or harmful drinking using one or two quantity–frequency questions. Alcohol Alcohol. 2005;40(3):208–13.
19. **Williams R, Vinson DC.** Validation of a single screening question for problem drinking. J Fam Pract. 2001;50(4):307–12.
20. **Vinson DC.** Screening question for alcohol abuse. Am Fam Physician. 2002;66(2):209.
21. **Taj N, Devera-Sales A, Vinson DC.** Screening for problem drinking: does a single question work? J Fam Pract. 1998;46(4):328–35.
22. National Center For Chronic Disease Prevention and Health Promotion, Center for Disease Control. Percentage of adults who were current, former, or never smokers, overall and by sex, race, Hispanic origin, age, and education, National Health Interview Surveys, selected years—United States, 1965–2000. Available at: http://www.cdc.gov/tobacco/research_data/adults_prev/adstat1print.htm. Accessed October 5, 2006.
23. Centers for Disease Control and Prevention. Smoking-attributable mortality and years of potential life lost—United States, 1984. MMWR Morb Mortal Wkly Rep. 1997;46(20):444–51.
24. **Critchley J, Capewell S.** Smoking cessation for the secondary prevention of coronary heart disease. Cochrane Database Syst Rev. 2004;(1):CD003041.
25. **Friedenreich CM.** Methods for pooled analyses of epidemiologic studies. Epidemiology. 1993;4(4):295–302.
26. **Downs SH, Black N.** The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52(6):377–84.
27. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Lancet. 2002;360(9326):7–22.
28. National Cholesterol Education Program Expert Panel on Detection EaToHBCiAATPI. Third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. 2002;106(25):3143–421.
29. **Collins R, Peto R, MacMahon S, et al.** Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335(8693):827–38.
30. **Grundy SM, Cleeman JI, Merz CN, et al.** Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. Circulation. 2004;110(2):227–39.
31. The Tobacco Use and Dependence Clinical Practice Guideline Panel SaCR. A clinical practice guideline for treating tobacco use and dependence: a US Public Health Service report. JAMA. 2000;283(24):3244–54.
32. **Thorndike AN, Rigotti NA, Stafford RS, Singer DE.** National patterns in the treatment of smokers by physicians. JAMA. 1998;279(8):604–8.
33. **Gonzales D, Rennard SI, Nides M, et al**. Varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs sustained-release bupropion and placebo for smoking cessation: a randomized controlled trial. JAMA. 2006;296(1):47–55.
34. **Jorenby DE, Hays JT, Rigotti NA, et al**. Efficacy of varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs placebo or sustained-release bupropion for smoking cessation: a randomized controlled trial. JAMA. 2006;296(1):56–63.
35. **Tonstad S, Tonnesen P, Hajek P, et al.** Effect of maintenance therapy with varenicline on smoking cessation: a randomized controlled trial. JAMA. 2006;296(1):64–71.
36. **Johnston LD, O'Malley PM, Backman JG, Schulenberg JE.** Monitoring the Future National Results on Adolescent Drug Use: Overview of Key Findings. Bethesda, MD: National Institute on Drug Abuse; 2005.
37. **Grant BF, Hasin DS, Stinson FS, et al.** Prevalence, correlates, comorbidity, and comparative disability of DSM-IV generalized anxiety disorder in the USA: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Psychol Med. 2005;35(12):1747–59.
38. **Gilson AM, Ryan KM, Joranson DE, Dahl JL**. A reassessment of trends in the medical use and abuse of opioid analgesics and implications for diversion control: 1997–2002. J Pain Symptom Manage. 2004;28(2):176–88.
39. **Turner BJ, Laine C, Lin YT, Lynch K.** Barriers and facilitators to primary care or human immunodeficiency virus clinics providing methadone or buprenorphine for the management of opioid dependence. Arch Intern Med. 2005;165(15):1769–76.
40. Institute of Medicine. Improving the Quality of Healthcare for Mental and Substance-use Conditions: The Quality Chasm Series. Washington, DC: The National Academies Press; 2005.
41. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: The National Academies Press; 2001.

MAJOR ARTICLE

Effect of Exposure to Injection Drugs or Alcohol on Antigen-Specific Immune Responses in HIV and Hepatitis C Virus Coinfection

Camilla S. Graham,[1] Annalee Wells,[1a] Erika M. Edwards,[2] Timothy Herren,[2] Sheila Tumilty,[3] Sherri O. Stuver,[2] Jeffrey H. Samet,[3] David Nunes,[3] C. Robert Horsburgh, Jr.,[2,3] and Margaret James Koziel

[1]Beth Israel Deaconess Medical Center and Harvard Medical School, [2]Boston University School of Public Health, and [3]Boston Medical Center and Boston University School of Medicine, Boston, Massachusetts

Background. Ongoing substance use is a potential confounder for immunological studies on hepatitis C virus (HCV), but there is little in the literature regarding the effects of injection drug use (IDU) or alcohol on HCV-specific immune responses. We wanted to determine whether IDU or alcohol affected immune responses in HCV-infected and human immunodeficiency virus (HIV)/HCV coinfected subjects.

Methods. Eight-four subjects with HIV/HCV and 57 with HCV were classified as either injection drug users, drinkers, or nonusers based on questionnaire results. Immune responses were studied with enzyme-linked immunosorbent spot assay for interferon (IFN)–γ, interleukin (IL)–10, and tumor necrosis factor (TNF)–α against HCV proteins Core, NS3, and NS5 and recall antigens.

Results. Subjects with HIV/HCV, in aggregate, had significantly lower HCV-specific IFN-γ and TNF-α responses than subjects with HCV. However, HIV/HCV injection drug users had HCV-specific IFN-γ and IL-10 responses that were similar to those of HCV injection drug users and were significantly higher than in nonusers with HIV/HCV. Conversely, subjects who drank alcohol had similar immune responses to those who were abstinent, among both subjects with HIV/HCV and subjects with HCV.

Conclusions. Studies that examine IFN-γ or IL-10 immune responses in HIV/HCV-coinfected or HCV-infected persons need to consider current IDU. Alcohol, at levels consumed in this cohort, does not appear to have as much of an effect on antigen-specific immune responses.

Hepatitis C virus (HCV) is not considered cytopathic in most cases, and the liver inflammation and progressive fibrosis found in many patients with chronic viremia is believed to be due to an ongoing, ineffectual host immune response to persistent infection. It is paradoxical that HIV infection, an immunocompromised state, is associated with accelerated liver disease, and this suggests that the type of immune response may be as relevant as the quantity of immune response when defining specific immunological profiles that are associated with progressive HCV-related liver disease.

Many persons with HIV/HCV coinfection acquired both infections via parenteral injection drug use (IDU), and ongoing polysubstance abuse is common [1–3]. Persons with HIV/HCV coinfection and heavy alcohol consumption (>50 g/day) are at high risk of severe liver fibrosis [4, 5]. Alcohol modifies cellular and humoral immune responses in cell culture models as well as animal models; studies show an increase in tumor necrosis factor (TNF)–α production or a shift from a type 1 (interferon [IFN]–γ and interleukin [IL]–2) to a type 2 (IL-4 and IL-10) immune response in the presence of alcohol [6–10]. However, little has been reported on the effect of alcohol on HCV-specific cellular immune

Received 17 October 2006; accepted 9 November 2006; electronically published 2 February 2007.

Potential conflicts of interest: none reported.

Presented in part: 12th Conference on Retroviruses and Opportunistic Infections, Boston, 22–25 February 2005 (abstract 920); American Association for the Study of Liver Disease, San Francisco, 11–15 November 2005 (abstract 912).

Financial support: National Institutes of Health (grants DA14495-01 to C.S.G., DK56410 to M.J.K. and C.R.H., and AI49508 to M.J.K.).

[a] Present affiliation: Chicago College of Osteopathic Medicine of Midwestern University, Downers Grove, Illinois.

Reprints or correspondence: Dr. Camilla S. Graham, Beth Israel Deaconess Medical Center, Harvard Institutes of Medicine, Rm. 221, 4 Blackfan Circle, Boston, MA 02115 (cgraham@bidmc.harvard.edu).

The Journal of Infectious Diseases 2007;195:847–56

0022-1899/2007/19506-0013$15.00
DOI: 10.1086/511990

responses in humans or about potential immune mechanisms by which alcohol modifies the disease course of HCV-related liver damage.

Similarly, in animal models, chronic morphine treatment leads to decreased IFN-γ and IL-2 and increased IL-4 and IL-5, which appears to be mediated by increasing GATA 3 expression and decreasing T-bet expression, thus shifting T helper cells to a type 2 phenotype [11]. The opioid antagonist, naloxone, can counteract the type 2 shift in cytokine production and increase IFN-γ and IL-2 production [12]. These data are consistent with an endogenous role of opioids in inhibiting excessive inflammatory responses with injury or other pain-inducing processes by augmenting anti-inflammatory type 2 responses and limiting proinflammatory type 1 responses [13, 14]. However, the effect of opiates on antigen-specific immune responses in humans with chronic HCV infection has not been studied.

It is important to understand how substance abuse affects antigen-specific immune responses because these immune responses may affect HCV-related liver disease outcomes. For example, early after acute HCV infection, it has been shown that persons who develop chronic HCV infection but maintain HCV-specific IFN-γ responses have a slower progression of liver fibrosis [15]. We have shown that persons with HIV/HCV coinfection who have higher HCV and *Candida*-specific IFN-γ immune responses are significantly more likely to have lower inflammation and fibrosis scores on liver biopsy [16].

Our long-term goal has been to define the immunological alterations in cellular immune responses to HCV that may impact on liver disease progression. However, because many of our subjects with HIV/HCV have a history of IDU, we sought to determine whether current substance abuse affects the results of our standard immunologic assays. Therefore, this study was designed to determine whether current IDU or any alcohol exposure was associated with alterations in HCV-specific immune responses and, if so, whether the effects were the same in HCV-infected and HIV/HCV-coinfected subjects. We examined HCV-specific and recall antigen responses in subjects who consumed alcohol or were active injection drug users and who were HCV-infected or HIV/HCV-coinfected, compared with responses in subjects who did not use injection drugs or consume alcohol, to determine the relative impact of HIV status and alcohol or injection drug exposure on cellular immune responses.

SUBJECTS, MATERIALS, AND METHODS

Subjects. Subjects were recruited from the Boston Medical Center Hepatitis C, HIV, and Related Morbidity (CHARM) cohort, a prospective natural history cohort that includes demographics, alcohol and drug use histories, clinical and laboratory information, and collection of peripheral blood mononuclear cells (PBMCs) every 12 months. All subjects who were HCV-antibody positive by ELISA and had detectable HCV RNA completed a questionnaire regarding drug and alcohol use. Those who had PBMCs collected simultaneously for immune assays were included in the cohort from which subjects for this study were selected. Injection drug users were defined as subjects who had injected heroin in the last month before collection of PBMCs and denied any alcohol use in the past year. Drinkers were defined as subjects who had consumed any alcoholic beverages in the last month and denied any IDU in the past year. Subjects who both drank alcohol and were injection drug users were excluded. All subjects completed the Alcohol Use Disorders Identification Test (AUDIT) questionnaire, and subjects with scores $\geqslant$8 were considered to have hazardous drinking [17, 18]. We also used the alcohol consumption ranges in AUDIT to define "heavy drinkers" as subjects who consumed 3 or more drinks 4 or more times a week or at least 6 drinks in one sitting. "Nonusers" were defined as subjects who denied alcohol consumption and IDU for at least 1 year; for the present analysis, a subgroup of nonusers were randomly selected and matched by HIV status to the drinker cohort (the larger subject group). This study was reviewed and approved by the Institutional Review Board of Boston University Medical Center and Beth Israel Deaconess Medical Center, human experimentation guidelines of the respective institutions were followed in the conduct of this research, and all subjects provided written, informed consent.

Recombinant HCV proteins. The recombinant HCV proteins used were derived from HCV genotype 1b and included Core protein (aa 1–115) and nonstructural (NS) proteins NS3 (aa 1007–1534) and NS5 (aa 2622–2868) at 1 μg/mL (Mikrogen).

Enzyme-linked immunosorbent spot (ELISPOT) assay. ELISPOT assays were performed on viable, cryopreserved PBMCs. One hundred microliters of primary monoclonal antibody (MAb) (anti–IFN-γ or anti–IL-10 [Mabtech] or anti–TNF-α [BD Pharmingen]) was added to 96-well polyvinylidene difluoride plates (Millipore) at a concentration of 10 μg/mL in phosphate-buffered saline and incubated overnight at 4°C. Excess antibody was removed, and wells were blocked with complete media (RPMI 1640 [Gibco] supplemented with 10 mmol/L HEPES, penicillin [100 IU/mL], and streptomycin [100 μg/mL]), supplemented with 10% heat-inactivated human AB serum (RPMI-AB; Biowhittaker) for 1 h; 2.5×10^5 PBMCs were cultured for 40 h in the presence of the recombinant HCV proteins Core, NS3, and NS5 at 1 μg/mL. Positive control wells consisted of phytohemaglutinin (5 μg/mL; Sigma), *Candida* cellular antigen (20 μg/mL; Greer Labs), and tetanus toxoid (5 μg/mL; Accurate Chemicals). Negative control wells were media alone and buffer (used to prepare HCV antigens; Mikrogen). After 40 h, plates were washed with PBS followed by PBS containing 0.05% Tween20. Biotin-conjugated secondary MAbs

(100 μL of 1 μg/mL anti–IFN-γ or anti–IL-10 [Mabtech] or 50 μL of 1 μg/mL anti–TNF-α [BD Pharmingen]) were added and incubated for 2 h at 37°C. After rinsing, 100 μg/mL avidin-peroxidase (1:7.5 ratio; ABC Kit; Vector Labs) was added for 30 min. After washing, 100 μL of substrate (3′, 9′ aminoethylcarbazole [Sigma] in dimethyl formamide and sodium acetate buffer [0.05 mol/L; pH 5.0] and 0.3% H_2O_2) was added and incubated until the appearance of red-brown spots, and then rinsed with tap water. The numbers of spots per well were scored using an Automated ELISpot Reader System with KS4.5.21 software (Carl Zeiss). Averaged numbers of spot-forming cells in control wells were subtracted from antigen-stimulated wells to correct for spontaneous cytokine production. Persons performing the ELISPOT assays were blinded to all subject data except unique patient identification numbers and dates of sample collection.

Statistical methods. Data were analyzed in SAS (version 8.2; SAS Institute) using the Mann-Whitney *U* test and Fisher's exact test. Crude versus stratified odds ratios were compared to determine whether immune response effects were confounded by $CD4^+$ T cell count or HIV virological control (HIV RNA <400 copies/mL) among subjects with HIV/HCV coinfection or estimated duration of HCV infection for all subjects. To determine the likelihood of having an immune response, continuous immune response values were converted to dichotomous "responder" and "nonresponder" variables. We determined the cutoff value for responders to be ⩾10 sfc/10^6 PBMCs based on responses in 9 healthy controls to HCV proteins Core, NS3, and NS5 and recall antigens and by determining the bimodal break point between lower frequency and higher frequency responses in subjects for each antigen and cytokine combination and averaging (data not shown). This cutoff value was determined before any comparative analyses were performed. All tests were 2-tailed and evaluated at a significance level of $P \leqslant .05$.

RESULTS

Subject characteristics. In total, we examined 84 subjects with HIV/HCV and 57 subjects with HCV alone. There were no significant differences in age, sex, and race/ethnicity between the HIV/HCV and HCV groups (data not shown). We then examined the cohort by substance use status. In total, we studied 35 injection drug users (16 with HIV/HCV and 19 with HCV), 53 drinkers (34 with HIV/HCV and 19 with HCV), and 53 nonusers (34 with HIV/HCV and 19 with HCV). These substance use–defined groups were also well balanced for age, sex, and race/ethnicity with no significant differences between groups (data not shown). Ninety-four percent of the IDU group injected heroin 5 or more times per week. Forty-six percent of drinkers in this cohort had an AUDIT score ⩾8, indicating hazardous drinking (15/34 subjects with HIV/HCV and 10/19 subjects with HCV). Eight subjects with HIV/HCV and 5 subjects with HCV were further classified as heavy drinkers.

Table 1 compares demographic characteristics between subjects with HIV/HCV and HCV within the IDU, drinkers, and nonusers groups. The HIV/HCV IDU group had significantly lower $CD4^+$ T cell counts than HIV/HCV nonusers group, but there were no other significant differences between the HIV/HCV IDU or drinkers groups and the HIV/HCV nonusers group. Because $CD4^+$ T cell counts are not routinely measured in persons with HCV alone, these data were not available. Because HCV genotype and viral load are typically measured in persons under consideration for treatment of HCV infection and because active substance use is considered a contraindication for treatment, many subjects in this cohort were missing these data. Thus, those HCV-related factors could not be examined.

Cellular immune responses. Unlike the relatively high-frequency T cell responses observed in chronic HIV infection, most studies of HCV-specific responses in persons with HIV/HCV coinfection observe very low-frequency responses that may not be significantly lower than those observed in subjects with HCV alone [19–24]. The HCV proteins used for our ELISPOT assays primarily stimulate $CD4^+$ T cells (based on depletion studies; data not shown). In figure 1, all subjects with HIV/HCV were compared with all subjects with HCV alone (irrespective of substance use status). IFN-γ immune responses to HCV proteins Core and NS5, summed HCV proteins (the sum of responses to HCV proteins Core, NS3, and NS5), *Candida,* and tetanus were significantly higher in HCV subjects, compared with HIV/HCV subjects. In contrast, there were no significant differences in IL-10 secretion between subjects with HCV and subjects with HIV/HCV. Subjects with HCV had significantly higher TNF-α responses to Core and NS5 than subjects with HIV/HCV, whereas there were no significant differences when recall antigens were examined. In summary, HCV-specific IFN-γ responses were low in both HIV/HCV and HCV groups, although statistically higher in the HCV-alone group. Of interest, HCV-specific TNF-α responses were also higher in HCV monoinfection.

Next, we compared antigen-specific immune responses in all subjects with IDU (the HIV/HCV and HCV groups combined) and all drinkers to all nonusers to determine whether there were any overall effects of IDU or alcohol on immune responses (table 2). The IDU group had broadly higher IL-10 responses to both HCV proteins and *Candida* as well as numerically higher TNF-α responses (HCV NS5-4 and *Candida* responses were significantly higher), compared with nonusers. In contrast, drinkers had only higher IL-10 responses to NS5 and *Candida,* compared with nonusers.

We then determined whether the immune response differences seen in subjects with HIV/HCV versus subjects with HCV

Table 1. Demographic and clinical data.

Characteristic	Injection drug users			Drinkers			Nonusers		
	HIV/HCV (n = 16)	HCV (n = 19)	P	HIV/HCV (n = 34)	HCV (n = 19)	P	HIV/HCV (n = 34)	HCV (n = 19)	P
Age, median (IQR), years	44 (37–51)	42 (32–46)	.17	45 (42–48)	48 (43–50)	.21	45 (41–47)	44 (40–48)	.88
Male	9 (56)	10 (53)	.83	26 (76)	11 (58)	.16	25 (74)	12 (63)	.43
Race			.95			.39			.21
White	6 (38)	8 (42)		10 (29)	2 (10)		10 (29)	10 (53)	
Black	4 (25)	5 (26)		19 (56)	13 (68)		16 (47)	7 (37)	
Hispanic	5 (31)	6 (32)		4 (12)	3 (16)		8 (24)	2 (10)	
Other	1 (6)	0 (0)		1 (3)	1 (5)		0 (0)	10 (53)	
Ever used injection drugs	16 (100)	19 (100)		24 (71)	14 (74)	.81	30 (88)	15 (79)	.44
Duration of HCV infection,[a] median (IQR), years	24 (20.5–27.5)	23 (6–31)	.39	29 (15–31)	31.5 (31–34)	.04	25.5 (21–29.5)	22 (15–29)	.25
HIV load <400 copies/mL	5 (36)	NA		15 (46)	NA		13 (41)	NA	
$CD4^+$ T cell count									
Nadir, median (IQR), cells/mm^3	110 (26–195)	ND		265 (162–390)	ND		178 (96–309)	ND	
Current,[b] median (IQR), cells/mm^3	226 (110–387)	ND		476 (290–655)	ND		398 (239–630)	ND	

NOTE. Data are no. (%) unless otherwise indicated. Statistical comparisons were made using the Mann-Whitney U test for continuous data or Fisher's exact test for categorical data. HCV, hepatitis C virus; IQR, interquartile range; NA, not applicable; ND, not done.

[a] If actual date of acquisition of HCV was not known, then duration of HCV infection was estimated as the first year of injection drug use (IDU) or receipt of blood products before 1991.

[b] In addition, the HIV/HCV and HCV IDU groups were compared with the HIV/HCV and HCV nonusers groups, respectively; and the HIV/HCV and HCV drinkers groups were compared with HIV/HCV and HCV nonusers groups, respectively. The HIV/HCV IDU group had significantly fewer $CD4^+$ T cells/mm^3 than the HIV/HCV nonusers group (P < .03). The HCV drinkers group had a significantly longer duration of HCV infection than the HCV nonusers group (P < .001). All other comparisons were not significantly different.

(figure 1) were affected by IDU. We compared responses to summed HCV proteins and *Candida* between (1) the HIV/HCV IDU group and the HIV/HCV nonusers group, (2) the HIV/HCV IDU group and the HCV IDU group; (3) the HIV/HCV nonusers group and the HCV nonusers group; and (4) the HCV IDU group and the HCV nonusers group. As shown in figure 2, the HIV/HCV IDU group had significantly higher IFN-γ secretion to summed HCV proteins than the HIV/HCV nonusers group, whereas the HCV-specific IFN-γ responses were not different between the HIV/HCV IDU and HCV IDU groups. In contrast, the HCV nonusers group had significantly higher HCV-specific IFN-γ responses than the HIV/HCV nonusers group. When we examined IFN-γ responses to *Candida,* we found that both the HCV IDU and HCV nonusers groups had significantly higher responses than the HIV/HCV IDU group and HIV/HCV nonusers group, respectively.

When we examined IL-10 responses, we found that the HIV/HCV IDU group had significantly higher IL-10 secretion to summed HCV proteins than the HIV/HCV nonusers group, whereas the HCV IDU group had similar IL-10 responses to summed HCV proteins, compared with the HCV nonusers group. In contrast, the HCV IDU group had significantly higher IL-10 secretion to *Candida,* compared with the HCV nonusers group. There were no significant differences in TNF-α secretion to summed HCV proteins, whereas both the HIV/HCV nonusers group and the HCV IDU group had significantly higher TNF-α secretion to *Candida* than the HCV nonusers group.

The HIV/HCV IDU group had significantly lower $CD4^+$ T cell counts than the HIV/HCV nonusers group, and it has been described in the HIV literature that advanced immunosuppression is associated with a shift to a type 2–like response [25]. To evaluate whether the higher HCV-specific IL-10 responses in the HIV/HCV IDU group were actually a function of lower $CD4^+$ T cell counts as opposed to IDU, we examined the likelihood of having an HCV-specific IL-10 response in subjects with $CD4^+$ T cell counts <359 cells/mm^3 (the median value for the HIV/HCV IDU and nonusers groups), compared with >359 cells/mm^3. The associations between HCV-specific IL-10 responses in the HIV/HCV IDU group and the HIV/HCV nonusers group were similar in low versus high $CD4^+$ T cell count strata; thus, these responses were more likely related to IDU than to the degree of immunosuppression (data not shown).

Next, we examined the effect of exposure to alcohol on cellular immune responses in the group of drinkers compared with the same nonusers group to which we had compared the IDU group (figure 3). In contrast to findings in figure 1, the HIV/HCV drinkers group had similar HCV-specific IFN-γ responses to the HCV drinkers group, whereas the HIV/HCV nonusers group had significantly lower IFN-γ responses than the HCV nonusers group. However, similar to figure 1, we

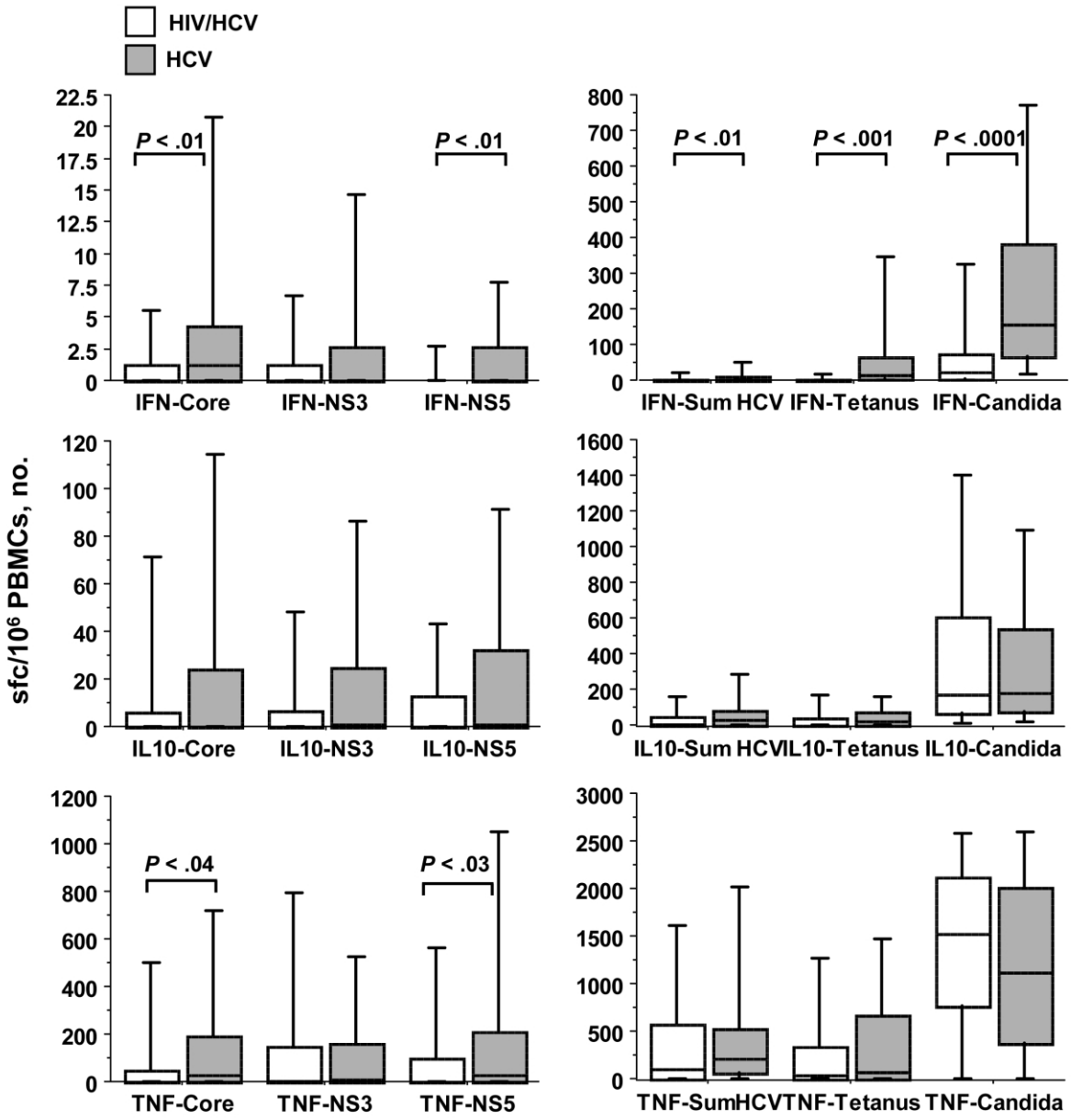


Figure 1. Box and whisker plots demonstrating nos. of spot-forming cells (sfc)/10^6 peripheral blood mononuclear cells (PBMCs) in HIV/hepatitis C virus (HCV; *white bars*) vs. HCV groups *(gray bars)* for interferon (IFN)–γ *(top panels),* interleukin (IL)–10 *(middle panels),* and tumor necrosis factor (TNF)–α *(bottom panels).* Responses to individual HCV antigens (Core, NS3, and NS5) are shown in left panels, and summed HCV-specific (sum of responses to HCV proteins Core, NS3, and NS5), as well as recall antigen responses, are shown in right panels. The box represents the 25th, 50th *(middle line),* and 75th percentile values, whereas the lower and upper lines represent the 10th and 90th percentile values, respectively. Note the different scales for each panel. Significant differences between HIV/HCV and HCV groups ($P \leq .05$, Mann-Whitney *U* test) are indicated.

found that the HCV drinkers and nonusers groups had significantly higher IFN-γ secretion to *Candida* than the HIV/HCV drinkers and nonusers groups, respectively. HCV-specific and *Candida* responses between the HIV/HCV drinkers group and the HIV/HCV nonusers group were also not significantly different for IFN-γ, IL-10, or TNF-α secretion. There were no significant differences in any immune responses between the HCV drinkers group and the HCV nonusers group.

When we examined the immune responses in hazardous drinkers (AUDIT scores $\geq$8) versus nonusers, we found similar patterns of immune responses as when all drinkers were examined, although numerical differences were still not significant (data not shown). We then compared the subgroup of heavy drinkers ($n = 13$) with the nonusers and found that heavy drinkers had significantly higher IL-10 to NS5 (as found when all drinkers were examined; table 2) and TNF-α secretion to

Table 2. Comparisons of immune responses in injection drug users or drinkers vs. nonusers.

Cytokine, antigen	Injection drug users (n = 35)	Drinkers (n = 53)	Nonusers (n = 53)
IFN-γ			
Core	0 (0–3) [9]	0 (0–4) [17]	0 (0–1) [9]
NS3	0 (0–3) [11]	1 (0–1) [8]	0 (0–3) [6]
NS5	0 (0–1) [6]	1 (0–1) [4]	0 (0–1) [6]
Sum-HCV	4 (1–7)	3 (0–11)	1 (0–7)
Candida	156 (41–281)[a] [86]	45 (7–148) [74]	43 (5–153) [66]
Tetanus	11 (0–105) [51]	1 (0–12) [30]	3 (1–11) [26]
IL-10			
Core	5 (0–25)[a] [40]	0 (0–13) [26]	0 (0–5) [21]
NS3	4 (0–21) [37]	0 (0–4) [11]	0 (0–21) [34]
NS5	9 (0–52)[b] [46]	3 (0–16)[a] [34]	0 (0–4) [21]
Sum-HCV	43 (8–106)[b]	7 (0–44)	0 (0–65)
Candida	468 (56–1061)[a] [91]	226 (85–500)[a] [91]	93 (61–302) [89]
Tetanus	24 (3–66) [66]	7 (0–73) [43]	1 (0–76) [38]
TNF-α			
Core	22 (0–134) [53]	1 (0–132) [47]	0 (0–131) [36]
NS3	15 (0–288) [53]	5 (0–117) [45]	0 (0–162) [44]
NS5	10 (0–224)[a] [50]	26 (0–177) [53]	0 (0–36) [35]
Sum-HCV	210 (30–1206)	138 (0–558)	87 (0–424)
Candida	2132 (1318–2572)[b] [97]	1430 (345–1738) [82]	880 (412–1728) [83]
Tetanus	123 (18–1043) [82]	55 (0–343) [57]	34 (0–491) [58]

NOTE. Data are median nos. of spot-forming cells (sfc)/10^6 peripheral blood mononuclear cells (PBMCs) (interquartile range) [percentage of subjects who had at least 10 sfc/10^6 PBMCs for each antigen]. Injection drug users and drinkers were each compared with nonusers. IFN, interferon; IL, interleukin; TNF, tumor necrosis factor.

[a] $P < .05$, Mann-Whitney U test.
[b] $P < .01$, Mann-Whitney U test.

NS5 than nonusers, but all other comparisons were not significantly different (data not shown).

More subjects in the HIV/HCV drinkers group had HIV loads <400 copies/mL than in the HIV/HCV nonusers group. Stratified odds ratios were compared for the likelihood of having an HCV-specific or *Candida* immune response with IFN-γ, IL-10, or TNF-α between the HIV/HCV drinkers and nonusers groups, by HIV load level (stratified by <400 copies/mL or >400 copies/mL). There was no increased likelihood of immune responses associated with being a drinker in the virologically well-controlled group versus the detectable group; so, immunological responses were not a function of the degree of HIV virological control (data not shown). Of note, estimated duration of HCV infection was also not associated with immune responses for any group in this cohort (data not shown).

In summary, the significantly lower IFN-γ HCV-specific responses found when all subjects with HIV/HCV, in aggregate, were compared with all subjects with HCV seem driven by the much lower responses found in the HIV/HCV nonusers group, because the HIV/HCV IDU group and the HIV/HCV drinkers group had similar frequencies of cytokine secretion as the HCV IDU group and HCV drinkers group, respectively. Overall, IDU appeared to affect HCV-specific IFN-γ and IL-10 responses more than alcohol exposure did in this cohort.

DISCUSSION

We have shown that the impact of HIV status on cellular immune responses differs by the cytokine response being tested. IFN-γ secretion was most affected by HIV status, with subjects with HIV/HCV having lower frequency responses for nearly all antigens tested than subjects with HCV alone. Most other studies have not shown a significant difference in HCV-specific IFN-γ secretion between HIV/HCV-infected and HCV-infected groups, although in these studies response frequencies were low in both groups, cohorts were smaller, and current substance use histories were not taken into account [20, 22–24]. One study, using proteins and peptides to stimulate both $CD4^+$ and $CD8^+$ T cells, also showed lower aggregate IFN-γ T cell responses in HIV/HCV-infected versus HCV-infected groups [19]. This finding is of interest because, in a different cohort, we have shown that persons with HIV/HCV coinfection who

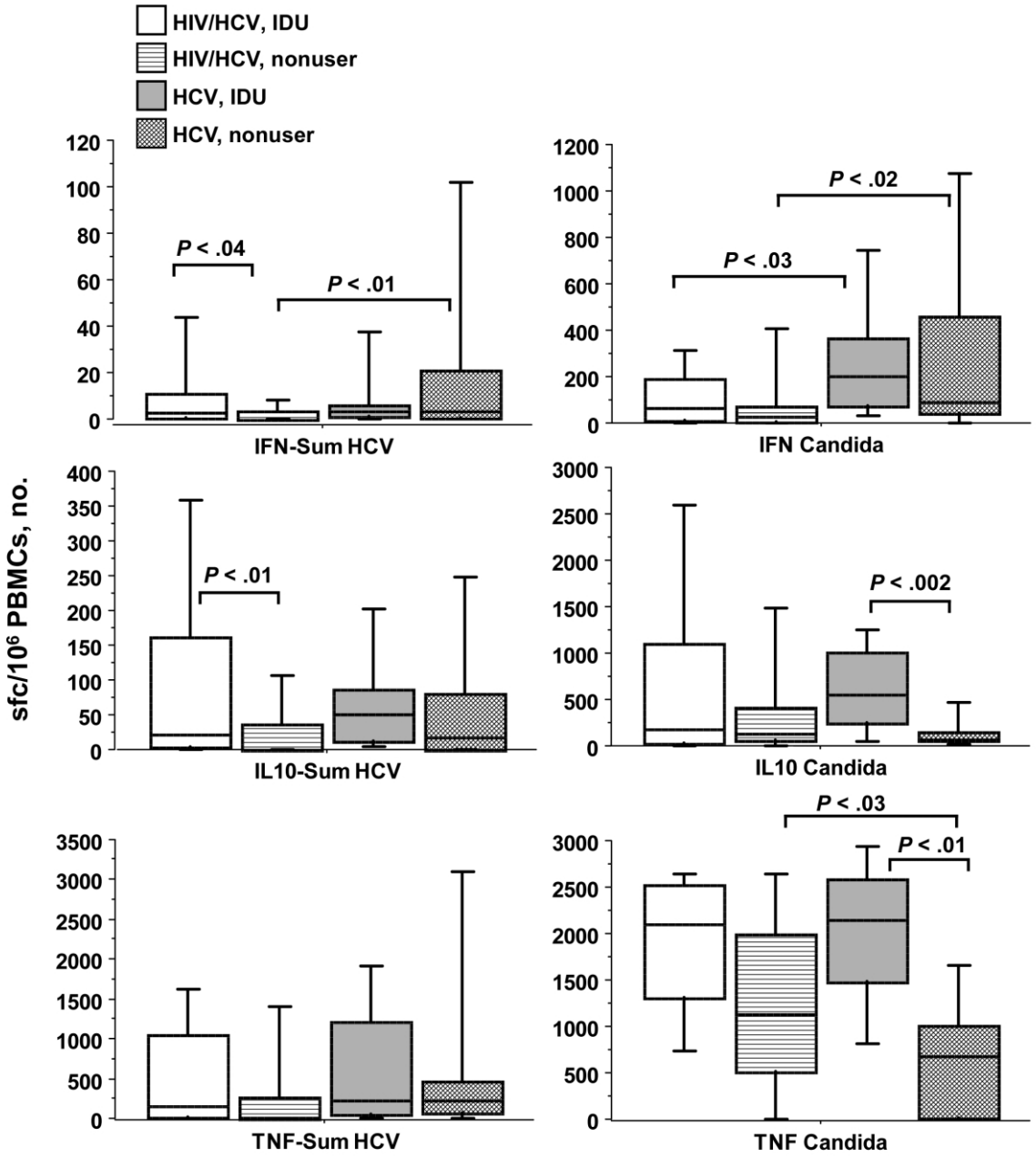


Figure 2. Box and whisker plots demonstrating nos. of spot-forming cells (sfc)/10^6 peripheral blood mononuclear cells (PBMCs) in the HIV/hepatitis C virus (HCV) injection drug use (IDU) group *(white bars),* the HIV/HCV nonusers group *(striped bars),* the HCV IDU group *(gray bars),* and the HCV nonusers group *(hatched bars)* for interferon (IFN)–γ *(top panels),* interleukin (IL)–10 *(middle panels),* and tumor necrosis factor (TNF)–α *(bottom panels)* in response to summed HCV proteins (responses to Core, NS3, and NS5 were summed; *left panels)* and *Candida (right panels).* Note the different scales for each panel. Statistical comparisons were made between the HIV/HCV IDU group and the HIV/HCV nonusers group, the HIV/HCV IDU group and the HCV IDU group, the HCV IDU group and the HCV nonusers group, and the HIV/HCV nonusers group and the HCV nonusers group. Significant differences ($P \leq .05$, Mann-Whitney *U* test) are indicated.

had higher IFN-γ responses to HCV antigens or *Candida* had significantly lower fibrosis and inflammatory scores on liver biopsy [16]. Because many subjects in the present cohort were active substance users and did not have liver biopsies performed, we cannot perform a similar analysis for this study. However, we can speculate that the overall lower IFN-γ responses seen in HIV/HCV-coinfected subjects may be related to the accelerated fibrosis progression seen clinically [26].

Few studies have compared HCV-specific IL-10 responses in HIV/HCV-infected versus HCV-infected groups. One study

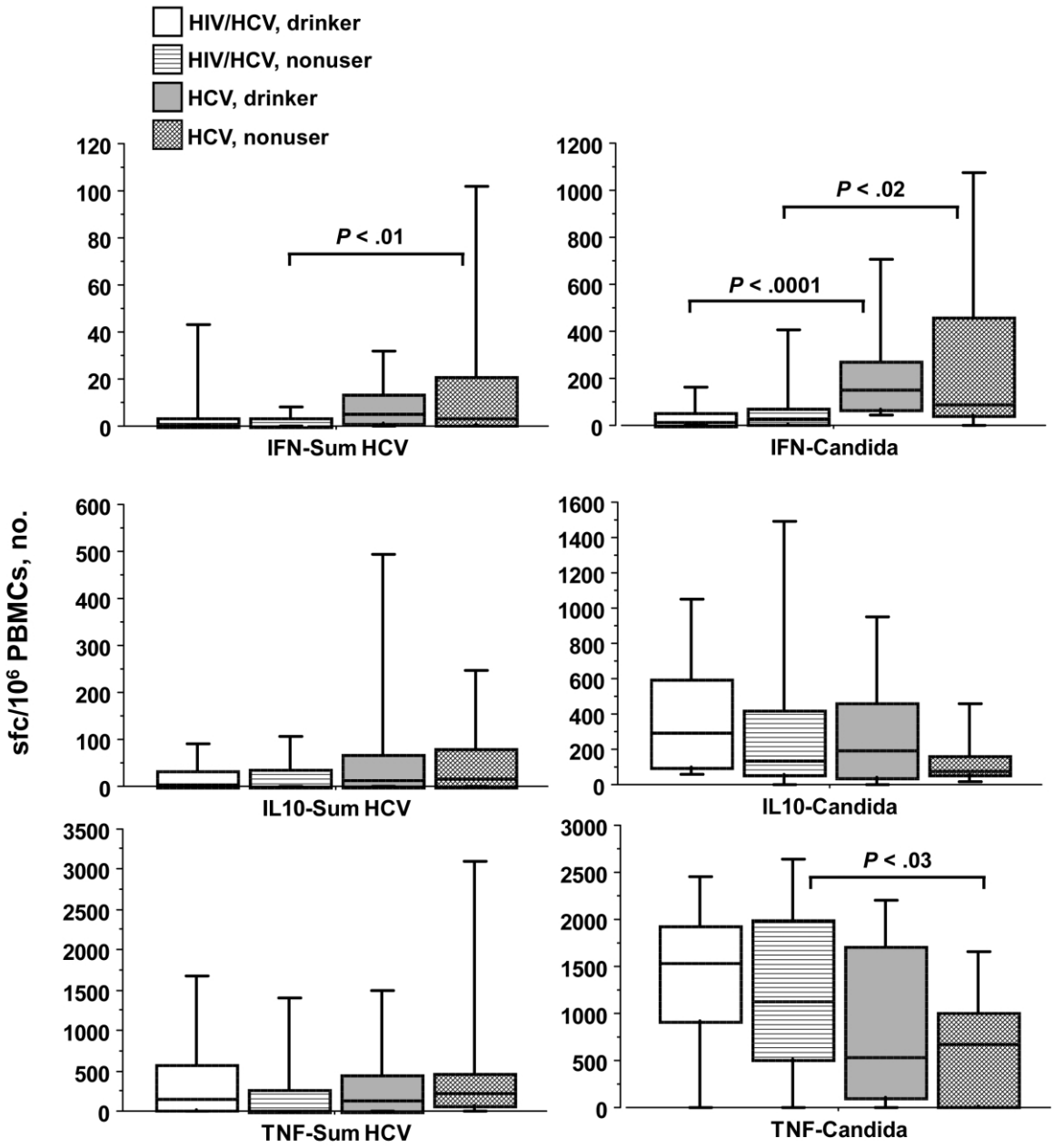


Figure 3. Box and whisker plots demonstrating nos. of spot-forming cells (sfc)/10^6 peripheral blood mononuclear cells (PBMCs) in the HIV/hepatitis C virus (HCV) drinkers group *(white bars),* the HIV/HCV nonusers group *(striped bars),* the HCV drinkers group *(gray bars),* and the HCV nonusers group *(hatched bars)* for interferon (IFN)–γ *(top panels),* interleukin (IL)–10 *(middle panels),* and tumor necrosis factor (TNF)–α *(bottom panels)* in response to summed HCV proteins (responses to Core, NS3, and NS5 were summed; *left panels*) and *Candida (right panels).* Note the different scales for each panel. Statistical comparisons were made between the HIV/HCV drinkers group and the HIV/HCV nonusers group, the HIV/HCV drinkers group and the HCV drinkers group, the HCV drinkers group and the HCV nonusers group, and the HIV/HCV nonusers group and the HCV nonusers group. Significant differences ($P \leq .05$, Mann-Whitney *U* test) are indicated.

found that women with HIV/HCV had low but similar IL-10 responses to HCV and HIV antigens compared with HCV-monoinfected subjects, although HIV/HCV-coinfected women had significantly lower IL-10 to CMV antigens, compared with HIV- or HCV-monoinfected women [24]. We found that secretion of IL-10 to HCV or recall antigens was not significantly different between the HIV/HCV-infected and HCV-infected groups. With overall depressed IFN-γ responses, the HIV/HCV-coinfected subjects in this cohort have a relative shift to type 2–like responses, as has been seen in other HIV cohorts with and without other coinfections [25, 27]. This immune state has been associated with an increased rate of liver fibrosis pro-

gression in HCV/schistosomiasis coinfection and has been associated with poorer response to IFN-based treatment of HCV infection [15, 28]

Based on previous animal models of the effect of opiates on immune responses, we had hypothesized that subjects who use infection drugs would have a shift to lower IFN-γ and higher IL-10 responses than nonusers, regardless of HIV status. When we examined all injection drug users in aggregate, we did find broadly higher IL-10 responses in injection drug users compared with nonusers. Unexpectedly, the subset of the IDU group with HIV/HCV coinfection had significantly higher IFN-γ and IL-10 HCV-specific responses than the HIV/HCV nonusers group, whereas comparable HCV-specific responses did not differ between the HCV IDU group and the HCV nonuser group. It is unknown how this pattern of cytokine production may affect progression of HCV-related liver disease, but this will be examined in more detail in future prospective studies.

We had also hypothesized that drinkers, regardless of HIV status, would have lower IFN-γ secretion than nonusers because of data supporting a shift to a type 2–like immune response with alcohol exposure, although previous studies have not examined HCV-specific cellular immune responses [10, 29, 30]. Surprisingly, in the present study, alcohol had little effect on antigen-specific IFN-γ, IL-10, or TNF-α immune responses when the HCV drinkers group was compared with the HCV nonusers group. In addition, drinkers with HIV/HCV had similar HCV-specific IFN-γ responses, compared with drinkers with HCV, whereas nonusers with HIV/HCV had significantly lower IFN-γ responses, compared with nonusers with HCV. Therefore, these data do not support a strong shift to type 2 cellular immune responses in the presence of alcohol consumption.

For this study, our aim was to determine whether usual patterns of alcohol consumption would affect our immune response assays. In fact, we found few immune response differences between drinkers and nonusers. However, a limitation of the present study is the lack of quantified daily measurements of alcohol consumption. Therefore, we cannot determine whether there is a threshold effect of alcohol consumption on immune responses. Approximately 25% of our subjects were considered heavy drinkers based on categories of alcohol consumption, and there was a suggestion that, in this small group, there may be higher HCV-specific IL-10 and TNF-α secretion, at least for NS5. This question would need to be addressed in cohorts with higher rates of heavy alcohol use.

In conclusion, antigen-specific IFN-γ immune responses are broadly diminished in persons with HIV/HCV coinfection, and HCV-specific TNF-α responses are also weaker in subjects with HIV/HCV, compared with subjects with HCV. This effect is modified by IDU and, to a lesser extent, by alcohol consumption, for which we found a relative increase in immune responses in substance users with HIV/HCV, compared with nonusers with HIV/HCV. Because most studies examining cellular immune responses in HIV/HCV and HCV focus on IFN-γ as the representative T cell effector function, these results suggest that IDU may be a confounding variable in these analyses. Future studies should examine mechanisms of action of IDU on HCV-specific immune responses as well as the effects of immune response alterations on HCV-related liver disease.

References

1. Di Martino V, Rufat P, Boyer N, et al. The influence of human immunodeficiency virus coinfection on chronic hepatitis C in injection drug users: a long-term retrospective cohort study. Hepatology **2001**; 34:1193–9.
2. Kwiatkowski CF, Fortuin Corsi K, Booth RE. The association between knowledge of hepatitis C virus status and risk behaviors in injection drug users. Addiction **2002**; 97:1289–94.
3. Romeo R, Rumi MG, Donato MF, et al. Hepatitis C is more severe in drug users with human immunodeficiency virus infection. J Viral Hepat **2000**; 7:297–301.
4. Telfer P, Sabin C, Devereux H, Scott F, Dusheiko G, Lee C. The progression of HCV-associated liver disease in a cohort of haemophilic patients. Br J Haematol **1994**; 87:555–61.
5. Pol S, Lamorthe B, Thi NT, et al. Retrospective analysis of the impact of HIV infection and alcohol use on chronic hepatitis C in a large cohort of drug users [comments]. J Hepatol **1998**; 28:945–50.
6. Bhattacharya R, Shuhart MC. Hepatitis C and alcohol: interactions, outcomes, and implications. J Clin Gastroenterol **2003**; 36:242–52.
7. McClain CJ, Barve S, Deaciuc I, Kugelmas M, Hill D. Cytokines in alcoholic liver disease. Semin Liver Dis **1999**; 19:205–19.
8. Neuman MG, Shear NH, Bellentani S, Tiribelli C. Role of cytokines in ethanol-induced cytotoxicity in vitro in Hep G2 cells. Gastroenterology **1998**; 115:157–66.
9. Perlemuter G, Letteron P, Carnot F, et al. Alcohol and hepatitis C virus core protein additively increase lipid peroxidation and synergistically trigger hepatic cytokine expression in a transgenic mouse model. J Hepatol **2003**; 39:1020–7.
10. Peterson JD, Herzenberg LA, Vasquez K, Waltenbaugh C. Glutathione levels in antigen-presenting cells modulate Th1 versus Th2 response patterns. Proc Natl Acad Sci USA **1998**; 95:3071–6.
11. Roy S, Wang J, Gupta S, Charboneau R, Loh HH, Barke RA. Chronic morphine treatment differentiates T helper cells to Th2 effector cells by modulating transcription factors GATA 3 and T-bet. J Neuroimmunol **2004**; 147:78–81.
12. Sacerdote P, Manfredi B, Gaspani L, Panerai AE. The opioid antagonist naloxone induces a shift from type 2 to type 1 cytokine pattern in BALB/cJ mice. Blood **2000**; 95:2031–6.
13. Stefano GB, Burrill JD, Labur S, Blake J, Cadet P. Regulation of various genes in human leukocytes acutely exposed to morphine: expression microarray analysis. Med Sci Monit **2005**; 11:MS35–42.
14. Tegeder I, Geisslinger G. Opioids as modulators of cell death and survival—unraveling mechanisms and revealing new indications. Pharmacol Rev **2004**; 56:351–69.
15. Kamal SM, Graham CS, He Q, et al. Kinetics of intrahepatic hepatitis C virus (HCV)–specific $CD4^+$ T cell responses in HCV and *Schistosoma mansoni* coinfection: relation to progression of liver fibrosis. J Infect Dis **2004**; 189:1140–50.
16. Graham CS, Wells A, Liu T, et al. Antigen-specific immune responses and liver histology in HIV and hepatitis C coinfection. AIDS **2005**; 19: 767–73.
17. Bradley KA, Bush KR, McDonell MB, Malone T, Fihn SD. Screening for problem drinking: comparison of CAGE and AUDIT. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. J Gen Intern Med **1998**; 13:379–88.

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. Anthony DD, Yonkers NL, Post AB, et al. Selective impairments in dendritic cell-associated function distinguish hepatitis C virus and HIV infection. J Immunol **2004**; 172:4907–16.
20. Graham CS, Curry M, He Q, et al. Comparison of HCV-specific intrahepatic CD4+ T cells in HIV/HCV versus HCV. Hepatology **2004**;40: 125–32.
21. Harcourt GC, Donfield S, Gomperts E, et al. Longitudinal analysis of CD8 T-cell responses to HIV and hepatitis C virus in a cohort of coinfected haemophiliacs. AIDS **2005**; 19:1135–43.
22. Lauer GM, Nguyen TN, Day CL, et al. Human immunodeficiency virus type 1-hepatitis C virus coinfection: intraindividual comparison of cellular immune responses against two persistent viruses. J Virol **2002**; 76:2817–26.
23. Valdez H, Anthony D, Farukhi F, et al. Immune responses to hepatitis C and non-hepatitis C antigens in hepatitis C virus infected and HIV-1 coinfected patients. AIDS **2000**; 14:2239–46.
24. Villacres MC, Literat O, Degiacomo M, et al. Reduced type 1 and type 2 cytokines in antiviral memory T helper function among women coinfected with HIV and HCV. J Clin Immunol **2005**; 25:134–41.
25. Kedzierska K, Crowe SM. Cytokines and HIV-1: interactions and clinical implications. Antivir Chem Chemother **2001**; 12:133–50.
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. Sindhu S, Toma E, Cordeiro P, Ahmad R, Morisset R, Menezes J. Relationship of in vivo and ex vivo levels of TH1 and TH2 cytokines with viremia in HAART patients with and without opportunistic infections. J Med Virol **2006**; 78:431–9.
28. Kamal SM, Fehr J, Roesler B, Peters T, Rasenack JW. Peginterferon alone or with ribavirin enhances HCV-specific CD4 T-helper 1 responses in patients with chronic hepatitis C. Gastroenterology **2002**; 123:1070–83.
29. Szabo G, Mandrekar P, Dolganiuc A, Catalano D, Kodys K. Reduced alloreactive T-cell activation after alcohol intake is due to impaired monocyte accessory cell function and correlates with elevated IL-10, IL-13, and decreased IFNgamma levels. Alcohol Clin Exp Res **2001**; 25:1766–72.
30. Dolganiuc A, Kodys K, Kopasz A, Marshall C, Mandrekar P, Szabo G. Additive inhibition of dendritic cell allostimulatory capacity by alcohol and hepatitis C is not restored by DC maturation and involves abnormal IL-10 and IL-2 induction. Alcohol Clin Exp Res **2003**; 27:1023–31.

Independent predictors of stroke in patients with atrial fibrillation

A systematic review

The Stroke Risk in Atrial Fibrillation Working Group*

CME

Address correspondence and reprint requests to Dr. R.G. Hart, University of Texas Health Science Center, 7703 Floyd Curl Dr., San Antonio, TX 78284
hartr@uthscsa.edu

ABSTRACT

Background: Absolute stroke rates vary widely among patients with nonvalvular atrial fibrillation. To balance the benefits and risks of chronic antithrombotic prophylaxis, it is important to estimate the absolute risk of stroke for individual patients.

Methods: Systematic review of studies using multivariate regression techniques to identify independent risk factors for stroke in patients with atrial fibrillation was conducted, and reports of absolute stroke rates in subgroups of patients with these risk factors collected. A summary estimate of the relative risk associated with each independent risk factor was calculated using maximum likelihood methods.

Results: Seven studies (including six entirely independent cohorts) were identified. Prior stroke/TIA (relative risk 2.5, 95% CI 1.8 to 3.5), increasing age (relative risk 1.5 per decade, 95% CI 1.3 to 1.7), a history of hypertension (relative risk 2.0, 95% CI 1.6 to 2.5), and diabetes mellitus (relative risk 1.7, 95% CI 1.4 to 2.0) were the strongest, most consistent independent risk factors. Observed absolute stroke rates for nonanticoagulated patients with single independent risk factors were in the range of 6 to 9% per year for prior stroke/TIA, 1.5 to 3% per year for history of hypertension, 1.5 to 3% per year for age >75, and 2.0 to 3.5% per year for diabetes. Female sex was inconsistently associated with stroke risk, whereas the evidence was inconclusive that either heart failure or coronary artery disease is independently predictive of stroke.

Conclusions: Four clinical features (prior stroke/TIA, advancing age, hypertension, diabetes) are consistent independent risk factors for stroke in atrial fibrillation patients. Prior stroke/TIA is the most powerful risk factor and reliably confers a high stroke risk (>5% per year, averaging 10% per year). Absolute stroke rates associated with other individual risk factors are difficult to precisely estimate from available data. ***Neurology*® 2007;69:546-554**

Among patients with atrial fibrillation, the absolute risk of stroke varies 20-fold depending on patient age and other clinical features. Accurate risk stratification is critical when balancing the potential benefits and risks of chronic antithrombotic therapy to prevent stroke. We systematically review clinical studies that identified patient features independently associated with stroke in patients with atrial fibrillation and summarize the absolute stroke rates associated with individual patient features.

METHODS The primary aim was to identify independent risk factors for ischemic stroke in nonanticoagulated patients with atrial fibrillation. We searched MEDLINE from January 1985 to October 2005 for all studies of patients with nonvalvular atrial fibrillation that determined factors independently associated with stroke using multivariate regression techniques. The search strategy (appendix E-1 on the *Neurology* Web site [www.neurology.org]) yielded 81 abstracts for initial consideration and was supplemented by personal files supplied by the members of the expert Working Group. Both prospective and retrospective studies were considered; studies in which >10% of the cohort received oral anticoagulants (e.g., oral vitamin K antagonists, ximelagatran) and studies in which >10% of patients had mitral stenosis or prosthetic cardiac valves were excluded. Studies in which participants took aspirin or other antiplatelet agents were included. The focus was to identify risk factors for ischemic stroke, but outcome events varied among studies and variably included all strokes (including hemorrhagic), ischemic strokes, strokes plus TIAs, and ischemic strokes plus non-CNS embolic events (together termed thrombo-

Supplemental data at www.neurology.org

*A list of members of the Stroke Risk in Atrial Fibrillation Working Group can be found in the Appendix.

From the University of Texas Health Science Center, San Antonio, TX.

The literature search was sponsored by the American Academy of Neurology Quality Standards Subcommittee as part of an update of its practice parameter on stroke prevention in atrial fibrillation.

Disclosure: The author reports no conflicts of interest.

embolism); these are described in the text and tables. Studies restricted to surrogate outcomes (e.g., left atrial thrombi or neuroimaging evidence of subclinical strokes) were not considered.

Analyses of independent echocardiographic and hematologic predictors were included only if analyzed with clinical risk factors. Studies with fewer than 30 stroke outcome events were excluded because accurate characterization of the predictive strength of multiple independent features would be weakened (i.e., potential for "overfitting" of the models). Cohorts that were subsequently included in larger analyses were not considered individually (e.g., the Stroke Prevention in Atrial Fibrillation (SPAF) I trial[1] placebo participants were included as part of the subsequent Atrial Fibrillation Investigators (AFI) 1994 pooled analysis)[2] with exceptions noted in the text and footnotes to the tables. Excluded studies and the reasons for exclusion are listed in appendix E-II. Variables included in the multivariate models for each study are given in appendix E-III; for selected variables (heart failure, coronary artery disease), specific definitions are also provided when available. Reported measures of effect (i.e., risk ratio, hazard ratio, odds ratio [OR], rate ratio) differed across studies and are specified in appendix E-III; because these measures are very similar when the outcome is infrequent, the term "relative risk" is used throughout. When not provided, the 95% CI for the relative risk was estimated using the Z statistic corresponding to the reported p value to estimate the variance.

The prevalence of individual risk factors across studies was estimated by combining the prevalences in each study weighted by its number of participants. To obtain an overall estimate of the relative risk associated with each risk factor, a maximum likelihood estimate that weighted each study inversely to its variance was computed after first testing for statistical homogeneity across studies assuming a fixed effects model.[3] All reported relative risks from multivariate models were included in the computation regardless of the statistical significance. In the few cases in which a relative risk was reported as nonsignificant in the multivariate model and not reported, the reported univariate relative risk estimate and 95% CI were used instead. To minimize bias, an overall estimate of relative risk was not computed for features that were not analyzed in or assessed by the majority of studies. Statistical significance was accepted at the 0.05 level, and analyses were done using SPSS 10.5 and Excel 2002 software.

A second aim was to comprehensively assess absolute stroke rates in the absence of anticoagulation for subgroups of atrial fibrillation patients with differing clinical risk factors. The clinical risk factors were taken from both the multivariate analyses and additional variables used in nine stroke risk prediction schemes.[3-11] Published rates (and 95% CIs when available), the number of participants, mean follow-up interval, and the number of outcome events were recorded (estimated for some rates, as indicated in the tables). Published stroke rates were not included when the number of outcome events and patient-years of exposure could not be determined unless a CI was also provided. Outcome rates from data sets that were subsequently included in larger analyses are not reported individually (e.g., patients involved in the SPAF III derivation scheme[4] were all included in the larger SPAF aspirin analysis)[5] with exceptions for overlapping cohorts noted in footnotes to the tables.

RESULTS Seven studies meeting the above criteria were identified (table 1). There were six entirely separate cohorts, whereas the seventh study had partial participant overlap (AFI 1994[2] with AFI echo).[13] These six patient cohorts included two from randomized clinical trials,[2,5] two prospective clinical cohorts,[14,16] one population-based epidemiologic study,[11] and one hospital-based case-control study[12](table 1). In all except the case-control study,[12] clinical and echocardiographic variables were assessed and recorded prior to the occurrence of stroke events. The largest study limited its multivariate analysis to assessing gender as an independent predictor of stroke.[16]

Mean patient ages ranged from 62 in the study of outpatients undergoing transesophageal echocardiography[14] to 79 in the hospital-based case-control study,[12] with an overall mean age of 71; 40% were women (table 1). Paroxysmal atrial fibrillation was present in 13 to 42% (mean = 30%) of the study cohorts. About half received aspirin (range 5 to 100%). Events rates in the longitudinal studies ranged from 2.5 to 4.7% per year during average follow-up intervals of 1.4 to 4.8 years. The number of thromboembolic events (>90% of which were strokes) ranged between 50 and 394 per study.

Independent risk factors for stroke. Eight clinical features predictive of stroke emerged from seven multivariate analyses of six cohorts (two studies had overlapping cohorts, in which echocardiographic features were considered in one[2,13] (appendix E-IV).

Prior stroke/TIA, present in 0 to 14% (mean 7%) of the cohort participants, was the strongest independent risk factor for subsequent stroke and was significant in all five studies (involving four separate cohorts) in which it was evaluated. Stroke risk increased 2.5-fold (95% CI 1.8 to 3.5) in patients with prior stroke/TIA (table 2).

Patient age was a consistent independent risk factor for stroke in each of the six studies in which it was assessed (mean ages ranged from 62 to 79 years in the cohorts), associated with an incremental increased risk of 1.5 (95% CI 1.3 to 1.7) per decade (table 2).

Blood pressure variables were consistent stroke predictors in all studies. A history of hypertension was present in 42 to 53% (mean = 48%) of the cohorts and was an independent risk factor for stroke in five of five studies (relative risk 2.0, 95% CI 1.6 to 2.5) and systolic blood pressure in two others (table 2). Both a history of hypertension and systolic blood pressure >160

Table 1 Characteristics of key studies

Author (ref. no.)	Study type	Population	n	Mean age, y/% women	2° PVT, %	Mean f/u, y	APT Rx,* %	Event type†	No. of events	Event rate, %/y
Multivariate analyses to identify independent stroke predictors										
Moulton et al. (12)	CC‡	Hospital based	N/A	79/54	0	N/A	13	IS	134	N/A
AFI 1994§ (2)	RTC	Untreated pts from 5 trials	1,593	69/27	6	1.4	~5	IS	108	~4.5
AFI Echo‖ (13)	RTC	Untreated pts from 3 trials with echo data	1,066	67/22	7	1.6	~10	IS	78	4.7
SPAF Aspirin (5)	RTC	SPAF I-III aspirin pts	2,012	69/28	8	2.0	100	IS	130	3.3
Stollberger et al. (14)	PC	TEE cohort	409	62/36	6	4.8	98	TE	50	2.5
Wang et al. (11)	ES	Framingham Study cohort	705	75/48	14	4.0	22	S	83	2.9
Fang et al. (16)	PC	HMO outpts	6,222	72/45	6	2.4	~48	TE	394	2.5
Additional studies reporting absolute event rates in patient subgroups										
Gage et al. (7)	HDC	Medicare pts	1,733	81/58	25	1.2	31	IS	71	3.3
Evans et al. (15)	HDC	Recent stroke	172	78/55	100	2.1	100	S	34	9.5
Go et al. (17)	PC	HMO outpts	5,089	71/45	4	2.4	~48	TE	249	2.0
AFI 2004 (18)	RTC	Aspirin pts from 6 trials, including SPAF III	2,580	72/37	22	1.9	100	IS	207	4.2
SPAF III (4)¶	RTC	SPAF I and II aspirin pts	854	69/31	7	2.3	100	TE	73	3.7
SPAF III High-Risk TEE cohort (19)	RTC	TEE subset of SPAF III RTC	382	71/30	40	1.1	100	IS	16	7.8
AFI 2002# (20)	RTC	Aspirin pts from six trials	2,113	72/39	24	1.9	100	IS	174	4.3
van Walraven et al.** (10)	RTC	Aspirin pts from 6 trials, including SPAF I-III	2,501	70/33	3	1.9	100	S	122	2.6
Tsivgoulis et al. (21)	PC	Recent stroke, >75y	135	81/56	100	7.2	100	TE	~45	16.3

*Percentage of the cohort receiving antiplatelet therapy. In Fang et al.,[16] 48% of patients were estimated to be taking aspirin based on a nested case-control assessment of a small number of patients.[17]

†S = ischemic, hemorrhagic, and unknowns; IS = ischemic stroke only, TE = thromboembolic events consisting of ischemic stroke and non-CNS emboli.

‡Retrospective design.

§The available data provided in this table are for the 1,236 controls for the warfarin-control comparison, but 1,593 patients (1,236 plus 357 SPAF group II controls) were used for the multivariate analysis.

¶These 854 participants are included in the larger SPAF aspirin cohort.[5]

‖Subset of patients included in AFI 1994[2] from three trials that collected precordial echocardiographic data.

**Includes 1,695 participants from the SPAF aspirin cohort,[5] comprising 68% the total; there were 1,322 patients in common between the AFI 2002 cohort[20] (total n = 2,113) and the van Walraven et al. cohort[10] (total n = 2,501).

AFI = Atrial Fibrillation Investigators collaboration; SPAF = Stroke Prevention in Atrial Fibrillation randomized trials; 2° PVT = secondary prevention (i.e., the fraction with prior stroke or TIA); outpt = outpatient derived; pts = patients; TEE = transesophageal echocardiography; APT Rx = antiplatelet treatment; f/u = follow-up; CC = case-control; PC = prospective cohort; ES = epidemiologic study; RTC = randomized trial cohort; HDC = hospital discharge cohort; NA = not applicable, y = year; ~ = approximate.

mm Hg were independently and separately associated with stroke in the SPAF aspirin cohort[5](appendix E-IV).

Diabetes mellitus was present in 14 to 18% (mean = 15%) of study cohorts and was a significant independent risk factor for stroke in four studies involving three separate cohorts (relative risk 1.7, 95% CI 1.4 to 2.0)[2,5,11,13] but not significant in another small study that did not report results that could be included in the pooled estimate[14] (table 2). No study included analyses of the impact of the severity, duration, or control of diabetes on stroke outcomes.

Clinical heart failure was present in 19 to 54% of study cohorts and was not an independent predictor of stroke in any of the three independent

Table 2 Independent clinical and echocardiographic predictors of stroke in patients with atrial fibrillation

Variable (ref. no.)	Multivariate RR (95% CI)	p Value	Comment
Prior stroke/TIA			
AFI 1994 (2)	2.5 (NR)	<0.05	
AFI Echo* (13)	3.5 (1.8-6.7)	<0.001	
SPAF Aspirin (5)	2.9 (NR)	<0.001	
Stollberger et al. (14)	3.7 (1.5-7.5)	0.002	
Wang et al. (11)	1.9 (1.1-3.3)	NR	
Moulton et al. (12)	NA	—	
Fang et al. (16)	NA	—	
Pooled estimate	2.5 (1.8-3.5)	—	$p = 0.59$ for heterogeneity
Hx hypertension			
AFI 1994 (2)	1.6 (NR)	<0.05	
AFI Echo* (13)	1.5 (0.9-2.5)	0.13	
SPAF Aspirin (5)	2.0 (NR)	<0.001	
Stollberger et al. (14)	3.6 (1.8-8.4)	0.001	
Wang et al. (11)	NA	—	Hx hypertension not assessed; use of antihypertensive drugs NS; see below for systolic BP >160 mm Hg.
Moulton et al.(12)	1.9 (1.2-3.1)	<0.05	
Fang et al. (16)	NA	—	
Pooled estimate	2.0 (1.6-2.5)	—	$p = 0.29$ for heterogeneity
Age			
AFI 1994 (2)	1.4/decade	<0.05	
AFI Echo* (13)	1.5/decade	0.006	
SPAF Aspirin (5)	1.8/decade	<0.001	
Stollberger et al. (14)	1.1 (1.0-1.1)	<0.001	Unclear whether per decade or per year
Wang et al. (11)	1.3/decade	<0.05	
Moulton et al. (12)	1.7 (1.0-2.8)	<0.05	For age >75
Fang et al. (16)	NA	—	
Pooled estimate	1.5 (1.3-1.7)/decade	—	$p = 0.59$ for heterogeneity
Diabetes			
AFI 1994 (2)	1.7 (NR)	<0.05	
AFI Echo* (13)	1.7 (1.0-2.9)	0.05	
SPAF Aspirin (5)	1.9 (NR)	0.02	
Stollberger et al. (14)	NR	NS	Univariate RR = 1.4 (0.7-2.8)
Wang et al. (11)	1.8 (1.4-3.1)	NR	
Moulton et al. (12)	NA	NA	Univariate OR = 1.2 (0.7-2.4)
Fang et al. (16)	NA	—	
Pooled estimate	1.7 (1.4-2.0)	—	$p = 0.69$ for heterogeneity
Heart failure			
AFI 1994 (2)	1.4 (NR)	>0.05	22% of the cohort; univariate RR = 1.7 (1.1-2.5)
AFI Echo* (13)	1.4 (0.8-2.3)	0.16	Multivariate HR = 1.7 (1.1-2.7) ($p = 0.03$) without echo data
SPAF Aspirin (5)	NR	NS	
Stollberger et al. (14)	NR	NS	Univariate RR = 3.1 (1.4-7.1) ($p = 0.008$) for class III-IV (10% of cohort); univariate RR = 1.3 (0.7-2.4) for class I-II (44% of cohort)
Wang et al. (11)	NA	—	Heart failure was combined with prior myocardial infarct
Moulton et al. (12)	NA	NA	29% of cases; univariate odds ratio = 1.7 (0.9-2.9)
Fang et al. (16)	NA	—	

—Continued

Table 2 *Continued*

Variable (ref. no.)	Multivariate RR (95% CI)	*p* Value	Comment
Abnormal left ventricular function by echocardiography			
AFI Echo* (13)	2.5 (1.5-4.4)	<0.001	Moderate-severe by 2D echocardiography
SPAF Aspirin (5)			
M mode	NR	NS	Fractional shortening ≤25%; univariate RR = 1.2 (p = 0.6)
2D	NR	NS	Moderate-severe by 2D; univariate RR = 1.2 (p = 0.5)
Stollberger et al. (14)	NR	NS	Fractional shortening continuous; univariate RR p = 0.008
Female gender			
AFI 1994 (2)	NR	NS	Univariate RR = 1.2 (0.8-1.8); not significant in multivariate models
AFI Echo* (13)	NR	NS	22% women in this subset
SPAF Aspirin (5)	1.6 (NR)	0.01	
Stollberger et al. (14)	NR	NS	
Wang et al. (11)	1.9 (1.2-3.1)	NR	
Moulton et al. (12)	NR	NS	Univariate OR = 1.6 (1.0-2.7); apparently NS in multivariate model
Fang et al. (16)	1.6 (1.3-1.9)	NR	
Women > 75 y			
SPAF Aspirin (5)	3.0 (NR)	0.002	Compared with low-risk men over age 75
Fang et al. (16)			
Women ≤75 y	1.6 (1.0-2.3)	NR	Rate ratio compared with men, adjusted for several other risk factors; the interaction between female sex and age was NS (p = 0.38)
Women > 75 y	1.8 (1.4-2.3)	NR	
Estrogen HRT			
SPAF Aspirin (5)	3.1 (NR)	0.007	HRT used in 33% of women at entry
Fang et al. (16)	1.0 (0.7-1.4)	NS	HRT used in 22% of women
Systolic BP > 160 mm Hg			
SPAF Aspirin (5)	2.3 (NR)	<0.001	
Wang et al. (11)	1.1/10 mm	<0.05	
Coronary artery disease			
AFI 1994 (2)			
Hx of MI	1.7 (NR)	NS	Hx MI (14% of cohort) has univariate RR = 1.7 (p = NR);
Hx of angina	NR	NS	Hx angina (23% of cohort) had univariate RR = 1.5 (p = NR)
AFI Echo* (13)	NA	—	
SPAF Aspirin (5)	NR	NS	Neither prior MI nor coronary artery disease had significant association.
Stollberger et al. (14)	NR	NS	Definite (25% of cohort) had univariate RR = 1.4 (0.7-2.8)
Wang et al. (11)	NA	—	Prior myocardial infarct combined with heart failure
Moulton et al. (12)	NA	NA	Hx MI (14% cases, 13% controls) had univariate OR = 1.1 (0.5-2.3); hx angina (24% cases, 15% controls) had univariate OR = 1.9 (1.0-3.5)
Fang et al. (16)	NA	—	
Carotid artery stenosis			
SPAF II (22)	1.8 (0.5-3.6)	0.55	
Left atrial thrombus			
Stollberger et al. (14)	2.4 (0.9-6.9)	0.09	Half of the 10 patients with left atrial thrombus were among the handful given anticoagulants, likely blunting the predictive value

Pooled estimates are calculated (see Methods) only when results from all studies that assessed the variable are available with one exception: for diabetes, a single small study[14] did not report a nonsignificant relative risk, but results of four larger studies were pooled.

*Subset of patients included in AFI 1994[2] from three trials that recorded precordial echocardiographic data and is not included separately in calculation of the pooled estimate.

NR = not reported; NS = not significant (p > 0.5); NA = not assessed; Hx = history of; pts = patients, 2D = two-dimensional; HRT = hormone replacement therapy; echo = echocardiography; HR = hazards ratio; RR = relative risk; OR = odds ratio; MI = myocardial infarction; SPAF = Stroke Prevention in Atrial Fibrillation clinical trial; AFI = Atrial Fibrillation Investigators.

cohorts in which it was analyzed (table 2). In a subgroup of the AFI pooled clinical trials cohort consisting of three trials with echocardiographic data, heart failure was independently associated with ischemic stroke (relative risk 1.7, 95% CI 1.1 to 2.7, $p = 0.03$), but this association was no longer significant after echocardiographic evidence of left ventricular dysfunction was added to the model[13] (appendix E-IV). The combination variable of recent (within 3 months) congestive heart failure or fractional shortening ≤25% by M-mode echocardiography (together classified as "abnormal left ventricular function") was a significant independent risk factor in the SPAF III derivation analysis,[4] but not in the larger SPAF aspirin cohort.[5] Moderate to severe left ventricular systolic dysfunction assessed by two-dimensional echocardiography was a powerful independent predictor in placebo-treated patients in the AFI analysis of echocardiographic features (relative risk = 2.5, 95% CI 1.5 to 4.4).[13] This was not confirmed by the SPAF aspirin analysis,[5] nor was reduced fractional shortening independently predictive of subsequent stroke in another study.[14] In short, clinical heart failure was not a consistent independent predictor in atrial fibrillation patients, and although one study found left ventricular systolic dysfunction by two-dimensional echocardiography to be strongly predictive in placebo-treated patients,[13] this was not confirmed in another study in which patients received aspirin.[5]

Female sex (22 to 54% of the study cohorts, mean = 40%) was an independent significant predictor of stroke in three studies involving entirely separate cohorts (range of individual relative risk 1.6 to 1.9),[5,11,16] but not in three others[2,14,23] (table 2). In the SPAF aspirin cohort,[5] women over age 75 were at particularly high risk, but an age/sex interaction was not confirmed in another larger study.[16]

Coronary artery disease (mean frequency = 25%) was not independently predictive of stroke in any of the three separate patient cohorts in which it was assessed[2,5,14] (table 2), although variably defined as prior myocardial infarction (mean frequency = 11%) or angina pectoris (mean frequency = 18%) with or without revascularization procedures (appendix E-III). Of note, coronary artery disease was a relatively weak, but significant, independent predictor of stroke (hazard ratio 1.3, 95% CI 1.0 to 1.5, $p = 0.02$) in one large study that did not meet criteria for inclusion in this analysis[24] (see the footnote to appendix E-IV).

Absolute stroke rates associated with high-risk features. All previously published observed stroke rates associated with specific risk factors and their combinations are given in appendix E-V. Several caveats apply. These rates are not adjusted for the influence of other concomitant risk factors. Quantitative pooling of rates from several studies may lead to misleading results due to varying prevalence of patients with prior stroke/TIA and other independent risk factors. Some rates are based on participants in randomized trials, whereas others are from clinical cohorts. In addition, different frequencies of antiplatelet therapy use and different outcome constellations (i.e., ischemic strokes only vs all strokes vs strokes plus systemic emboli vs strokes plus TIAs) introduce uncertainty. Some rates are based on small numbers of stroke events (e.g., the 5.9% per year stroke rate among diabetics in one report was based on three events among 29 patients).[10] Finally, lowest absolute rates of stroke arose in the largest study in which stroke detection during follow-up was based on administrative databases.[16,17]

Stroke rates for patients with prior stroke/TIA were consistently high in all reports, with an aggregate estimate of about 10% per year with aspirin therapy (appendix E-V). Several studies attempted to identify patients with atrial fibrillation with prior stroke/TIA who had relatively low stroke risks[5,18,25,26]: Stroke rates were 5.9% per year in men without hypertension or diabetes.[5,18] Whereas the time interval from the most recent stroke/TIA was inversely related to stroke rate, prior stroke/TIA occurring 1 to 3 years before study entry conferred high stroke rates.[5]

Patients with a history of hypertension had reported stroke rates ranging from 2.6[17] to 5.8% per year,[20] but the latter analysis included nearly one-fourth of patients with prior stroke/TIA (appendix E-V). In one study of 612 otherwise low-risk patients with hypertension (age ≤75, systolic blood pressure <160 mm Hg, no diabetes or prior stroke/TIA), the observed stroke rate was 2.6% per year.[5] In a large outpatient-derived cohort with low frequencies of prior stroke/TIA (4%) and diabetes (15%), the observed stroke rate was 2.6% (CI 2.2 to 3.0) per year for those with hypertension.[17]

For patients without prior stroke/TIA, stroke rates for those age >75 ranged from 3.2[17] to 5.2%[5] per year (appendix E-V). Among men age >75 without prior stroke/TIA, two studies found relatively low stroke rates (2.8[16] and 3.2%[5] per year), and for such patients without hypertension

or diabetes, the stroke rate was 1.6% per year (95% CI 0.7 to 3.9%).[5] In two studies of patients older than age 80 without other risk factors (including coronary artery disease), stroke rates were 1.5[10] and 3.0%[2] per year, albeit based on relatively few events (appendix E-V). For those ages 70 to 79 with "lone atrial fibrillation," the ischemic stroke rate was 2.1% per year (95% CI 0.8 to 5.6) among patients in the placebo or control arms of clinical trials.[2]

Diabetic patients with atrial fibrillation had reported stroke rates ranging from 3.6[17] to 8.6%[2] per year (appendix E-V). In one study of 196 diabetic individuals without hypertension or prior stroke/TIA and who were not women over age 75, the observed stroke rate was 2.6% per year.[5]

The stroke rate among women with atrial fibrillation ranged from 3.5[16] to 5.8% per year[2] (appendix E-V); in two studies of otherwise low-risk women under age 75, the stroke rates were 0.6% per year[16] and 1.7% per year.[10]

Stroke rates among those with coronary artery disease ranged from 2.9[17] to 8.2%[2] per year, and patients with heart failure had stroke rates ranging from 3.5[17] to 6.8%[2] per year. The stroke rate associated with coronary artery disease and with heart failure in the absence of other risk factors could not be estimated from available data.

DISCUSSION Prior stroke/TIA, hypertension, advancing age, and diabetes are consistent independent predictors of stroke in patients with atrial fibrillation. Despite differences in the six cohorts (e.g., hospital based vs outpatients vs pooled clinical trial participants), there is overall consistency in the effects of these independent risk factors (table 2). Evidence is less robust for female sex. Although several studies have reported that diagnosed heart failure and/or left ventricular systolic dysfunction by echocardiography are risk factors for stroke in patients with atrial fibrillation, the overall evidence of an independent predictive effect is inconsistent. There is little compelling evidence supporting coronary heart disease as an independent effect predictor of stroke in atrial fibrillation patients. Of note, paroxysmal (or intermittent) atrial fibrillation did not emerge as independently predictive of stroke in any of the four studies in which it was assessed.[2,5,12,14] Varying frequencies of the use of antiplatelet drugs among studies did not appear to influence the results of multivariate analyses to identify independent predictors (e.g., aspirin was used in about 5% of patients in the AFI cohort[2] vs 100% of those in the SPAF aspirin cohort,[5] yet significant independent predictors were similar [table 2]).

Prior stroke/TIA is associated with high rates of stroke even in atrial fibrillation patients without other factors.[5,21,25] Because prior stroke/TIA is a marker of an underlying pathophysiologic cause and not the cause per se, this observation implies that additional important factors contributing to stroke risk in patients with atrial fibrillation remain to be elucidated. All patients with atrial fibrillation with prior stroke/TIA should be considered at high risk and treated with anticoagulation if it can be given safely. The absolute all-cause stroke risk reduction by adjusted-dose warfarin (about 6% per year) substantially exceeds the incremental increase in major hemorrhage (about 1.5% per year) based on pooled results of clinical trials.[20]

Taken individually, the remaining risk factors conferred stroke rates of 1.5 to 3.5% per year, but available data are limited. It is unlikely that a single rate adequately characterizes the magnitude of risk for all patients with a given risk factor: The severity of hypertension and duration of diabetes, for example, probably influence the absolute stroke rates in patients with these disorders.

The higher absolute stroke rates among very elderly patients with atrial fibrillation appear to be explained in part by the coexistence of other independent risk factors associated with age. Although it may be possible to identify relatively low-risk patients over age 75, the wide CIs around existing rate estimates warrant caution.

Female sex has been associated with an increased risk of stroke in atrial fibrillation patients in additional studies that did not meet inclusion criteria for these analyses[25,28-30] (appendix E-II). In one study, female sex lost its independent predictive effect on stroke risk when adjusted for systolic blood pressure[24]; this adjustment could not be done in one large recent study[16] and was associated with an age interaction in another.[5] Full characterization of the independent contribution of female sex to stroke risk in atrial fibrillation patients requires additional study.

Although the presence of heart failure would be expected to increase the risk of stroke in atrial fibrillation patients based on current pathophysiologic concepts, the available evidence does not provide support for this association (table 2). The clinical diagnosis of heart failure in elderly people with atrial fibrillation is particularly difficult, possibly resulting in misclassification and blunting of the predictive power. Although cardiologists were involved in the diagnosis of heart

failure in two of the studies that failed to demonstrate an association,[5,14] clinical criteria per se may be insufficient to reliably diagnose this condition (particularly diastolic dysfunction). Hypothetically, stringent diagnostic criteria that include echocardiographic evidence of left ventricular systolic impairment may be required.[5,13]

The stroke rates in recent clinical trials[31-34] appear to be lower than in clinical trials completed 15 years ago.[2] Better control of blood pressure may be contributing to lower stroke rates,[24] as even modest blood pressure lowering appears to have a substantial impact on stroke risk in atrial fibrillation patients.[27] Whether absolute stroke rates in high-risk subgroups of atrial fibrillation patients are now lower than 10 to 15 years ago cannot be definitively addressed by examining the reported rates owing to confounding by uncertain combinations of risk factors, varying frequencies of the use of aspirin and other cardiovascular medications, and the potential play of chance.

The most consistent independent risk factors for stroke emerging from these multivariate analyses (prior stroke/TIA, advancing age, hypertension, and diabetes) are also risk factors for stroke among persons without atrial fibrillation and consequently do not specifically implicate a cardioembolic mechanism for ischemic strokes. However, the absolute stroke rates associated with these risk factors are several times higher among patients with atrial fibrillation as compared with their effects in patients without atrial fibrillation. Hypertension is associated with stasis of flow and thrombus in the left atrial appendage and with cardioembolic strokes in atrial fibrillation patients.[35-37] Advancing age is independently associated with reduction of left atrial appendage flow velocities.[37] Diabetes hypothetically could be a predictor of noncardioembolic strokes that elderly atrial fibrillation patients experience, supported by a relatively smaller reduction in stroke in diabetic atrial fibrillation patients by warfarin in some studies,[20] but not in others.[17] Overall, the substantial reduction of ischemic strokes by anticoagulation over aspirin in atrial fibrillation patients with each of these risk factors[17,20] differs from warfarin's lack of relative benefit in patients with noncardioembolic cerebrovascular disease[38-40] and lends indirect support to a cardioembolic etiology.

Although these analyses summarize the best available evidence about independent risk factors for stroke in patients with atrial fibrillation, it would be premature to alter or abandon existing risk stratification schemes before wide validation of modifications or of new schemes. Optimal identification of low-risk atrial fibrillation patients who may not benefit substantially from anticoagulation is the focus of ongoing analyses by the Stroke Risk in Atrial Fibrillation Working Group. In addition, it is worth emphasizing that clinical decisions regarding antithrombotic therapy for patients with atrial fibrillation should incorporate other factors in addition to the estimated stroke risk, including bleeding risks if anticoagulated, the patient's ability to tolerate sustained therapy, access to high-quality anticoagulation monitoring, and patient-specific values and preferences.

APPENDIX

Stroke Risk in Atrial Fibrillation Working Group: Robert G. Hart, M.D. (San Antonio, TX, USA); Lesly A. Pearce, M.S. (Minot, ND, USA); Gregory W. Albers, M.D. (Palo Alto, CA, USA); Stuart J. Connolly, M.D. (Hamilton, ON, Canada); Gary H. Friday, M.D. (Philadelphia, PA, USA); Brian F. Gage, M.D. (St. Louis, MO, USA); Alan S. Go, M.D. (Oakland, CA, USA); Larry B. Goldstein, M.D. (Durham, NC, USA); Gary Gronseth, M.D. (Kansas City, KN, USA); Jonathan L. Halperin, M.D. (New York, NY, USA); Elaine M. Hylek, M.D., M.P.H. (Boston, MA, USA); Gregory Y. H. Lip, M.D. (Birmingham, UK); David G. Sherman, M.D. (San Antonio, TX, USA); Daniel E. Singer, M.D. (Boston, MA, USA); Carl van Walraven, M.D., M.Sc. (Ottawa, ON, Canada).

Received November 6, 2006. Accepted in final form March 12, 2007.

REFERENCES

1. Stroke Prevention in Atrial Fibrillation Investigators. The Stroke Prevention in Atrial Fibrillation Study: final results. Circulation 1991;84:527–539.
2. 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;154:1449–1457.
3. Normand ST. Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med 1999;19:321–359.
4. Stroke Prevention in Atrial Fibrillation Investigators. Risk factors for thromboembolism during aspirin therapy in patients with atrial fibrillation: the Stroke Prevention in Atrial Fibrillation Study. J Stroke Cerebrovasc Dis 1995;5:147–157.
5. Hart RG, Pearce LA, McBride R, et al. Factors associated with ischemic stroke during aspirin therapy in atrial fibrillation: analysis of 2012 participants in the SPAF I-III clinical trials. The Stroke Prevention in Atrial Fibrillation (SPAF) Investigators. Stroke 1999; 30:1223–1229.
6. Fuster V, Ryden LE, Asinger RW, et al. ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: executive summary. Circulation 2001;104:2118–2150.
7. Gage BF, Waterman AD, Shannon W, et al. Validation of clinical classification schemes for predicting stroke: results of the National Registry of Atrial Fibrillation. JAMA 2001;285:2864–2870.
8. Albers GW, Dalen JE, Laupacis A, et al. Antithrombotic therapy in atrial fibrillation. Chest 2001;119:194S–206S.

9. Singer DE, Albers GW, Dalen JE, et al. Antithrombotic therapy in atrial fibrillation. The Seventh ACCP Conference. Chest 2004;126(suppl):429S–456S.
10. van Walraven C, Hart RG, Wells GA, et al. A clinical prediction rule to identify patients with atrial fibrillation and a low risk for stroke while taking aspirin. Arch Intern Med 2003;163:936–343.
11. Wang TJ, Massaro JM, Levy D, et al. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community. The Framingham Study. JAMA 2003;290:1049–1056.
12. Moulton AW, Singer DE, Haas JS. Risk factors for stroke in patients with nonrheumatic atrial fibrillation: a case-control study. Am J Med 1991;91:156–161.
13. Atrial Fibrillation Investigators. Echocardiographic predictors of stroke in patients with atrial fibrillation. A prospective study of 1,066 patients from 3 clinical trials. Arch Intern Med 1998;158:1316–1320.
14. Stollberger C, Chnupa P, Kronik G., et al. Transesophageal echocardiography to assess embolic risk in patients with atrial fibrillation. Ann Intern Med 1998;128: 630–638.
15. Evans A, Perez I, Yu G, et al. Should stroke subtype influence anticoagulation decisions to prevent recurrence in stroke patients with atrial fibrillation? Stroke 2001;32:2828–2832.
16. Fang MC, Singer DE, Chang Y, et al. Gender differences in the risk of ischemic stroke and peripheral embolism in atrial fibrillation: the AnTicoagulation and Risk factors In Atrial fibrillation (ATRIA) Study. Circulation 2005;112:1687–1692.
17. Go AS, Hylek EM, Chang Y, et al. Anticoagulation therapy for stroke prevention in atrial fibrillation. How well do randomized trials translate into clinical practice? JAMA 2003;290:2685–2692.
18. Gage BF, van Walraven C, Pearce LA, et al. Selecting patients with atrial fibrillation for anticoagulation. Stroke risk stratification in patients taking aspirin. Circulation 2004;110:2287–2292.
19. Stroke Prevention in Atrial Fibrillation Investigators Committee on Echocardiography. Transesophageal echocardiographic correlates of thromboembolism in high-risk patients with atrial fibrillation. Ann Intern Med 1998;128:639–647.
20. van Walraven C, Hart RG, Singer DE, et al. Oral anticoagulants vs. aspirin in nonvalvular atrial fibrillation. An individual patient meta-analysis. JAMA 2002;288: 2441–2448.
21. Tsivgoulis G, Spengos K, Zakapoulos N, et al. Efficacy of anticoagulation for secondary stroke prevention in older people with nonvalvular atrial fibrillation: a prospective case series study. Age Ageing 2005;34:35–40.
22. Kanter MC, Tegeler CH, Pearce LA, et al. Carotid stenosis in atrial fibrillation: prevalence, risk factors and relationship to stroke. Arch Intern Med 1994;154:1372–1377.
23. Hu DY, Sun YH, Zhou ZQ, et al. Risk factors for stroke in Chinese with nonvalvular atrial fibrillation: a case control study [Chinese with English abstract]. Chin J Intern Med (Chung-Hua Nei Ko Tsa Chih) 2003;42:157–161.
24. Miyasaka Y, Barnes ME, Gersh BJ, et al. Time trends of ischemic stroke incidence and mortality in patients diagnosed with first atrial fibrillation in 1980 to 2000. Report of a community-hased study. Stroke 2005;36:2362–2366.
25. van Latum JC, Koudstaal PJ, Venables GS, et al. Predictors of major vascular events in patients with a transient ischemic attack or minor ischemic stroke with nonrheumatic atrial fibrillation. Stroke 1995;16:801–806.
26. Hart RG, Pearce LA, Koudstaal PJ. Transient ischemic attacks in patients with atrial fibrillation: implications for secondary prevention. Stroke 2004;35:948–951.
27. Arima H, Hart RG, Colman S, et al. Perindopril-based blood pressure lowering reduces major vascular events in patients with atrial fibrillation and prior stroke or TIA. Stroke 2005;36:2164–2169.
28. Boysen G, Nyhoe J, Appleyard M, et al. Stroke incidence and risk factors for stroke in Copenhagen, Denmark. Stroke 1988;19:1345–1353.
29. Cabin HS, Clubb KS, Hall C, et al. Risk for systemic embolization of atrial fibrillation without mitral stenosis. Am J Cardiol 1990;65:1112–1116.
30. Friberg J, Scharling H, Gadsboll N, et al. Comparison of the impact of atrial fibrillation on the risk of stroke and cardiovascular death in women versus men (Copenhagen Heart Study). Am J Cardiol 2004;94:889–894.
31. ACTIVE Writing Committee on behalf of the ACTIVE Investigators. Clopidogrel plus aspirin versus oral anticoagulation for atrial fibrillation. Atrial fibrillation Clopidogrel Trial with Irbesartan for prevention of Vascular Events (ACTIVE-W). Lancet 2006;367:1903–1912.
32. Sherman DG, Kim SG, Boop BS, et al. The occurrence and characteristics of stroke events in the AFFIRM Study. Arch Intern Med 2005;165:1185–1191.
33. Olsson SB. Executive Steering Committee on behalf of the SPORTIF III Investigators. Stroke prevention with the oral direct thrombin inhibitor ximegalatran compared with warfarin in patients with non-valvular atrial fibrillation: randomized controlled trial. Lancet 2003;362:1691–1698.
34. SPORTIF Executive Steering Committee for the SPORTIF V Investigators. Ximelagatran vs. warfarin for stroke prevention in patients with nonvalvular atrial fibrillation. JAMA 2005;293:690–698.
35. Miller VT, Rothrock JF, Pearce LA, et al. Ischemic stroke in patients with atrial fibrillation: effect of aspirin according to stroke mechanism. Neurology 1993;43:32–36.
36. Zabalgoitia M, Halperin JL, Pearce LA, et al. Transesophageal echocardiographic correlates of clinical risk of thromboembolism in nonvalvular atrial fibrillation. J Am Coll Cardiol 1998;31:1622–1626.
37. Goldman ME, Pearce LA, Hart RG, et al. Pathophysiologic correlates of thromboembolism in nonvalvular atrial fibrillation. Reduced flow velocity in the left atrial appendage. J Am Soc Echo 1999;12:1080–1087.
38. Mohr J, Thompson JLP, Lazar RM, et al. A comparison of warfarin and aspirin for prevention of recurrent ischemic stroke. N Engl J Med 2001;345:1444–1451.
39. Stroke Prevention in Reversible Ischemia Trial (SPIRIT) Study Group. A randomized trial of anticoagulants versus aspirin after cerebral ischemia of presumed arterial origin. Ann Neurol 1997;42:857–865.
40. Chimowitz MI, Lynn MJ, Howlett-Smith H, et al. Comparison of warfarin and aspirin for symptomatic intracranial arterial stenosis. N Engl J Med 2005;352: 1305–1316.

BMC Medical Research Methodology

BioMed Central

Research article

Open Access

A cautionary note regarding count models of alcohol consumption in randomized controlled trials

Nicholas J Horton*[1], Eugenia Kim[1] and Richard Saitz[2,3,4]

Address: [1]Department of Mathematics and Statistics, Smith College, Northampton, MA, USA, [2]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA, [3]Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA and [4]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

Email: Nicholas J Horton* - nhorton@email.smith.edu; Eugenia Kim - ekim@email.smith.edu; Richard Saitz - rsaitz@bu.edu

* Corresponding author

Published: 15 February 2007

BMC Medical Research Methodology 2007, **7**:9 doi:10.1186/1471-2288-7-9

Received: 24 July 2006
Accepted: 15 February 2007

This article is available from: http://www.biomedcentral.com/1471-2288/7/9

Abstract

Background: Alcohol consumption is commonly used as a primary outcome in randomized alcohol treatment studies. The distribution of alcohol consumption is highly skewed, particularly in subjects with alcohol dependence.

Methods: In this paper, we will consider the use of count models for outcomes in a randomized clinical trial setting. These include the Poisson, over-dispersed Poisson, negative binomial, zero-inflated Poisson and zero-inflated negative binomial. We compare the Type-I error rate of these methods in a series of simulation studies of a randomized clinical trial, and apply the methods to the ASAP (Addressing the Spectrum of Alcohol Problems) trial.

Results: Standard Poisson models provide a poor fit for alcohol consumption data from our motivating example, and did not preserve Type-I error rates for the randomized group comparison when the true distribution was over-dispersed Poisson. For the ASAP trial, where the distribution of alcohol consumption featured extensive over-dispersion, there was little indication of significant randomization group differences, except when the standard Poisson model was fit.

Conclusion: As with any analysis, it is important to choose appropriate statistical models. In simulation studies and in the motivating example, the standard Poisson was not robust when fit to over-dispersed count data, and did not maintain the appropriate Type-I error rate. To appropriately model alcohol consumption, more flexible count models should be routinely employed.

Background

Count outcomes are common in randomized studies of alcohol treatment. Subjects may be queried about their daily consumption of alcohol, measured as a number of drinks over a recent period [1] (typically 30 days), and these values are used to estimate average drinking per day. In this setting, estimating differences between treatment group and control group is of primary interest.

A challenge in modeling consumption outcomes is to appropriately account for the distribution of drinking. These distributions are characterized by a large number of zeros (abstinent subjects) along with a long right tail

(heavy drinking subjects). An extensive literature describes models for counts [2-8], and they have been commonly applied in economic analyses, traffic accidents, and health services utilization. Many routines are now available in general purpose statistical software (e.g. Stata) [8]. A natural model for counts is the single-parameter Poisson distribution. One disadvantage of the Poisson is that it makes strong assumptions regarding the distribution of the underlying data (in particular, that the mean equals the variance). While these assumptions are tenable in some settings, they are less appropriate for alcohol consumption. Extensions of the Poisson, such as the over-dispersed Poisson, negative binomial and two stage (hurdle) or zero inflated models have been proposed [2-5].

Our methods are motivated by the analysis of the ASAP (Addressing the Spectrum of Alcohol Problems) study, a randomized clinical trial comparing a brief motivational interview to usual care for a sample of inpatients with unhealthy alcohol use at an urban hospital [9]. These subjects were followed to see if there were differences in drinking outcomes that could be attributed to randomized group assignment.

In this paper, we will demonstrate the limitations of the standard Poisson model in the presence of over-dispersion. We begin by describing several count models for alcohol outcomes, compare their performance in a series of simulated randomized trials, apply them to the ASAP study, and conclude with some general recommendations.

Methods

Statistical methods for the analysis of count outcomes

We begin by introducing notation to be used throughout. Let Y_{ij} denote the number of events for the jth subject (j = 1,..., n_i) in the ith group (i = 1, 2), where n_i is the number of subjects in the ith group. Typically in a randomized trial n_1 and n_2 are approximately equal.

The Poisson distribution is one of the simplest models for count data. Let λ_{ij} indicate the average number of events (in this case drinks consumed) in a given time interval for subject j in group i, where $f(Y_{ij} = k|\lambda_{ij})$ is the probability of observing k events. The Poisson distribution [8,10] is denoted:

$$P(Y_{ij} = k \mid \lambda_{ij}) = \frac{\exp(-\lambda_{ij})\lambda_{ij}^{k}}{k!}$$

for k = 0, 1, 2, ..., i = 1, 2, and j = 1,..., n_i where $\lambda_{ij} > 0$ and we assume that $\lambda_{ij} = \lambda_i$ for all j (i.e. all subjects in a given group have the same rate of drinking). The λ parameter uniquely specifies this distribution, and is equal to the expected value (mean) and variance (i.e. $E[Y_{ij}] = Var(Y_{ij}) = \lambda_{ij}$ for all i and j). The maximum likelihood estimate (MLE) of $\hat{\lambda}_i$ is given by $\overline{Y}_i$. In this setting, the test of randomized group effects for the Poisson model is a test of the null hypothesis that $\lambda_1 = \lambda_2$.

One limitation of this model is that it may be overly simplistic and may not provide an adequate fit to consumption data of the type that we consider. The constraint that the variance is equal to the mean may lead to incorrect test results.

Consider as an example the data from the ASAP study control group at 3 months. For this dataset, non-integer count values are possible. These arise when subjects consume a number of drinks not divisible by 30 (in the case of 30-day assessments). One approach in this situation would be to model the number of drinks consumed in a 30 day period, or utilize the non-integer values. Sometimes even the 30 day value is non-integer because people report a drink size that is then translated into standard drinks. The maximum likelihood estimates of the probability distributions remains the same for non-integer values, though it is necessary to move each non-integer observed value to the next integer (using a ceiling function) to be plotted. For the models that we discuss, we can plug non-integer values into the software and still get sensible results.

Figure 1 displays the observed distribution and superimposed Poisson with $\hat{\lambda}_1 = \overline{Y}_1 = 4.98$ using the prcounts routine in Stata [8]. The axis for the number of drinks per day after 3 months was limited to 25 drinks to improve readability (the maximum observed count was 48.6). There is a pronounced lack of fit to this model, particularly for values of less than 10 drinks per day. For the ASAP data, the assumption that the mean is equal to the variance is not tenable. In fact, the observed variance (71.7) is more than an order of magnitude larger than the mean. Also, note that there is some evidence for digit preference (even numbers are more common than odd numbers).

One approach to loosen the restrictive variance assumption involves use of an empirical (or *robust* or *sandwich*) variance estimator [11-13] to account for the over-dispersion. This more flexible extension of the Poisson allows the variance to be unconstrained. The over-dispersed Poisson option is available in a number of general purpose statistics packages (e.g. the robust option in Stata).

Another approach is to fit a negative binomial (two parameter) count model (NB) [5-8,10]. One common

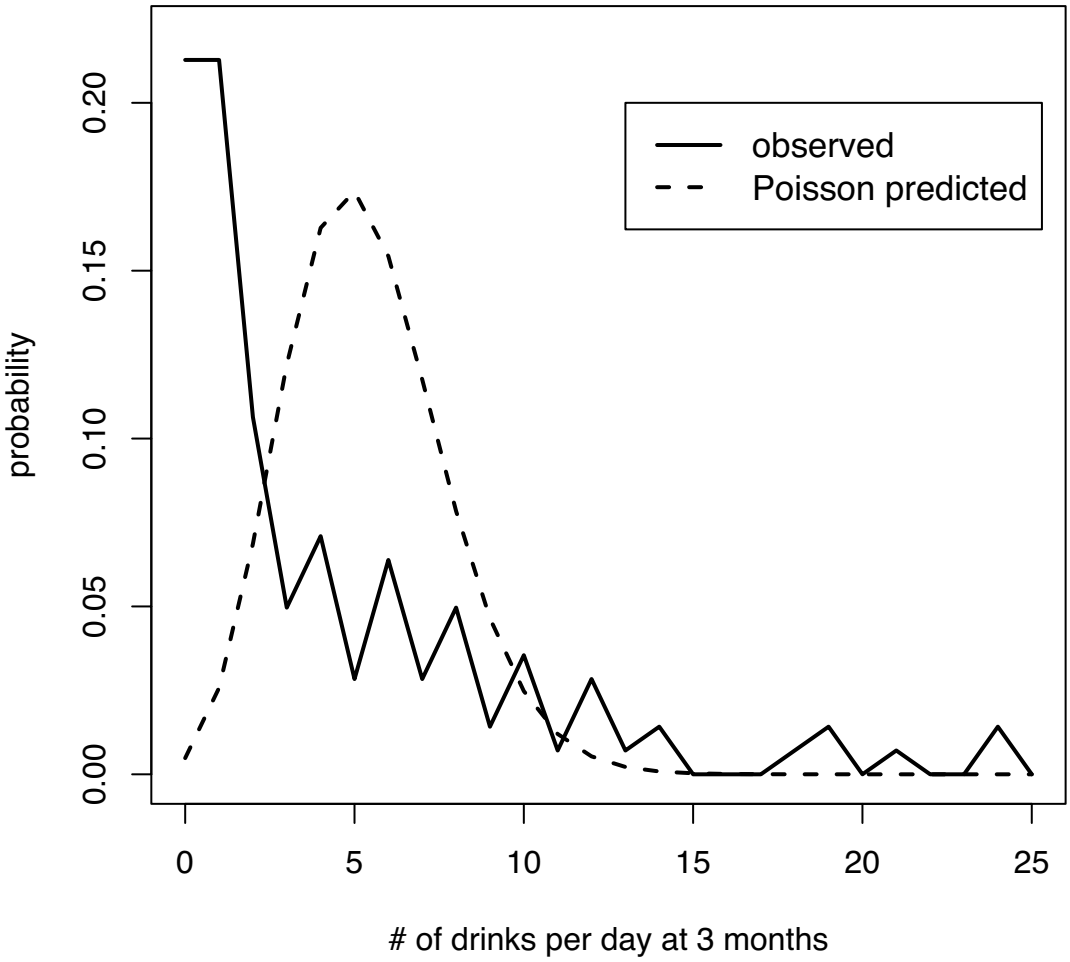


Figure 1

Observed value of drinks per day for the control group of the ASAP study at 3 months, plus the estimated Poisson fit to these data ($\hat{\lambda}_1 = 4.98$).

parametrization of the negative binomial distribution is given by:

$$P(Y_{ij} = k \mid \lambda_i, \theta_i) = \frac{\Gamma(\theta_i^{-1} + k)}{\Gamma(\theta_i^{-1})\Gamma(k+1)} \left(\frac{\theta_i^{-1}}{\theta_i^{-1} + \lambda_i} \right)^{\theta_i^{-1}} \left(\frac{\lambda_i}{\theta_i^{-1} + \lambda_i} \right)^{k}$$

where $\Gamma(\cdot)$ denotes the Gamma function, $\lambda_i > 0$ and $\theta_i > 0$. We note that $E[Y_{ij}] = \lambda_i$ and $Var(Y_{ij}) = \lambda_i + \lambda_i^2 * \theta_i = \lambda_i * (1 + \lambda_i * \theta_i)$ for all i and j and that $Var(Y_{ij}) > E[Y_{ij}]$. It can be shown that the negative binomial can be derived in terms of a Poisson random variable where the parameter λ_i varies according to a gamma distribution.

The negative binomial model is attractive because it allows the relaxation of strong assumptions regarding the relationship between the mean and the variance. This flexibility comes at some cost, since a two-parameter model is inherently more complicated to interpret.

Other models have been proposed that allow for an extra abundance of subjects with no consumption. In alcohol consumption outcomes, there may be subjects who are "non-susceptible" (e.g. abstinent). These "zero-inflation" (or "hurdle") models account for subjects who are structural zeros (e.g., abstinent subjects thought of as "non-susceptible") [2,3]. Conditional on being susceptible (with some probability), the distribution is assumed to be Poisson or negative binomial.

Zero-inflated Poisson (ZIP) models [3] separately estimate a parameter p_i that governs the proportion of non-susceptible subjects in the ith group:

$$f(Y_{ij} = k \mid \lambda_i, p_i) = I(k = 0)p_i + (1 - p_i)\frac{\exp(-\lambda_i)\lambda_i^k}{k!},$$

for $0 < p_i < 1$ and $\lambda_i > 0$ where $I(k = 0)$ is equal to 1 when $k = 0$, and equal to 0 otherwise. By distinguishing *Always-0* (with probability p_i) and *Not Always-0* group (with probability $(1 - p_i) * exp(-\lambda_i)$) for abstainers and drinkers who didn't drink during the reporting period, respectively, it can incorporate an overabundance of zeros [8]. Conditional on being a *Not Always-0*, counts are given by the Poisson distribution. This approach has been generalized to a regression framework, and implemented in general purpose statistical software (e.g. zip in Stata).

In many settings, the assumption that after accounting for the zeros the remaining counts are Poisson may not be tenable. The zero-inflated negative binomial (ZINB) allows for over-dispersion in this manner, though at the cost of more parameters.

Another approach to the modeling of count data involves use of a linear model (assuming that the observations are approximately Gaussian). While this is an extremely flexible model that is typically robust to misspecification (since the mean and variance are not linked), the linear model is less attractive because it may predict negative values of drinking given the skewness of the distribution. Use of a linear model is also inefficient if the variance is a function of the mean.

Simulation study

To better understand the behavior of these methods in a known situation, we conducted a series of simulation studies with parameters derived from the motivating example. These simulation studies were designed to address the question of whether or not the models were robust to misspecification of the underlying count distribution. More formally, we wanted to assess whether these models preserved the appropriate Type-I error rate (the probability of rejecting the null hypothesis when it is true) when there are no true differences between groups (i.e. do they reject the null at the appropriate α level).

For each set of parameters within a simulation, 100 observations were generated in each of two groups, to mimic a randomized clinical trial setting. The amount of alcohol consumption, in drinks per day was the outcome. For each simulated dataset a series of models (Poisson, negative binomial and zero-inflated Poisson) were fit. This process was repeated 2500 times for each set of parameters, where $E[Y_i] = \lambda = 5$ (taken from the ASAP control group) and an α level of 0.05 was used. For the simulation of Poisson data the variance was equal to the mean. Negative binomial distributions were simulated using three arbitrary variances (13.3, 40 and 70), with the latter value comparable to the observed variance from the ASAP control group. The zero-inflated model had a probability of 0.2 of being a structural zero, and Poisson with $\lambda = 5$ otherwise. The true distributions for the simulations are displayed in Figure 2. Models were fit using the Poisson distribution, over-dispersed Poisson using an empirical variance estimator, negative binomial and zero inflated Poisson. We estimated the probability that each model rejected the null hypothesis and constructed a 99% confidence interval around this estimate. The code for the simulations is available upon request from the first author.

ASAP study

The ASAP study was a randomized clinical trial of the effectiveness of a brief motivational intervention [14] on alcohol consumption among a group of hospitalized patients at Boston Medical Center. Details of the recruitment procedures, inclusion criteria, description of sample and results of the RCT have been published [15]. The Institutional Review Board of Boston University Medical Center approved this study, and the Institutional Review Board of Smith College approved the secondary analyses. After consenting to enroll, all subjects received an interviewer-administered baseline assessment prior to randomization into the control or intervention group. Subjects were randomly assigned to control or intervention group using a blocked randomization procedure. Intervention subjects participated in a brief motivational interview with a counselor (less than half an hour). Control subjects received usual care.

Follow-up was planned at 3-month and 12-month timepoints. Because the subjects came from a transient and hard-to-reach population, the researchers employed exhaustive techniques to track subjects over the follow-up period. The two primary alcohol-related outcomes were measures of alcohol consumption and linkage to appropriate alcohol treatment; for these secondary analyses we focus solely on treatment differences in alcohol consumption. The outcome of interest was the average number of standard drinks consumed per day in the past thirty days as reported using the Timeline Followback method [1] at the 3 and 12-month interviews. For the purpose of this secondary analysis we consider the 3 month time point; similar results were seen utilizing 12 month data (not reported here).

Eight models were fit comparing treatment to control for the ASAP study:

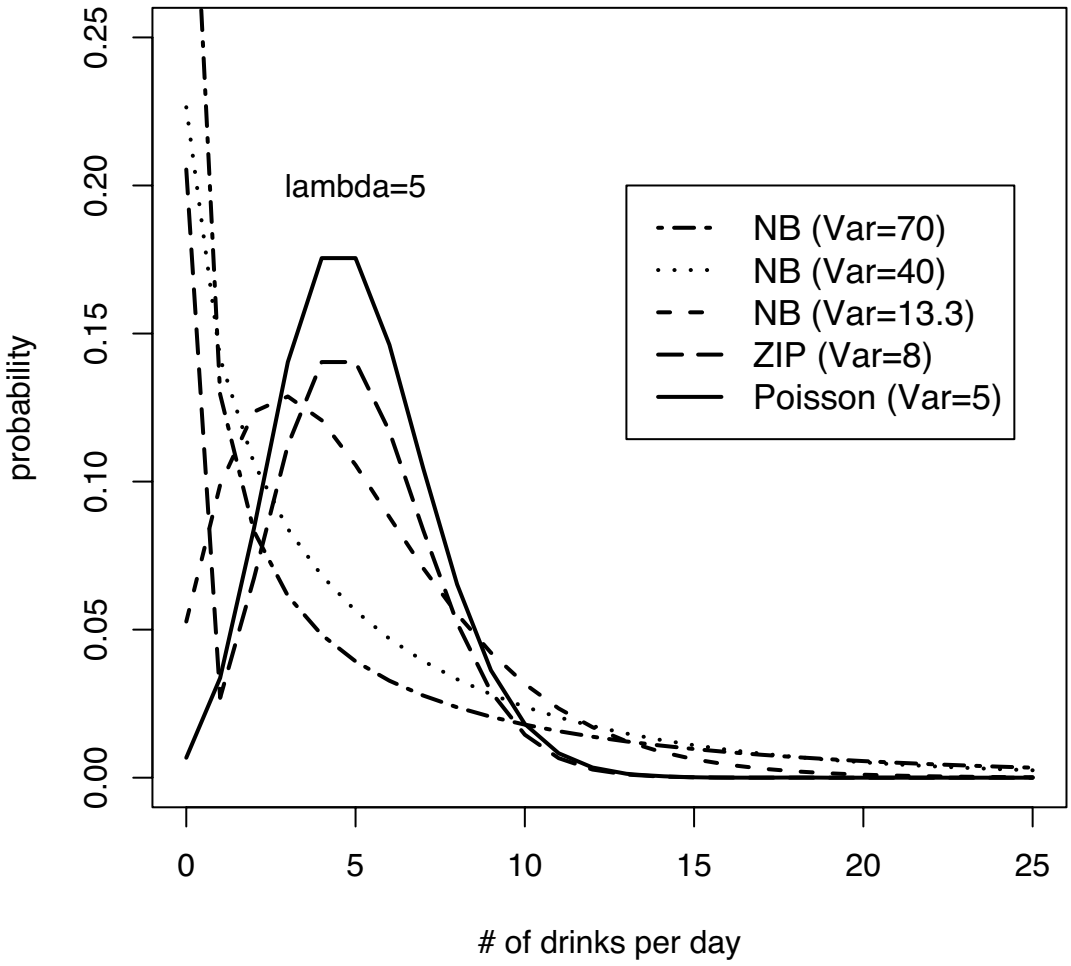


Figure 2
Graphical display of the five distributions, all with rate parameter 5, used in the simulations (Poisson [Var = 5], negative binomial [NB13, Var = 13], negative binomial [NB40, Var = 40], negative binomial [NB 70, Var = 70] and zero-inflated Poisson [ZIP, p = 0.2, Var = 8]).

Poisson standard Poisson model,

Over-dispersed Poisson Poisson model with empirical ("robust") variance estimator,

NB negative binomial,

ZIP zero-inflated Poisson, shared inflation parameter estimated for both randomized groups ($p_1 = p_2$),

ZINB zero-inflated negative binomial, shared inflation parameter estimated for both randomized groups ($p_1 = p_2$),

TTEST two-sample unequal variance t-test,

WILCOXON Wilcoxon-Mann-Whitney, a non-parametric two-sample comparison procedure suitable for ordinal data, and

PERMUTE two-sample permutation test.

Results

Simulation studies

In the simulation studies we assessed the behavior of models when the null hypothesis was true (there were no differences between alcohol consumption for groups 1 and 2). We note that the ZIP model failed to converge for more than a quarter of the simulations from the standard Poisson distribution. This is likely due to the fact that many datasets had no zeros whatsoever (for the Poisson distribution with $\lambda = 5$, the probability that a dataset has no zeros whatsoever is equal to $(1 - \exp(-5))^{100} = 0.51$).

Table 1: Estimated probability (and 99% CI) of rejecting the null hypothesis when there is no true difference between groups for a variety of statistical models and underlying distributions (results that do not include the alpha level of 0.05 are bolded)

	Analysis model fit			
True Distribution:	Poisson	ODP	NB	ZIP
Poisson (Var = 5)	.053 (.041,.064)	.054 (.042,.066)	.047 (.036,.058)	.055* (.043,.067)
NB (Var = 13)	**.225** (.204,.247)	.049 (.038,.060)	.049 (.038,.060)	.050 (.039,.061)
NB (Var = 40)	**.467** (.441,.493)	.047 (.036,.058)	.044 (.033,.055)	.046 (.036,.057)
NB (Var = 70)	**.584** (.558,.609)	.052 (.041,.063)	.048 (.037,.059)	.062 (.049,.074)
ZIP (Var = 8)	**.179** (.159,.199)	.058 (.046,.070)	**.031** (.022,.040)	.051 (.040,.063)

all distributions except ZIP have $E[Y_i] = \lambda = 5$, for ZIP $E[Y_i] = 0.8 * 5 = 4$.
ODP (over-dispersed Poisson); NB (negative binomial); ZIP (zero-inflated Poisson)
* For the true distribution under the Poisson, the ZIP model failed to converge for n = 672 of the simulations.

Table 1 displays the estimated Type I error rate (when there is no difference between the groups) when α was set to 0.05. The negative binomial model was conservative when the underlying data were zero-inflated. When the underlying distributions were not Poisson, the Poisson model did not maintain the appropriate Type I error rate. When the count models were over-dispersed by a factor of more than 2 (i.e. $Var(Y_i) > 2 * E[Y_i]$), the Poisson model rejected more than 22% of the time. When the over-dispersion was more extreme (factor of 8 and 14), the Type I error rate was 47% and 58%, respectively. The severe lack of robustness of the Poisson model in this setting is a serious concern.

ASAP study

Of 341 subjects enrolled in the clinical trial, 169 subjects were randomized to the control group and the other 172 into the intervention group. The mean age of the subjects was 44.3 (SD = 10.7). Twenty-nine percent were women, 45% were Black, 39% White, 9% Hispanic, and 7% Other. Sixty-three percent were unemployed during the past three months and 25% of the subjects were homeless at one point during the past three months. Four percent of the subjects met criteria for current (past year) alcohol abuse and 77% were alcohol dependent.

We analyze the 3-month follow-up data for which 271 subjects were observed (141 control, 130 treatment), for an overall response rate of 79%. Table 2 displays the distribution of drinks per day at baseline and 3-month follow-up separately for each group. As noted earlier, drinking outcomes are highly skewed to the right, with some extremely large values. These extreme values are plausible given the large number of dependent drinkers in the sample, many of whom have developed tolerance (the need to consume large amounts of alcohol to induce effects). We also note that reported drinking quantities decreased for both groups between baseline and 3-month outcome.

Table 3 displays the results from the ASAP study using a variety of count models. Use of the Poisson model yielded a statistically significant p-value, in contrast to the other methods (all other p-values > 0.45).

Figure 3 displays the observed and predicted counts for the Poisson, negative binomial, and ZIP models, while Figure 4 displays the plot of (observed minus expected) for the Poisson, negative binomial and ZIP models for the control group. The standard Poisson model provides an unsatisfactory fit, and is not appropriate for the analysis of this dataset. The fit of the zero-inflated Poisson is improved, particularly for modeling the probability of no drinking, but remains unsatisfactory over most of the remaining values. The negative binomial provides an excellent fit for these data, and that there is no indication

Table 2: Distribution of drinking outcome by timepoint and randomization group

	Base line		3 Months	
	C (n = 169)	T (n = 72)	C (n = 141)	T (n = 130)
MIN	0.17	0	0	0
25th percentile	1.14	1.32	0.17	0.13
MEDIAN	3.47	3.85	1.8	1.6
75th percentile	8.23	9.12	6.1	5.7
MAX	61.77	60	48.6	38.43
mean (SD)	6.95 (9.58)	6.68 (8.44)	4.98 (8.47)	4.36 (6.47)

Table 3: p-values for the ASAP randomization group effect at 3 months for a variety of count models

MODEL	p-value
Poisson	.018
over-dispersed Poisson	.489
Negative binomial	.458
zero-inflated Poisson	.542
zero-inflated negative binomial	.489
t-test	.495
Wilcoxon	.805
Permutation	.746

that any further zero-inflation is needed, since the model already overpredicts zeros (hence the predicted values for the NB and ZINB would be identical).

In this setting, there was little indication from the observed plots that there were significant group differences. As seen in the simulation studies, the Poisson may

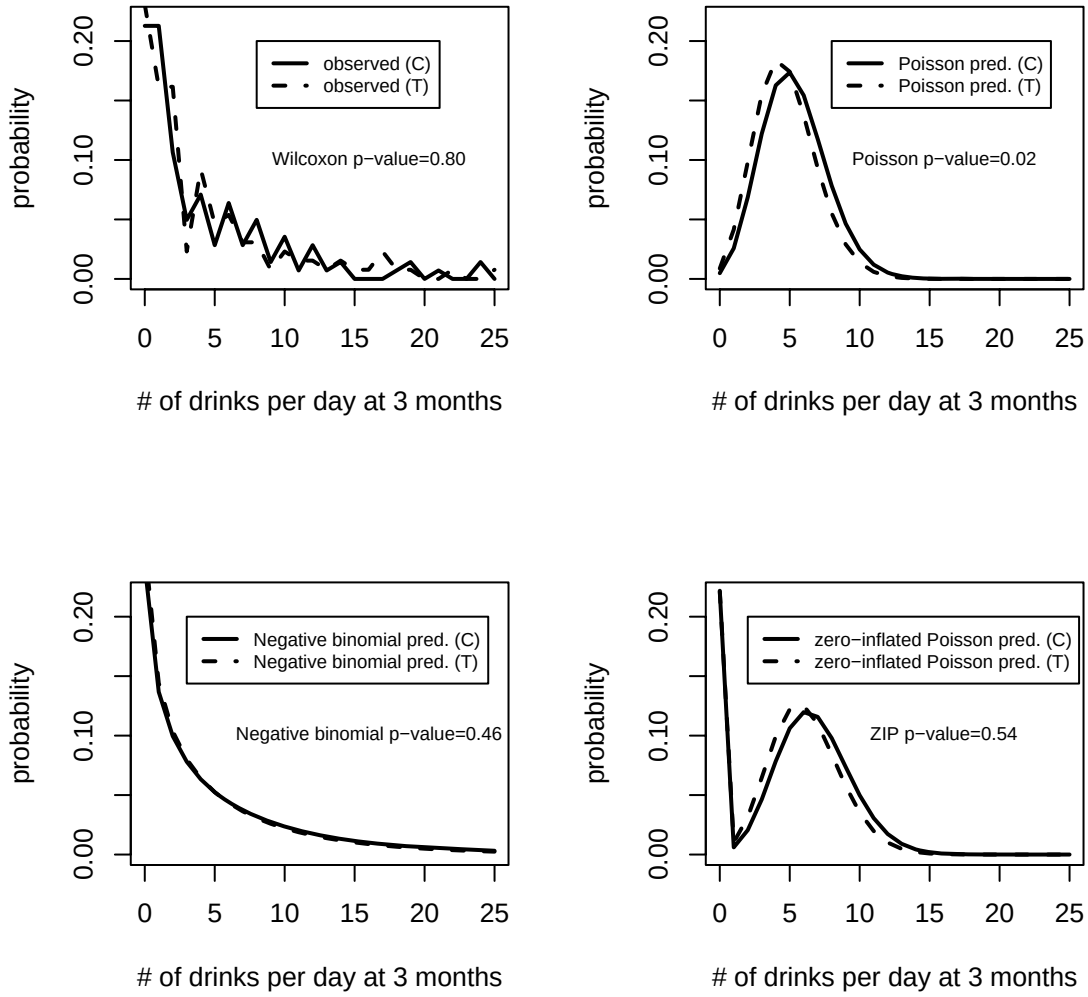


Figure 3
Observed and predicted values from the ASAP study at 3 months for control and treatment groups for each of four models: Wilcoxon, Poisson, negative binomial and zero-inflated Poisson.

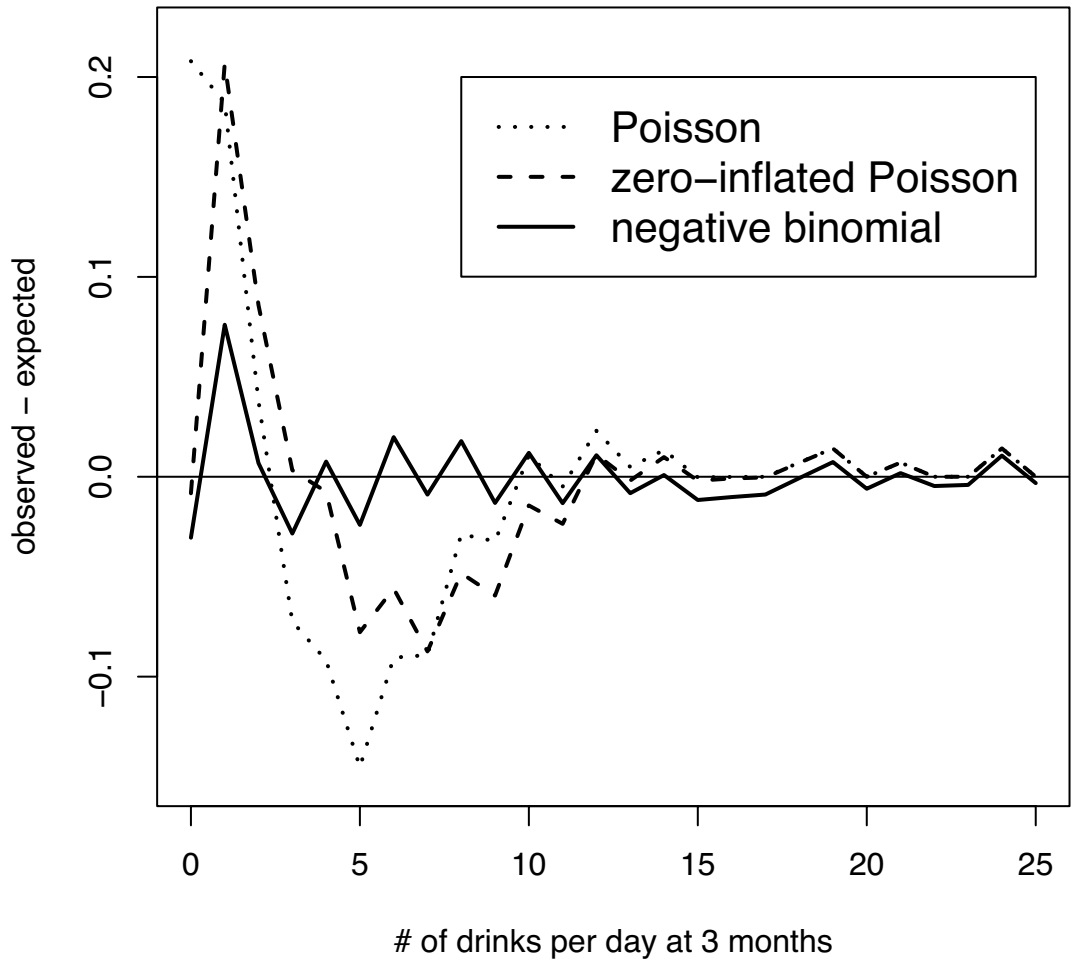


Figure 4
Observed minus expected values from the ASAP study at 3 months as a function of count for the Poisson, negative binomial and zero-inflated Poisson.

not have preserved the appropriate Type I error rate due to the extremely large values of drinking for some subjects. The Appendix includes the Stata commands to fit these models and the output, along with the code to generate observed and predicted plots using the prcounts routine.

Discussion and conclusion

A number of models have been proposed for the analysis of count data, and these models are now available in general purpose statistical packages. We have described these methods in the context of modeling reports of alcohol consumption, where a large proportion of respondents report no drinking, and a small number of respondents typically account for an extreme amount of drinking.

For the analysis of the ASAP study, we found that the standard Poisson had an extremely poor fit, and yielded a statistically significant p-value (in contrast to all of the other models, which had highly non-significant results). The unrealistic assumption that the expected rate of drinking is the same for all subjects may partially account for the poor fit of the Poisson distribution. We caution against use of the Poisson for this analysis. The negative binomial fit particularly well, and we saw no evidence for zero-inflation.

In settings where there are excess zeros, zero-inflation models are attractive. One advantage of these models is that they can estimate the probability of being a zero as a function of covariates, as well as allowing the rate parameter to be a function of covariates. In an alcohol study, the intervention may be hypothesized to affect the abstinence proportion as well as the rate parameter for drinkers. Ad-hoc methods in this setting might involve estimating the proportion of drinkers at follow-up, and in a separate model, estimating the amount of drinking amongst the

subset of subjects who reported any drinking. A more principled approach involves the simultaneous estimation of the zero-inflation factor (testing $p_1 = p_2$) and the rate parameter (testing $\lambda_1 = \lambda_2$). Slymen and colleagues [2] adopted this approach by simultaneously fitting separate models for what they describe as the "logistic" component and the "Poisson" component, and this approach is also detailed in books by Winkelmann [7] as well as Cameron and Trivedi [4].

The results of the simulation studies and the secondary analyses of the ASAP study demonstrated the importance of appropriately modeling count outcomes. We caution against the use of the standard Poisson model when the mean and variance are not equal. Extensions of the Poisson (incorporating an over-dispersion parameter or use of the negative binomial distribution and/or zero-inflated models) are now available in general purpose statistical software, and address many of the shortcomings of the overly simplistic Poisson model.

As always, analysts are obliged to look at their data and utilize models that provide an appropriate fit in their situation. In particular, for models of alcohol consumption, attention should be paid to the functional form of the outcome to ensure that underlying assumptions of the methods utilized are met.

Authors' contributions

NH conceived of the project and provided overall guidance, in addition to reviewing and interpreting analyses, and drafting the manuscript. EK participated in the drafting of the manuscript, and carried out analyses and simulations. RS led the ASAP study and participated in the drafting of the manuscript. All authors read and approved the final version of the manuscript.

Additional material

Additional File 1

Appendix. Stata code and results for count models.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2288-7-9-S1.pdf]

Acknowledgements

This research was supported in part by the National Institute on Alcohol Abuse and Alcoholism R01-AA12617, the Smith College Summer Research Program and the Howard Hughes Medical Institute. Thanks to Jessica Richardson for editorial assistance, Emily Shapiro and Min Zheng for assistance with simulations and Joseph Hilbe and Jeffrey Samet for helpful comments on an earlier draft.

References

1. 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 RZ, Allen JP. Totowa: Humana Pr, Inc.; 1992:41-69.
2. Slymen DJ, Ayala GX, Arredondo EM, Elder JP: **A demonstration of modeling count data with an application to physical activity.** *Epidemiologic Perspectives & Innovations* 2006, **3(3):**1-9.
3. Lambert D: **Zero-inflated Poisson regression, with an application to defects in manufacturing.** *Technometrics* 1992, **34:**1-14.
4. Cameron AC, Trivedi PK: *Regression analysis of count data* Cambridge, UK: Cambridge University Press; 1998.
5. Gardner W, Mulvey EP, Shaw EC: **Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models.** *Quantitative Methods in Psychology* 1995, **118(3):**392-404.
6. Hilbe JM: *Negative binomial regression: modeling overdispersed count data* Cambridge: Cambridge University Press in press.
7. Winkelmann R: *Econometric analysis of count data* fourth edition. Berlin: Springer-Verlag; 2003.
8. Long JS, Freese J: *Regression models for categorical dependent variables using Stata* Texas: Stata Press Publication; 2003.
9. Saitz R, Freedner N, Palfai TP, Horton NJ, Samet JH: **The severity of unhealthy alcohol use in hospitalized medical patients: the spectrum is narrow.** *Journal of General Internal Medicine* 2006, **21(4):**381-5.
10. McCullagh P, Nelder JA: *Generalized linear models* Chapman & Hall; 1989.
11. Zeger SL, Liang KY: **Longitudinal data analysis for discrete and continuous outcomes.** *Biometrics* 1986, **42:**121-130.
12. Liang KY, Zeger SL: **Longitudinal data analysis using generalized linear models.** *Biometrika* 1986, **73:**13-22.
13. Huber PJ: **The behavior of maximum likelihood estimates under non-standard conditions.** *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1967, **1:**221-233.
14. Miller WR, Rollnick S: *Motivational interviewing: Preparing people to change addictive behavior* second edition. New York: Guilford Press; 2002.
15. Saitz R, Palfai TP, Cheng DM, Horton NJ, Freedner N, Dukes K, Kraemer KL, Roberts MS, Guerriero RT, Samet JH: **Brief intervention for medical inpatients with unhealthy alcohol use: A randomized-controlled trial.** *Annals of Internal Medicine* 2007, **146(3):**167-176.

Pre-publication history

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

http://www.biomedcentral.com/1471-2288/7/9/prepub

Drug evaluation: DU-176b, an oral, direct Factor Xa antagonist

Elaine M Hylek

Address
Boston University Medical Center
Department of Medicine
91 East Concord Street
Suite 200
Boston
MA 02118
USA
Email: ehylek@bu.edu

Current Opinion in Investigational Drugs 2007 8(9):778-783
 ISSN 1472-4472

Daiichi Sankyo Inc (formerly Daiichi Seiyaku Co Ltd) is developing DU-176b, a direct, orally active Factor Xa inhibitor, as an anticoagulant for the potential treatment of cardiovascular indications, including venous and arterial thrombosis. By January 2005, phase II studies had begun in the US and Europe, and by March 2006, phase II studies had begun in Japan. Phase II trials for the prevention of thromboembolism were ongoing in mid 2007.

Originator Daiichi Sankyo Inc (formerly Daiichi Seiyaku Co Ltd)

Status Phase II Clinical

Indication Thromboembolism

Actions Coagulation inhibitor, Factor Xa antagonist

Synonym Factor Xa inhibitor (oral formulation)

Introduction

Vitamin K antagonist drugs (eg, warfarin) are highly efficacious for the prevention and treatment of venous and arterial thrombosis. However, despite their proven efficacy these drugs have significant drawbacks that limit their effectiveness, including a variable dose response and narrow therapeutic window. These features, coupled with common drug and food interactions, mandate frequent monitoring of patients which constitutes a significant barrier to their use in clinical care [814315].

Factor Xa (FXa) has a pivotal role in the coagulation cascade, common to both the extrinsic and intrinsic pathways, and is therefore a rational and potent target for new antithrombotic therapies [484147], [815221]. It is estimated that a single molecule of FXa is capable of generating 1000 molecules of thrombin, highlighting the desirability of suppressing the coagulation cascade proximal to thrombin [814319]. Direct FXa inhibitors are not dependent on antithrombin to exert their antithrombotic effect, unlike the indirect FXa inhibitor fondaparinux [323346]. Thus, direct activity combined with high selectivity enables the inhibition of FXa within the prothrombinase complex in addition to free FXa [814323], [816052].

DU-176b is an orally active, direct FXa inhibitor being developed by Daiichi Sankyo Inc [676138]. The first clinical candidate, DX-9065a, was developed by former subsidiary Daiichi Pharmaceuticals Co Ltd in Japan [158251]; however, clinical development of this compound was halted because of its poor oral bioavailability [638308]. DU-176b is currently in multinational phase IIb clinical trials for the potential prevention of deep vein thrombosis among individuals undergoing hip and knee replacement surgery [659999], [676138].

Synthesis and SAR

DX-9065a was shown to be an effective FXa inhibitor with low intestinal tract absorption; therefore, structural activity optimization was undertaken. X-ray crystallography revealed that 6-chloronaphthalene was a suitable replacement for the amidinonaphthyl group, the portion of the molecule which binds at the S1 site of FXa. Modification of the piperazine linker, and substitution of the pyrrolidin-1-yl-iminoimine with 4,5,6,7-tetrahydrothiazolo[5,4-*c*]pyridine, resulted in a compound that exhibited an activity of 4.4 μM in an assay of the concentration required to double prothrombin time (PT) in human plasma (the PTCT2 assay) [567056], [733062].

The low metabolic stability of this structure necessitated further SAR analysis. As a result, the 6-chloronaphthalene moiety was replaced with 6-chloro-benzo[*b*]thiophene, while the thiazolo[5,4-*c*]pyridine group at the S4 binding region of the molecule was substituted with thieno[3,2-*b*]pyridine 4-oxide [733058]. The resulting compound had an activity of 3.2 μM in a PTCT2 assay and a clearance of 6.97 ml/min/kg. In a subsequent study, the most active compound (PTCT2 = 0.54 μM) retained a pyridine-oxide group, while incorporating the structurally similar 6-chloro-benzo[*b*]pyrrole [733061].

Several potent oral FXa inhibitors were identified in these studies; however, DU-176b was not specifically identified. In addition, none of the lead compounds had properties sufficiently similar to those described in the limited published data on DU-176b; therefore, it is not possible to infer a structure or formula for DU-176b.

Preclinical development

DU-176b competitively inhibited human FXa *in vitro* (K_i = 0.56 nM) and displayed 10,000-fold selectivity for FXa relative to inhibition of thrombin [573107], [812208]. DU-176b did not affect the enzymatic activities of other serine proteases (ie, Factor VIIa, tissue plasminogen activator, plasmin, trypsin or chymotrypsin) [573107].

In *in vitro* studies with human plasma, DU-176b concentration-dependently increased PT and partial thromboplastin time (PTT), the standard measures of the extrinsic (PT) and intrinsic (PTT) coagulation pathways [573107], [812208]. Clotting times were doubled at concentrations of 0.256 at 0.508 μM, respectively.

Following oral administration of DU-176b to rats, anti-FXa activity was demonstrated in plasma over 4 h [573107]. In a

rat model of thrombogenesis induced by insertion of a wire into the vena cava, thrombosis was inhibited by DU-176b (0.5 to 12.5 mg/kg, ED_{50} = 0.076 mg/kg/h) [573101], [573107], [812208]. A 10-fold dissociation was observed between the antithrombotic and bleeding effects. This margin of safety for DU-176b was wider than for heparin, low-molecular weight heparin and warfarin, which inhibited thrombosis in this model (ED_{50} = 56 and 66 U/kg/h and 0.16 mg/kg/day, respectively) but prolonged bleeding time at doses slightly higher than the effective doses (1.3-, 2.0- and 1.3-fold, respectively) [573101], [812208].

DU-176b was also compared with the indirect FXa inhibitor fondaparinux in rat models of arterial thrombosis induced by ferric chloride and venous thrombosis caused by the insertion of a wire [573105]. DU-176b was administered at 0.05 to 1.25 mg/kg/h in both models, while fondaparinux was administered at 1 to 10 mg/kg/h in the arterial model and 0.03 to 1 mg/kg/h in the venous model. DU-176b was effective at preventing both arterial and venous thrombosis; however, fondaparinux required a 100-fold higher dose to inhibit arterial thrombosis compared with venous thrombosis. In a study of venous thrombosis induced in heterozygous antithrombin-deficient ($AT^{+/-}$) mice, the antithrombotic effect of DU-176b was more potent than fondaparinux and heparin [640490]. While DU-176b was equipotent in heterozygous and wild-type $AT^{+/+}$ mice, fondaparinux and heparin were less potent. Efficacy was evaluated in rat models under different blood sheer rates to simulate venous flow and flow through a narrowed artery using an *ex vivo* capillary perfusion chamber [812206]. In the same dose range (0.3 to 3.0 mg/kg/h), intravenous DU-176b significantly reduced thrombus formation under both venous and arterial blood sheer rates. Significantly higher doses of intravenous fondaparinux (values not stated) were necessary to inhibit thrombus formation at the higher arterial sheer rate (~ 66-fold higher dose than for the lower sheer rate). In addition, scanning electron microscopy revealed inhibition of fibrin net formation by DU-176b at both sheer rates and a limited effect on fibrin by fondaparinux at the high sheer rate [811466].

Because of its high selectivity and direct action, DU-176b has been shown to suppress FXa bound within the prothrombinase complex (K_i = 2.98 nM) [811474]. DU-176b inhibited the generation of thrombin by prothrombinase, a physiological complex comprising FXa, FVa, calcium and phospholipids. This was in contrast with indirect FXa inhibitors such as fondaparinux, where the effect was limited to free FXa [811474], [816052].

Additional animal studies compared DU-176b to melagatran, a direct thrombin inhibitor used as a research tool. There are hypothetical advantages of direct FXa inhibition compared with direct thrombin inhibition, including the potential for decreased rebound hypercoagulability upon anticoagulant withdrawal. In an *in vivo* model of hypercoagulability induced by a tissue factor (thromboplastin, a protein that initiates thrombin formation), rats were administered an intravenous bolus of melagatran (2 mg/kg) or DU-176b (0.3 mg/kg) [640496]. Both drugs inhibited platelet consumption and thrombin-antithrombin complex formation when tissue factor hypercoagulation was induced 5 min after drug administration. However, at 2 and 4 h after dosing, injection of tissue factor resulted in increased platelet consumption and formation of thrombin-antithrombin complexes with melagatran, indicating recurrent stimulation of coagulation.

Preliminary data in rat models of thrombosis (induced by arterial injection of ferric chloride, venous ligation or insertion of a wire into the vena cava) have also suggested an additive benefit of DU-176b when combined with other antithrombotic agents. A low dose of DU-176b (2 mg/kg po) combined with a low dose of the antiplatelet drug ticlopidine (30 mg/kg po) exhibited a more potent antithrombotic effect than either agent alone [812200]. Combination of the same DU-176b dose with the tissue plasminogen activator pamiteplase (1000 kU/mg iv) had a similarly additive thrombolytic effect.

Major hemorrhage, particularly intracranial hemorrhage, constitutes the most serious adverse effects of anticoagulants; therefore, reversibility of the anticoagulant effect is desirable particularly in an emergency situation. Recently presented data provide insight into the neutralizing effects on the activity of DU-176b of three approved reverse anticoagulants/factor replacements: prothrombin complex concentrate (Feiba; 0.03, 0.1 and 0.3 U/ml), recombinant Factor VIII (Kogenate-FS; 0.1, 1.0 and 6.7 U/ml) and Factor IX (Christmassin-M; 0.1, 1.0 and 6.7 U/ml) [811628]. In this *in vitro* study of human plasma and varying concentrations of DU-176b (150, 300 and 450 ng/ml), all three agents shortened PTT. The addition of a prothrombin complex concentrate reversed both PT and PTT in a concentration-dependent manner. Recombinant Factor VIIa also effectively corrected the prolongation of PT over a similar dose range [812203]. In a rat model of template bleeding, recombinant Factor VIIa (5, 50 and 500 ng/ml) given as a bolus injection corrected the bleeding time in a dose-dependent manner following infusion of DU-176b (1 mg/kg/h iv) over 2 h.

Animal data have also been presented on the effect of DU-176b infusion compared with melagatran infusion immediately following induction of intracerebral hemorrhage (ICH) by a collagenase solution [736678], [811467]. At a dose of 3 mg/kg/h, DU-176b had no effect on ICH volume. However, ICH volume increased by 180 and 230% at 6 and 18 mg/kg/h, and PT was prolonged by 2.8- and 6.5-fold, respectively. There were no deaths associated with these doses. Melagatran exerted no effect on ICH volume at the 0.3-mg/kg/h dose; however, at 1 and 3 mg/kg/h, ICH volume increased by 280 and 390%, and PT prolongation was 6.1- and > 30-fold, respectively. All rats in the 3-mg/kg/h melagatran group died. The safety margins for DU-176b and melagatran, the ratio of the dose required for ICH exacerbation to the ED_{50} value for thrombosis prevention, were 133 and 7, respectively.

Toxicity
No preclinical toxicology data were available at the time of publication.

Metabolism and pharmacokinetics
No data were available on the pharmacokinetics or metabolism of DU-176b at the time of publication. The oral bioavailability of DU-176b in monkeys was approximately 50% compared with 10% for DX-9065a; however, no further details were available [812208].

In a phase I trial in 12 healthy male volunteers administered a single dose of DU-176b (60 mg), drug levels at 1.5, 5.0 and 12.0 h post-dose were 240 (± 54), 127 (± 22) and 37 (± 10) ng/ml, respectively [733059]. However, no further details were provided.

Detailed pharmacokinetic and pharmacodynamic analyses are among the objectives of the ongoing phase II trials with DU-176b [815272], [815227].

Clinical development
Phase I
In a phase I trial, 12 healthy male volunteers (aged 28 ± 6 years) were administered single doses of DU-176b (60 mg) [733059]. Antithrombotic effects were assessed by serial measurement of thrombus size at 1.5, 5.0 and 12.0 h post-dose using a Badimon chamber under both venous and arterial flow conditions. At 1.5 and 5.0 h post-dose, venous thrombosis was reduced by 28 and 21%, respectively, and arterial thrombosis was reduced by 26 and 17%, respectively. At 12 h, the reduction in thrombosis was 3% under both flow conditions. Changes in clotting parameters, anti-FXa activity and thrombin generation paralleled the antithrombotic effects. Anti-FXa activity at 1.5, 5.0 and 12.0 h was 3.6, 1.6 and 0.3 IU/ml, respectively, and PTs were 20.3, 17.5 and 14.7 s, respectively, compared with 13.4 s at baseline.

Phase II
A phase IIa, open-label, dose-ranging trial assessed the efficacy, safety and tolerability of oral DU-176b for the prevention of venous thromboembolism (VTE) after total hip replacement [815225]. The trial enrolled 402 patients to receive once or twice daily oral doses of DU-176b. The trial was reportedly completed in July 2005; however, no data were available at the time of publication.

A phase IIb, double-blind trial of DU-176b for the prevention of VTE following total hip replacement commenced in the US, Europe, Canada, Russian Federation and Ukraine in May 2006 [815272]. This randomized, multiple dose trial was designed to compare the safety and efficacy of DU-176b with dalteparin for the prevention of VTE. The target enrollment was 950 patients, who were to receive 7 to 10 days of treatment. Secondary outcome measures were the incidence of VTE and major and clinically relevant non-major bleeds with DU-176b compared with dalteparin, and assessment of the pharmacokinetic and pharmacodynamic properties of DU-176b [815272].

Phase IIb trials were reportedly ongoing in Japan in 2006 [676138]; however, no details are available.

A phase II trial is also underway to compare DU-176b with warfarin in patients with non-valvular atrial fibrillation [815227]. Patients were to receive one of four fixed-dose regimens of DU-176b or warfarin (doses unstated) for 3 months followed by a 30-day follow-up. The primary endpoint was safety and secondary outcomes, including the assessment of major cardiac adverse events, pharmacokinetics and pharmacodynamics. The trial began recruiting patients in the US, Europe, South America, Mexico and the Russian Federation in June 2007.

No announcements have been made regarding phase III studies with DU-176b.

Side effects and contraindications
No data on clinical side effects were available at the time of publication.

Patent summary
Daiichi Sankyo has published several patents claiming FXa inhibitors. EP-00540051, published in May 1993, claims several series of thrombolytic compounds, one of which is DX-9065a. Claims were extended in WO-09854132 (December 1998), which exemplifies synthesis routes for 126 compounds, and WO-09933458 (July 1999) which discloses formulations for percutaneous administration. Subsequently, WO-03000657 (January 2003) claims diamine compounds and WO-2004058728 (July 2004) claims ethylenediamines. WO-2006106963, published in October 2006, claims diamino-substituted cyclic amine compounds.

DU-176b is not mentioned by name, and no structure or formula is apparently disclosed in the patent literature. It is not possible, therefore, to clarify which cases are important to DU-176b, or determine the relationship to DX-9065a. The only patent that mentions DU-176b (together with leading FXa inhibitors from other companies) is WO-2006106695, published by Kissei Pharmaceutical Co Ltd in October 2006, which discloses a coagulation system assay for quantifying the activity of an FXa inhibitor.

Current opinion
DU-176b is a novel, highly selective, oral direct FXa inhibitor. DU-176b has antithrombotic efficacy in preclinical models equivalent to or better than heparin, warfarin, melagatran or fondaparinux, with a significantly wider therapeutic window. Its antithrombotic effects have been demonstrated in both venous and arterial conditions in a preclinical model. Preclinical results suggest that the anticoagulant effect of DU-176b may be reversed (at least partially) by factor replacement. However, whether or not these agents will alter clinical outcomes in the presence of serious bleeding is unknown and requires further investigation. DU-176b is currently in phase II/IIb trials based on convincing preclinical data and positive results of phase I studies. However, until more clinical data become available, it is premature to extrapolate preclinical results to the clinical setting. The limited clinical data for DU-176b at the time of publication also prohibit comparisons to other direct FXa inhibitors.

The advantages of DU-176b over the current standards of care are true of other oral direct FXa inhibitors. This drug class shows great promise and has the capacity to revolutionize anticoagulant therapy [816290]. Other direct oral FXa inhibitors include rivaroxaban (Bayer AG/Ortho-McNeil Pharmaceutical Inc), apixaban (Bristol-Myers Squibb Co/Pfizer Inc), LY-517717 (Eli Lilly & Co), YM-150 (Astellas Pharma Inc), PRT-054021 (Portola Pharmaceuticals Inc) and 813893 (GlaxoSmithKline plc) [816291]. Of these, apixaban and rivaroxaban are more advanced in clinical development than DU-176b, as they are currently undergoing phase III trials. LY-517717, YM-150 and PRT-054021 are in phase II trials and 813893 is in phase I trials [816291].

The wide therapeutic window of the oral direct FXa inhibitors coupled with the limited potential for food and drug interactions will eliminate the need for monitoring in most cases, a situation that is currently mandated with vitamin K antagonists. The shorter half-life will also eliminate the need for bridging anticoagulants in invasive procedures. Despite these advances, several key issues remain unanswered until more robust clinical information is available for DU-176b, or for the other drugs. It is uncertain whether these agents will extend anticoagulant therapy to elderly individuals who are at the highest risk for major hemorrhage [816287]. Because renal function declines with age, safety in the elderly population will need to be rigorously demonstrated. Reversibility of the anticoagulant effect may be less of an issue with drugs characterized by shorter half-lives; however, reversibility in the setting of a life-threatening hemorrhage would be highly desirable, particularly if shown to improve clinical outcomes. The shorter half-life of newer agents will place greater emphasis on medication adherence. Large-scale phase III studies will answer these questions and, ultimately, the success of direct FXa inhibitors such as DU-176b will depend upon demonstrated efficacy and safety, particularly among older individuals for whom the need for antithrombotic therapy continues to expand.

Development status

Developer	Country	Status	Indication	Date	Reference
Daiichi Sankyo Inc	Canada	Phase II	Thromboembolism	01-MAY-06	815272
Daiichi Sankyo Inc	Japan	Phase II	Thromboembolism	31-MAR-06	676138
Daiichi Sankyo Inc	Mexico	Phase II	Thromboembolism	01-JUN-07	815227
Daiichi Sankyo Inc	Russian Federation	Phase II	Thromboembolism	01-MAY-06	815272
Daiichi Sankyo Inc	South America	Phase II	Thromboembolism	01-JUN-07	815227
Daiichi Sankyo Inc	Ukraine	Phase II	Thromboembolism	01-MAY-06	815272
Daiichi Sankyo Inc	US	Phase II	Thromboembolism	01-JAN-05	815225
Daiichi Sankyo Inc	Western Europe	Phase II	Thromboembolism	03-APR-06	659999

Literature classifications

Biology

Study type	Effect studied	Model used	Result	Reference
In vitro	Efficacy	Affinity of DU-176b for human FXa or other serine proteases.	DU-176b competitively inhibited human FXa (K_i = 0.56 nM) and had a 10,000-fold selectivity for FXa relative to inhibition of thrombin. DU-176b did not inhibit Factor VIIa, tissue plasminogen activator, plasmin, trypsin or chymotrypsin.	573107
In vitro	Efficacy	Human plasma treated with DU-176b (details unstated).	DU-176b concentration-dependently increased PT and PTT; clotting times were doubled at concentrations of 0.256 and 0.508 μM, respectively.	812208
In vivo	Efficacy	Rats with thrombogenesis (induced by insertion of a wire into the vena cava) were treated with a single oral dose of DU-176b (0.5 to 12.5 mg/kg).	DU-176b inhibited thrombogenesis (ED_{50} = 0.076 mg/kg/h). Heparin, low molecular weight heparin and warfarin also inhibited thrombosis (ED_{50} = 56 and 66 U/kg/h and 0.16 mg/kg/day, respectively). These drugs prolonged bleeding time at doses 1.3-, 2.0- and 1.3-fold above the effective dose, respectively, compared with a 10-fold safety margin for DU-176b.	573101
In vitro	Efficacy	Human plasma treated with DU-176b (150, 300 and 450 ng/ml) plus prothrombin complex concentrate (Feiba; 0.03, 0.1 and 0.3 U/ml), recombinant Factor VIII (Kogenate-FS; 0.1, 1.0 and 6.7 U/ml) or Factor IX (Christmassin-M; 0.1, 1.0 and 6.7 U/m).	All three reverse anticoagulants shortened the PTT. The addition of prothrombin complex concentrate reversed both the PT and PTT in a concentration-dependent manner.	811628

Biology (continued)

Study type	Effect studied	Model used	Result	Reference
In vivo	Efficacy	Rat models of arterial thrombosis induced by ferric chloride and venous thrombosis induced by wire insertion were treated with 0.05 to 1.25 mg/kg/h DU-176b or 1 to 10 mg/kg/h fondaparinux (1 to 10 mg/kg/h in the arterial model and 0.03 to 1 mg/kg/h in the venous model).	DU-176b inhibited arterial and venous thrombosis over a similar dose range (unstated), whereas fondaparinux required a 100-fold higher dose to inhibit arterial thrombosis compared with venous thrombosis.	573105

Metabolism

Study type	Effect studied	Model used	Result	Reference
In vivo	Bioavailability	Monkeys treated orally with DU-176b or DX-9065a.	The bioavailability of DU-176b was approximately 50% compared with 10% for DX-9065a.	812208
In vivo	Pharmacokinetics	A phase I trial of DU-176b (60 mg) administered to 12 healthy male volunteers.	Drug levels at 1.5, 5.0 and 12.0 h post-dose were 240 (± 54), 127 (± 22) and 37 (± 10) ng/ml, respectively.	733059

Clinical

Effect studied	Model used	Result	Reference
Efficacy	A phase I trial of DU-176b (60 mg) administered to 12 healthy male volunteers. Antithrombotic effects were assessed *ex vivo* using a Badimon chamber under both venous and arterial flow conditions.	At 1.5, 5.0 and 12.0 h post-dose, venous thrombosis was reduced by 28, 21 and 3%, respectively, and arterial thrombosis was reduced by 26, 17 and 3%, respectively. Changes in clotting parameters, anti-FXa activity and thrombin generation paralleled the antithrombotic effects.	733059

Associated patent

Title Aromatic amidine derivatives and salts thereof.

Assignee Daiichi Pharmaceutical Co Ltd

Publication EP-00540051 05-MAY-93

Priority JP-1991 286088 31-OCT-91

Inventors Nagahara T, Kanaya N, Inamura K, Yokoyama Y.

References

158251 **Dibasic (amidinoaryl) propanoic acid derivatives as novel blood coagulation Factor Xa inhibitors.** Nagahara T, Yokoyama Y, Inamura K, Katakura S, Komoriya S, Yamaguchi H, Hara T, Iwamoto M *J MED CHEM* 1994 **37** 8 1200-1207

323346 **Recent advances in inhibitors of Factor Xa in the prothrombinase complex.** Zhu BY, Scarborough RM *CURR OPIN CARDIOVASC PULM RENAL INVEST DRUGS* 1999 **1** 1 63-88

484147 **Factor Xa inhibitors: Today and beyond.** Walenga JM, Jeske WP, Hoppensteadt D, Fareed J *CURR OPIN INVEST DRUGS* 2003 **4** 3 272-281

567056 **Orally active Factor Xa inhibitors: 4,5,6,7-tetrahydrothiazolo[5,4-c] pyridine derivatives.** Haginoya N, Kobayashi S, Komoriya S, Hirokawa Y, Furugori T, Nagahara T *BIOORG MED CHEM LETT* 2004 **14** 11 2935-2939

573101 **Antithrombotic and hemorrhagic effects of DU-176b, a novel, potent and orally active direct Factor Xa inhibitor: A wider safety margin compared to heparins and warfarin.** Furugohri T, Honda Y, Matsumoto C, Isobe K, Sugiyama N, Morishima Y, Shibano T *BLOOD* 2004 **104** 11 Abs 1851

573105 **Antithrombotic properties of DU-176b, a novel, potent and orally active direct Factor Xa inhibitor in rat models of arterial and venous thrombosis: Comparison with fondaparinux, an antithrombin dependent Factor Xa inhibitor.** Fukuda T, Matsumoto C, Honda Y, Sugiyama N, Morishima Y, Shibano T *BLOOD* 2004 **104** 11 Abs 1852

573107 ***In vitro* characteristics, anticoagulant effects and *In vivo* antithrombotic efficacy of a novel, potent and orally active direct Factor Xa inhibitor, DU-176b.** Morishima Y, Furugohri T, Isobe K, Honda Y, Matsumoto C, Shiozaki Y, Muramatsu S, Sugiyama N, Shibano T *BLOOD* 2004 **104** 11 Abs 1862

638308 **Worldwide sales of olmesartan to top ¥ 100 billion in 2006: Daiichi Sankyo.** *PHARMA JPN* 2005 **1968** 1-3

640490 **Impact of antithrombin deficiency on efficacies of DU-176b, a novel orally active direct Factor Xa inhibitor, and antithrombin dependent anticoagulants, fondaparinux and heparin.** Fukuda T, Honda Y, Matsumoto C, Sugiyama N, Matsushita T, Yanada M, Morishima Y, Shibano T *BLOOD* 2005 **106** 11 Abs 1874

640496 **A thrombin inhibitor, but not a Factor Xa inhibitor, DU-176b, aggravates tissue factor-induced hypercoagulation in rats.** Morishima Y, Fukuda T, Tsuji N, Honda Y, Matsumoto C, Furugohri T, Shibano T *BLOOD* 2005 **106** 11 Abs 1878

659999 **Daiichi Sankyo Inc company website.** Daiichi Sankyo Inc *COMPANY WORLD WIDE WEB SITE* 2006 April 03

676138 **Dalichi Sankyo group research & development pipeline.** Daiichi Sankyo Co Ltd *COMPANY WORLD WIDE WEB SITE* 2006 March 31

733058 **Design, synthesis, and biological activity of novel Factor Xa inhibitors: Improving metabolic stability by S1 and S4 ligand modification.** Komoriya S, Kobayashi S, Osanai K, Yoshino T, Nagata T, Haginoya N, Nakamoto Y, Mochizuki A, Nagahara T, Suzuki M, Shimada T *et al BIOORG MED CHEM* 2006 **14** 5 1309-1330

733059 **A phase-I study to assess the antithrombotic properties of DU-176b: An orally active direct Factor-Xa inhibitor.** *J AM COLL CARDIOL* 2006 **47** 4 Suppl A Abs 908-161

733061 **Design, synthesis, and biological activity of non-basic compounds as Factor Xa inhibitors: SAR study of S1 and aryl binding sites.** Komoriya S, Haginoya N, Kobayashi S, Nagata T, Mochizuki A, Suzuki M, Yoshino T, Horino H, Nagahara T, Suzuki M, Isobe Y *et al BIOORG MED CHEM* 2005 **13** 12 3927-3954

733062 **Synthesis and conformational analysis of a non-amidine Factor Xa inhibitor that incorporates 5-methyl-4,5,6,7-tetrahydrothiazolo[5,4-c]pyridine as S4 binding element.** Haginoya N, Kobayashi S, Komoriya S, Yoshino T, Suzuki M, Shimada T, Watanabe K, Hirokawa Y, Furugori T, Nagahara T *J MED CHEM* 2004 **47** 21 5167-5182

736678 **A wide safety margin of a Factor Xa inhibitor, DU-176b, between antithrombotic effect and exacerbation of intracerebral hemorrhage in rats: Comparison with a thrombin inhibitor.** Morishima Y, Shirasaki Y, Kito F, Honda Y, Shibano T *CIRCULATION* 2006 **114** 18 Suppl Abs 632/B101

811466 **Comparison of antithrombotic properties of a direct FXa inhibitor, DU-176b, with an antithrombin dependent FXa inhibitor, fondaparinux, in a capillary perfusion chamber thrombosis model.** Shibano T, Tsuji N, Fukuda T, Morishima Y *CONGRESS INT SOC THROMB HAEMOSTASIS* 2007 **21** Abs P-T-640

811467 **Effect of a Factor Xa inhibitor DU-176b on intracerebral hemorrhage in rats: A wider safety margin compared with a thrombin inhibitor.** Morishima Y, Shirasaki Y, Kito F, Honda Y, Shibano T *CONGRESS INT SOC THROMB HAEMOSTASIS* 2007 **21** Abs P-T-641

811474 **Effects of DU-176b, a novel direct Factor Xa inhibitor, on prothrombinase activity and platelet aggregation *in vitro*.** Shibano T, Tsuji N, Kito F, Fukuda T, Furugohri T, Morishima Y *CONGRESS INT SOC THROMB HAEMOSTASIS* 2007 **21** Abs P-T-642

811628 **Activated prothrombin complex concentrate, recombinant Factor VIII and Factor IX reverse prolonged clotting time induced by DU-176b, a direct Factor Xa inhibitor, in human plasma.** Morishima Y, Fukuda T, Honda Y, Shibano Y *CONGRESS INT SOC THROMB HAEMOSTASIS* 2007 **21** Abs P-T-639

812200 **Antithrombotic properties of DU-176b, a novel orally active Factor Xa inhibitor: Inhibition of both arterial and venous thrombosis, and combination effects with other antithrombotic agents.** Morishima Y, Furugohri T, Honda Y, Matsumoto C, Isobe K, Fukuda T, Shiozaki Y, Muramatsu S, Sugiyama N, Shibano T *J THROMB HAEMOST* 2005 3 Suppl 1 Abs P0511

812203 **Recombinant Factor VIIa reverses Te prolonged bleeding time induced by high dose of DU-176b, a novel direct Factor Xa inhibitor, in rats.** Morishima Y, Honda Y, Matsumoto C, Fukuda T, Isobe K, Kumada T, Shibano T *J THROMB HAEMOST* 2005 **3** Suppl 1 Abs P0512

812206 **Antithrombotic effects of a direct Factor Xa inhibitor, DU-176b, and an antithrombin dependent Factor Xa inhibitor, fondaparinux, on an *ex vivo* capillary perfusion chamber thrombosis model in rats.** Fukuda T, Matsumoto C, Honda Y, Morishima Y, Shibano T *J THROMB HAEMOST* 2005 **3** Suppl 1 Abs P0612

812208 **Pharmacological characterization, antithrombotic and bleeding effects of DU-176b, a novel, potent and orally active direct inhibitor of Factor Xa: A wider safety margin of antithrombotic and bleeding effects compared to heparin, LMWH and warfarin.** Furugohri T, Isobe K, Honda Y, Matsumoto C, Sugiyama N, Morishima Y, Shibano T *J THROMB HAEMOST* 2005 **3** Suppl 1 Abs P1110

814315 **The pharmacology and management of the vitamin K antagonists: The seventh ACCP conference on antithrombotic and thrombolytic therapy.** Ansell J, Hirsh J, Poller L, Bussey H, Jacobson A, Hylek EM *CHEST* 2004 **126** 204S-233S

814319 **What is all that thrombin for?** Mann KG, Brummel K, Butenas S *J THROMB HAEMOST* 2003 **1** 7 1504-1514

814323 **Factor Xa is highly protected from antithrombin-fondaparinux and antithrombin-enoxaparin when incorporated into the prothrombinase complex.** Brufatto N, Ward A, Nesheim ME *J THROMB HAEMOST* 2003 **1** 6 1258-1263

815221 **Inhibition of Factor Xa: A potential target for the development of new anticoagulants.** Alexander JH, Singh KP *AM J CARDIOVASC DRUGS* 2005 **5** 5 279-290

815225 **NCT00107900: Study of the efficacy and safety of DU-176b in preventing blood clots in patients undergoing total hip replacement.** Daiichi Sankyo Inc *WWW.CLINICALTRIALS.GOV* 2006 April 04

815227 **NCT00504556: Study to assess the safety of a potential new drug in comparison to the standard practice of dosing with warfarin for a diagnosis of non-valvular atrial fibrillation.** Daiichi Sankyo Inc *WWW.CLINICALTRIALS.GOV* 2007 July 18

815272 **NCT00398216: A study of DU-176b in preventing blood clots after hip replacement surgery.** Daiichi Sankyo Inc *WWW.CLINICALTRIALS.GOV* 2007 March 27

816052 **Clot-bound thrombin is protected from inhibition by heparin-antithrombin III but is susceptible to inactivation by antithrombin III-independent inhibitors.** Weitz JI, Hudoba M, Massel D, Maraganore J, Hirsh J *J CLIN INVEST* 1990 **86** 2 385-391

816287 **Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation.** Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S *CIRCULATION* 2007 **115** 21 2689-2696

816290 **New anticoagulants.** Weitz JI, Bates SM *J THROMB HAEMOST* 2005 **3** 8 1843-1853

816291 **Oral, direct Factor Xa inhibitors in development for the prevention and treatment of thromboembolic diseases.** Turpie AG *ARTERIOSCLER THROMB VASC BIOL* 2007 **27** 6 1238-1247

nature CLINICAL PRACTICE
CARDIOVASCULAR
MEDICINE

PRACTICE POINT

Nature Clinical Practice Cardiovascular Medicine (2007) 4, 186-187
doi:10.1038/ncpcardio0800
Received 18 October 2006 | Accepted 29 November 2006 | Published online: 6 February 2007

Does preadmission anticoagulation therapy reduce stroke fibrillation?

Elaine M Hylek

Correspondence Department of Medicine, Boston University Medical Center, Suite 200, 91 East Concorc

Email elaine.hylek@bmc.org

This article has no abstract so we have provided the first paragraph of the full text.

Stroke is the leading cause of disability and the third most-common cause of death. mortality and more severe disability than other ischemic stroke subtypes. AF is a pc that emboli in AF originate in the left atrial appendage, where the confluence of sta predispose to thrombus formation. Factors that promote embolization of *in situ* thr

COMMENTARY

Stroke is the leading cause of disability and the third most common cause of death. AF is a powerful risk factor for stroke. Cardioembolic strokes are associated with higher mortality and more severe disability than other ischemic stroke subtypes. It is widely accepted that emboli in AF originate in the left atrial appendage where the confluence of stasis, hypercoagulability, and endothelial dysfunction predispose to thrombus formation. Factors that promote embolization of in-situ thrombus are poorly understood.

Vitamin K antagonists (VKA) have been shown to greatly reduce the risk of stroke in AF. The unanticipated effect of INR intensity on stroke severity and mortality has now been validated in three studies most recently by O'Donnell et al. These independent studies conducted within distinct AF populations in the United States[1], Norway[2], and now Canada have all shown that an INR ≥ 2 at the time of stroke is associated with decreased severity and reduced stroke-related mortality compared to patients presenting with INR <2. Although the effect of pre-stroke aspirin was also assessed, the pivotal comparison is within the warfarin group as comparison across treatment groups is subject to selection bias and possible confounding related to initial non-random assignment of antithrombotic drug.

These studies collectively provide powerful evidence for the optimal intensity of anticoagulation for preventing ischemic stroke and stroke-related disability and mortality in AF. Recently revised guidelines recommend VKAs (e.g., warfarin) with an INR target of 2.0-3.0 for patients with a history of stroke or TIA and for those with ≥ 2 moderate risk factors, i.e., hypertension, age ≥ 75, diabetes mellitus, or heart failure. Among patients with only 1 of these 4 risk factors, aspirin or VKAs may be considered based on patient preference and hemorrhagic risk given the expected lower rate of stroke.[3]

The clinical challenge arises when patients at high risk of stroke also have a propensity for hemorrhage, a scenario encountered more frequently as we care for older and more medically complicated patients. Although major extracranial hemorrhage causes significant morbidity and often precipitates cessation of warfarin, its long-term sequelae are not comparable in severity to cardioembolic stroke. From the patient's perspective, a disabling stroke is often viewed as equivalent as or worse than death.[4]

Interventions to reduce hemorrhage should be aggressively implemented to optimize the effectiveness of anticoagulant therapy. Revised guidelines newly advise against use of aspirin for stable coronary artery disease among patients taking warfarin.[3] Blood pressure control has been shown to decrease the risk of both ischemic and hemorrhagic stroke. Because the first 90 days of warfarin convey the highest risk, vigilant INR monitoring is essential. Elderly patients require lower doses of warfarin and are slower to normalize an elevated INR. Physicians and patients need to be cognizant of the most powerful precipitants of erratic control, e.g., decompensated heart failure, amiodarone, chemotherapy, to better anticipate changes in bleeding risk. Lower target intensity, INR 1.5-2.5, has not been shown to decrease major bleeding and therefore likely only exposes patients to unacceptable stroke risk.[5]

Finally, it is hoped that newer anticoagulant drugs with a wider therapeutic index and shorter half-life will translate into safer, yet equally efficacious, alternatives to warfarin.

1. Hylek EM, et al. (2003) Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. N Engl J Med 349:1019-1026.
2. Indredavik B, et al (2005) Frequency and effect of optimal anticoagulation before onset of ischaemic stroke in patients with known atrial fibrillation. J Intern Med 258:133-144.
3. Fuster V, et al. (2006) ACC/AHA/ESC Guidelines for the Management of Patients with Atrial Fibrillation: a report of the American College of Cardiology/American Heart Association Task Force and the European Society of Cardiology Committee on Practice Guidelines. Circulation 114:700-752.
4. Devereaux PJ, et al. (2001) Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. BMJ 323:1-7.
5. Fang MC, et al (2004) Advanced age, anticoagulation intensity, and risk for intracranial hemorrhage among patients taking warfarin for atrial fibrillation. Ann Intern Med 141:745-752.

Major Hemorrhage and Tolerability of Warfarin in the First Year of Therapy Among Elderly Patients With Atrial Fibrillation

Elaine M. Hylek, MD, MPH; Carmella Evans-Molina, MD; Carol Shea, RN; Lori E. Henault, MPH; Susan Regan, PhD

Background—Warfarin is effective in the prevention of stroke in atrial fibrillation but is under used in clinical care. Concerns exist that published rates of hemorrhage may not reflect real-world practice. Few patients ≥80 years of age were enrolled in trials, and studies of prevalent use largely reflect a warfarin-tolerant subset. We sought to define the tolerability of warfarin among an elderly inception cohort with atrial fibrillation.

Methods and Results—Consecutive patients who started warfarin were identified from January 2001 to June 2003 and followed for 1 year. Patients had to be ≥65 years of age, have established care at the study institution, and have their warfarin managed on-site. Outcomes included major hemorrhage, time to termination of warfarin, and reason for discontinuation. Of 472 patients, 32% were ≥80 years of age, and 91% had ≥1 stroke risk factor. The cumulative incidence of major hemorrhage for patients ≥80 years of age was 13.1 per 100 person-years and 4.7 for those <80 years of age ($P=0.009$). The first 90 days of warfarin, age ≥80 years, and international normalized ratio (INR) ≥4.0 were associated with increased risk despite trial-level anticoagulation control. Within the first year, 26% of patients ≥80 years of age stopped taking warfarin. Perceived safety issues accounted for 81% of them. Rates of major hemorrhage and warfarin termination were highest among patients with $CHADS_2$ scores (an acronym for congestive heart failure, hypertension, age ≥75, diabetes mellitus, and prior stroke or transient ischemic attack) of ≥3.

Conclusions—Rates of hemorrhage derived from younger noninception cohorts underestimate the bleeding that occurs in practice. This finding coupled with the short-term tolerability of warfarin likely contributes to its underutilization. Stroke prevention among elderly patients with atrial fibrillation remains a challenging and pressing health concern. **(*Circulation*. 2007;115:2689-2696.)**

Key Words: anticoagulants ■ atrial fibrillation ■ hemorrhage ■ stroke

Warfarin is highly effective in the prevention of stroke in atrial fibrillation (AF).[1,2] Despite its proven benefit, studies attest to its underutilization particularly among elderly individuals who face the highest risk.[3–6] In a recent study of 21 teaching, 13 community, and 4 Veterans Administration hospitals in 28 states, the use of warfarin at discharge was 54% even among patients considered to be at highest risk.[7] Age >80 years and perceived bleeding risk were negative predictors of warfarin use. In a recent prospective study, 51% of patients were discharged on warfarin. Of those not on warfarin, 83% had ≥2 risk factors for stroke and 23% had taken warfarin in the past but were unable to tolerate it in the long term.[8] Stroke is the leading cause of disability and the third leading cause of death in the United States with an estimated annual total cost of 57.9 billion dollars.[9] Given the aging of our population, it is projected that 7.5 million individuals will have AF in the United States by the year 2020 on the basis of an expected prevalence of 13.5% for individuals ≥75 years of age, and 18.2% for those ≥85 years.[10]

Editorial p 2684
Clinical Perspective p 2696

Rates of major hemorrhage from randomized trials and observational cohorts have been reassuringly low.[11–15] However, published rates may be underestimates, as few patients >80 years of age were enrolled and few cohort studies include the initial phase of therapy, which is reported to convey the highest risk.[13,16–19] Observational noninception cohort studies focus on prevalent warfarin use among individuals who have proven they can tolerate anticoagulant therapy. Moreover, recent trials have also largely enrolled patients on vitamin K antagonists at study entry.[20–22]

Continuing medical education (CME) credit is available for this article. Go to http://cme.ahajournals.org to take the quiz.
Received July 21, 2006; accepted March 13, 2007.

From the Department of Medicine (E.M.H., L.E.H.), Section of General Internal Medicine-Research Unit, Boston University School of Medicine, Boston Medical Center, and Department of Medicine (C.S., S.R.), General Medicine Division, Massachusetts General Hospital, Boston, Mass; and Department of Medicine (C.E.-M.), University of Virginia, Charlottesville, Va.

Guest Editor for this article was Douglas P. Zipes, MD.

Correspondence to Elaine M. Hylek, MD, MPH, Research Unit, Section of General Internal Medicine, Boston University School of Medicine, Boston Medical Center, 91 East Concord St, Suite 200, Boston, MA 02118. E-mail ehylek@bu.edu

Circulation **is available at http://www.circulationaha.org** **DOI: 10.1161/CIRCULATIONAHA.106.653048**

Higher rates of hemorrhage with incident warfarin use may in part underlie the reported underutilization of warfarin from cross-sectional studies and noninception cohorts. Long-term tolerability of anticoagulant therapy is largely unknown and may also be a contributing factor. We conducted an inception cohort study to define the rate of major hemorrhage in patients at warfarin initiation and to define the risk of bleeding in the early phase of therapy. We also focused on the tolerability of warfarin during the first year and determined the physician reason for discontinuation with attention to perceived safety issues.

Methods

Design and Patients

To be eligible, patients had to be ≥65 years of age, have AF verified by ECG, be new to warfarin (or had taken none within the previous 12 months), have their care established at the study institution, and have their warfarin managed by the on-site anticoagulation clinic. Potentially eligible patients were prospectively identified through daily searches of electronic admission notes of patients admitted to the medical service of Massachusetts General Hospital from January 2001 to June 2003. We intentionally did not screen patients admitted to the stroke or surgical service given their higher baseline risk of bleeding. Outpatients who started warfarin were identified at the time of the faxed physician referral to the anticoagulation clinic and were tracked to the first appointment. Referred ambulatory patients with AF-related stroke were eligible only if the event had occurred at least 4 weeks before the first appointment.

Outcomes

Patients were enrolled on the first day of warfarin and followed through the first year. Outcomes included major hemorrhage, time to termination of warfarin, and physician reason for discontinuation. Major hemorrhage was defined as fatal, hospitalization with transfusion of ≥2 units of packed red blood cells, or involvement of a critical site (ie, intracranial, retroperitoneal, intraspinal, intraocular, pericardial, or atraumatic intra-articular hemorrhage). The international normalized ratio (INR) at the time of the event was recorded in addition to other potentially contributing factors such as a fall. The reason for warfarin termination was recorded from the medical record. If the reason was not explicitly stated, the physician was queried directly. Reasons for termination included major hemorrhage, other hemorrhage, falls, patient nonadherence with medication or monitoring, coagulopathy, or sustained sinus rhythm.

TABLE 1. Clinical Characteristics of Patients at the Start of Warfarin Therapy

Variable	All (n=472)	≥80 Years of Age (n=153)	<80 Years of Age (n=319)	*P*
Age				
Mean, y (range)	77 (65 to 97)	84 (80 to 97)	73 (65 to 79)	—
Female, % (n)	47 (222)	55 (84)	43 (138)	0.21
Atrial fibrillation, % (n)				
First episode	59 (279)	59 (90)	59 (189)	
Recurrent	35 (165)	36 (55)	34 (110)	
Sustained	6 (28)	5 (8)	6 (20)	
Risk factors for stroke, % (n)				
Coronary artery disease	35 (166)	37 (57)	34 (109)	0.47
Diabetes mellitus	22 (106)	19 (29)	24 (77)	0.23
Heart failure	28 (130)	37 (57)	23 (73)	0.001
Hypertension	75 (354)	83 (127)	71 (227)	0.004
Prior stroke	5 (25)	9 (14)	3 (11)	0.010
$CHADS_2$ score (estimated stroke rate), % (n)[23]				<0.0001
0 (1.9%)	9 (42)	0 (0)	13 (42)	
1 (2.8%)	26 (121)	10 (15)	33 (106)	
2 (4%)	38 (181)	46 (70)	35 (111)	
3 (5.9%)	20 (94)	28 (43)	16 (51)	
≥4 (≥8.5%)	7 (34)	16 (25)	3 (9)	
Potential risk factors for hemorrhage, % (n)				
Active malignancy	5 (25)	5 (7)	6 (18)	0.63
Prior gastrointestinal bleed	7 (32)	10 (15)	5 (17)	0.07
Prior other bleed	3 (14)	3 (5)	3 (9)	0.79
Dementia	3 (15)	6 (9)	2 (6)	0.52
Liver disease	1 (3)	(0)	1 (3)	0.23
History of falling	4 (21)	8 (12)	3 (9)	0.01
Renal dysfunction*	11 (50)	15 (23)	8 (27)	0.03
Excessive alcohol use	1 (7)	(0)	2 (7)	0.07
Antiplatelet therapy	40 (190)	39 (59)	41 (131)	0.60

*Creatinine >1.5 mg/dL; creatinine value was unobtainable for 5 patients.

Patient Characteristics

Demographic and clinical characteristics extracted from the medical record included known risk factors for stroke (hypertension, prior stroke, heart failure, diabetes mellitus) and potential risk factors for hemorrhage (prior hemorrhage, liver disease, falls, active alcohol abuse, active malignancy, renal impairment, dementia). Stroke risk scores were calculated according to the $CHADS_2$ scheme (congestive heart failure=1 point, hypertension=1, age ≥75=1, diabetes mellitus=1, and prior stroke or transient ischemic attack=2).[23] Concurrent medications were obtained from the discharge summary, electronic medication list, and physician office record and were verified by the patient or family member at the first anticoagulation appointment. Use of aspirin or nonsteroidal antiinflammatory medications was also recorded.

Anticoagulation Intensity and Warfarin Management

Per clinic routine, all patients attended a mandatory 60-minute educational session on warfarin. Time spent in the INR ranges <2.0, 2.0 to 3.0, 3.1 to <4.0, and ≥4.0 was calculated with linear interpolation following the method described by Rosendaal et al.[24] This method does not interpolate an INR if the interval between 2 consecutive measurements exceeds 8 weeks.

Statistical Analysis

Baseline differences between age groups were evaluated with χ^2 tests for categorical variables and *t* tests for continuous variables. Crude event rates were calculated for different age groups, INR categories, time since warfarin initiation, and $CHADS_2$ categories by division of the number of first events by the total person-years of follow-up in the specified category. The INR at the time of event was unobtainable for 1 patient and this patient was excluded from analyses that involved this variable. Poisson regression models with generalized estimating equations were used to calculate incidence rate ratios (IRR) and 95% confidence intervals (CIs). The number of events precluded use of a multivariable regression model.

Cumulative incidence of major bleeding and intolerability of therapy, defined as stopping warfarin therapy for safety reasons, were both estimated with the Kaplan-Meier method. Time to major bleeding was censored at the time of cessation of warfarin therapy, death, or transfer of medical care, whereas time to cessation of warfarin was censored for reasons other than safety, death, or transfer of medical care. Log-rank tests were used to assess differences between the 2 age groups. For all analyses, a 2-sided probability value <0.05 was considered statistically significant. Analyses were performed with Stata statistical software, release 8.0 (Stata Corporation, College Station, Tex.).

Role of the Funding Source

The sponsor had no role in the design and conduct of the study, the collection, analysis, and interpretation of data, or in the preparation, review, and approval of the manuscript. The study was approved by the institutional review board at Massachusetts General Hospital. The observational nature of the study did not require written informed consent.

The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agree to the manuscript as written.

Results

During the study period, 533 patients were identified and 472 were enrolled; 61% of patients had their warfarin managed elsewhere (eg, at a skilled nursing facility) and were not included. A total of 47% of patients were female and 54% were >75 years of age (32% were ≥80 years of age) (Table 1). A total of 59% (*n*=279) of patients presented with the first documented episode of AF and 35% (*n*=165) with a recurrent episode. Thirty-five percent of patients had documented coronary artery disease, and 40% overall were on aspirin for primary or secondary prevention.[25] Compared with risk factors for stroke, potential risk factors for hemorrhage were less common except for older age. Of the 472 patients, 33% were enrolled at hospital discharge; 42% of these were ≥80 years of age. Symptoms related to an uncontrolled ventricular rate prompted admission for the majority of patients. A total of 90% of patients ≥80 years had a $CHADS_2$ score ≥2. With the use of the Outpatient Bleeding Risk Index, 95.3% of patients would have been classified as intermediate risk and 4.7% as high risk for major hemorrhage at the start of warfarin.[26]

Follow-up was complete for 100% of the cohort. During the first year of therapy, 134 patients were taken off warfarin, 16 died of unrelated causes, 16 transferred their warfarin management, and 306 patients remained on warfarin at the end of the 1-year observation period. The total observation time was 360 person-years.

Anticoagulation Control

The total number of INR measurements was 10 031. Person-time by INR category could not be determined for 21 person-years (5.7%) of the total time on warfarin. During the study, 58% of person-time was spent within the INR range 2.0 to 3.0, 29% below 2.0, 11% within 3.1 to <4.0, and 2% ≥4.0.

Major Hemorrhage

During the first year, 26 patients sustained a major hemorrhage (9 intracranial, 11 gastrointestinal, 1 retroperitoneal, 1 hemothorax after a fall, 1 ocular, 1 hemarthrosis, and 2 epistaxis that required transfusions). The rate of major hemorrhage was 7.2 per 100 person-years (95% CI 4.9 to 10.6), and the rate of intracranial hemorrhage was 2.5% (95% CI 1.1 to 4.7). Patients ≥80 years of age experienced higher rates of major bleeding compared with younger patients (13.08 per 100 person-years versus 4.75 per 100 person-years,

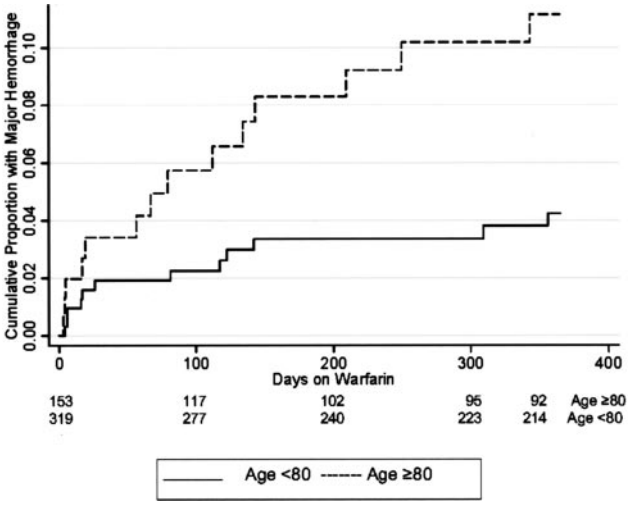


Figure 1. Cumulative incidence of major bleeding among patients aged ≥80 years and <80 years (n=472). Numbers below graph are the number of patients without bleeding who continued on warfarin at that time point (*P*=0.009, log-rank test).

TABLE 2. Incidence Rates and Incidence Rate Ratios for Major Hemorrhage Among Patients Newly Starting Warfarin

Risk Factor	Patient-Years of Follow-Up	Major Bleed (n)	IR per 100 Person-Years	IRR (95% CI) Unadjusted
Age, y				
≥80	107	14	13.08	2.75 (1.27 to 5.95)
<80	253	12	4.75	
Time spent in INR range*				
<2	97	4	4.11	19.34 (8.26 to 45.34)†
2 to 3	195	7	3.78	
3.1 to <4.0	38	6	15.78	
≥4	8	8	99.26	
Timing of events, d				
≤90	105	15	14.23	3.31 (1.51 to 7.25)
>90	254	11	4.13	
Stratification by INR*				
INR <4.0				
Age, years				
≥80	99	9	9.11	2.65 (1.01 to 6.95)
<80	233	8	3.44	
Timing of events, d				
≤90	97	10	10.31	3.47 (1.32 to 9.16)
>90	234	7	2.99	
INR ≥4.0				
Age, years				
≥80	2	5	230.39	3.80 (0.89 to 16.16)
<80	6	3	53.55	
Timing of events, d				
≤90	4	5	134.79	1.95 (0.46 to 8.24)
>90	4	3	68.94	

IR indicates incidence rates.
*The INR value of 1 patient (age 79 years) who sustained an intracerebral hemorrhage on day 117 after warfarin initiation was unobtainable.
†INR≥4 versus <4.

$P=0.010$). Figure 1 shows the cumulative major hemorrhage rate over time between the 2 age groups ($P=0.009$). Table 2 presents rates of hemorrhage by 3 risk factors, INR range, age group, and time since warfarin initiation. Risk of hemorrhage was increased by INR ≥4.0 (IRR 19.34, 95% CI 8.26 to 45.34), older age (IRR 2.75, 95% CI 1.27 to 5.95), and the first 90 days of therapy (IRR 3.31, 95% CI 1.51 to 7.25). INR ≥4.0 increased the risk, although only 2% of person-time was spent in this range. Fifteen of the 26 major hemorrhages (58%) occurred within 90 days of warfarin initiation, 11 within 30 days, and 7 within the first 2 weeks. Six of these 7 patients were enrolled at the time of hospital discharge; 2 were bridged with heparin. INR elevation was correlated with the other 2 risk factors. Of the 8 patients who experienced a major hemorrhage with an INR ≥4.0, 5 patients sustained the hemorrhage within 90 days, 5 patients were ≥80 years of age, and 3 patients had all 3 risk factors. Of the 15 hemorrhagic events that occurred within the first 90 days, the INR was ≥4.0 in 5 of them. However, age and time since warfarin initiation remained significant predictors when analysis was restricted to INR <4 (Table 2).

In the overall cohort, 40% of patients were on aspirin. Twelve of the 26 events (46%) occurred on aspirin and included 5 gastrointestinal and 4 intracranial bleeds. The dose was 81 mg for 8 of these patients. Nine of the 12 events occurred with an INR <4 and 6 patients were <80 years of age. Of the total 9 intracranial bleeds, 3 were fatal, 6 were intracerebral, and 2 were associated with documented falls. INR values were available for 8 patients and exceeded 4.0 in 3 of them. One intracerebral bleed occurred within the first 30 days, and 8 of 9 bleeds occurred in patients who were ≥75 years of age (3 were ≥80 years of age). Overall, the rate of major hemorrhage was higher in those with $CHADS_2$ scores of ≥3 (IRR 8.19, 95% CI 3.37 to 19.88) (Table 3).

Tolerability of Warfarin Among Patients ≥80 Years of Age

By the end of the first year, 134 patients had been taken off warfarin. Sustained sinus rhythm was the predominant reason for those <80 years of age, 63% (57 of 91 patients), compared with those ≥80 years of age, 19% (8 of 43 patients). Concerns related to safety accounted for 81% of those patients ≥80 years of age who stopped therapy during the first year (17 bleeding complications, 9 falls, 5 nonadherence with drug or monitoring, 3 coagulopathy, and 1 dermatologic reaction) compared with 37% of patients <80 years of

TABLE 3. Distribution of Major Hemorrhagic Events and Warfarin Terminations Due to Perceived Safety Concerns by $CHADS_2$ Score

	Overall		Major Bleed			Taken Off Therapy		
$CHADS_2$ Score	N	Person-Years	N	Rate (per 100 Person-Years)	95% CI	N	Rate (per 100 Person-Years)	95% CI
0	42	32	1	3.12	0.08 to 17.38	5	15.59	5.06 to 36.39
1	121	93	4	4.28	1.17 to 10.96	16	17.12	9.79 to 27.81
2	181	147	3	2.04	0.42 to 5.96	19	12.92	7.78 to 20.18
3	94	61	12	19.54	10.10 to 34.13	20	32.56	19.89 to 50.29
≥4	34	26	6	23.42	8.59 to 50.97	9	35.12	16.06 to 66.68
Total	472		26			69		

age ($P<0.001$). Figure 2 shows the smoothed hazard estimates over time for the 2 age groups. The risk of stopping warfarin peaked early and then, beginning at 6 months, approximated that of younger patients. With death and maintenance of sinus rhythm excluded as reasons for warfarin cessation, 26% of patients ≥80 years of age stopped warfarin within the first year. Similar to major hemorrhage, the rate of warfarin termination for perceived safety concerns was higher among patients with $CHADS_2$ scores of ≥3 (IRR 2.27, 95% CI 1.41 to 3.66) (Table 3).

Discussion

We found the rate of major hemorrhage on warfarin to be higher than previously reported. The aggregate rate was 7.2 per 100 person-years and 13.08% versus 4.75% for patients ≥80 years of age compared with patients <80 years of age. The first 90 days were associated with a 3-fold increased risk. Although only 2% of person-years were spent in an INR range of ≥4.0, INR was a strong risk factor. The long-term tolerability of warfarin among individuals ≥80 years of age has not previously been assessed. We found that 26% of these patients were taken off warfarin; concerns related to safety accounted for 81% of them.

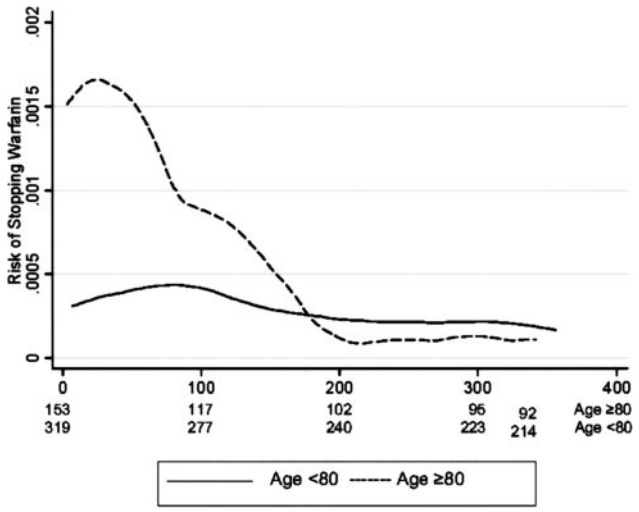


Figure 2. Risk of stopping warfarin in the first year on the basis of perceived safety concerns by age. Numbers below graph are the number of patients on warfarin at that time point ($P<0.001$, log-rank test).

The higher rate of major hemorrhage is likely attributable to the advanced age of our study population and restriction of our cohort to patients who started warfarin (Table 4). A total of 32% (n=153) of patients were ≥80 years of age compared with a total of 20 patients >75 in the pooled analysis of the first 5 randomized trials. Higher rates of hemorrhage were found in the Stroke Prevention in Atrial Fibrillation II Study, which consisted of 2 parallel trials of patients aged >75 years and ≤75 years.[27] Of note, the trials used an INR target range of 2.0 to 4.5. Annual rates of major bleeding were 4.2% and 1.7%, respectively. In our study, we reaffirmed the time-dependence of risk that has been shown in other studies. Underestimation of early adverse events and early cessation of warfarin among higher-risk patients likely contribute to the lower rates of bleeding reported from studies with prevalent user designs that largely reflect a warfarin-tolerant subset. This issue is particularly relevant given recent trials that have enrolled up to 84% of patients on warfarin at baseline.[20–22] Given the findings of our study, differences in treatment effect would be anticipated between prevalent users and those new to therapy.[22] New warfarin status may also help to explain the finding of increased bleeding among patients with a first episode of AF.[15] Our study highlights the importance of new-user designs in the evaluation of newer antithrombotic drugs to minimize survivor bias and to better assess drug effects in the early period.[28,29] The rate of major hemorrhage found in our study is similar to that of other inception cohorts with several caveats (Table 4). These studies included patients with different indications and higher INR target intensities and, in the Landefeld[16] and Beyth[26] studies, 100% of patients were identified at hospital discharge. In contrast, our study was restricted to patients with AF, target INR of 2.0 to 3.0, and 33% were identified at discharge. The distinct difference is the age of the patients enrolled: 6% versus 32% were ≥80 years of age. Elevated INR is a firmly established risk factor for hemorrhage. The percent of person-years spent in an INR range >3.0 in our study, 13%, was nearly identical to that reported in 2 recent trials, 12%[21] and 15%,[22] which emphasizes the quality of anticoagulation that was achieved.

A potential limitation of our study is that it was conducted at a single academic center that may not reflect other settings. However, the rate of bleeding found in our study may be an underestimate of that experienced in practice. We did not include patients whose warfarin was managed outside of our

TABLE 4. Rates of Major Hemorrhage by Study Design, Age, and Proportion of Patients New to Warfarin

Study Design	Year Published	Indication	Rate of Major Hemorrhage, % (n)*	Newly Starting Warfarin, %	Mean Age	Age ≥80 Years % (n)
Randomized trials						
AFI[2]	1994	AF	1.3 (24)	100	69	(20 > age 75)
SPAF II[27] (2 age strata)	1994	AF	1.7 (19)	100	NR	(358 ≤ age 75)
			4.2 (17)	100	80	(197 > age 75)
AFFIRM[15]	2002	AF	2.0 (136)†	NR	70	NR
SPORTIF III[20]	2003	AF	2.2 (50)	27	70	11 (195)
SPORTIF V[21]	2005	AF	3.4 (93)	15	72	19 (373)
Inception cohort						
Landefeld[16]	1989	All	7.4 (65)	100	61	6 (32)
Steffensen[18]	1997	All	6.0 (42)	100	59 F/66 M	14 (95)
Beyth[26]	1998	All	5.0 (22)	100	60	6 (16)
Current study	...	AF	7.2 (26)	100	77	32 (153)
Noninception cohort (prevalent warfarin use)						
van der Meer[14]	1993	All	2.7 (162)	NR	66	NR
Fihn[11]	1996	All	1.0 (37)	NR	58	4 (93)
ATRIA[12]	2003	AF	1.5 (196)	NR	71	23 (2651)‡

AFI indicates Atrial Fibrillation Investigators; AFFIRM, Atrial Fibrillation Follow-up Investigation of Rhythm Management; NR, not reported; SPORTIF, Stroke Prevention using an Oral Thrombin Inhibitor in Atrial Fibrillation; ATRIA, Anticoagulation and Risk factors in Atrial Fibrillation; F, female; and M, male.

*Major hemorrhage was most commonly defined as fatal, involving a critical site, or hospitalization with ≥2 units transfused.

†Data provided are for patients assigned to the rate-control strategy.

‡Includes patients on and off warfarin.

clinic. Because the clinic does not manage the warfarin of long-term care residents, these potentially frailer patients were not included. Patients had to have their longitudinal care established at the study institution to minimize referral bias, and those with higher baseline risk of bleeding (eg, acute stroke, postoperative AF) were not included. We do not believe that physicians were overly aggressive in the use of warfarin, given the lower prevalence of risk factors for hemorrhage at baseline compared with the high prevalence of risk factors for ischemic stroke. In addition, as we had previously reported, the prevalence of risk factors for bleeding was considerably higher among patients not prescribed warfarin at hospital discharge compared with that of patients enrolled in the current study: prior hemorrhage 32% versus 10%, prior fall 38% versus 4%, cognitive impairment 22% versus 3%, renal dysfunction 29% versus 11%.[8] Only 3 of 26 patients who sustained a major hemorrhage would have been classified as high risk by the Outpatient Bleeding Risk Index. However, it is of note that this index was derived from a cohort with a mean age of 61 years, in contrast to the mean age of 77 years in our study.

Despite these limitations, our study has several key strengths. We consecutively identified and followed all eligible patients from the first day of warfarin to ensure complete capture of adverse events. Warfarin management was provided by an experienced staff who worked in a long-established and well-organized clinic. We were able to directly ascertain use of aspirin, an important and often underreported confounder because of its over-the-counter status. Of the 472 patients, 32% were ≥80 years of age, which constitutes the largest proportion of patients in this age group represented in a prospective real-world study of AF.

Implications for Management of Anticoagulation in the Elderly Patient

In our study, patients at highest risk of stroke also experienced most of the bleeding, which illustrates the clinical complexity of these patients. Because cardioembolic strokes are associated with a 30-day mortality of 24% and significant disability among those who survive,[30,31] decisions not to prescribe warfarin for patients in this high-risk group should be largely influenced by hemorrhagic outcomes that are equal in magnitude to the irreversible sequelae of an ischemic stroke and not indiscriminate comparison of stroke rates to aggregate rates of bleeding. The rate of intracranial hemorrhage in our study was 2.5% (95% CI 1.1 to 4.7) and 1.7% for intracerebral hemorrhage. Strategies to decrease this risk need to be aggressively sought and implemented. Control of blood pressure has been shown to significantly reduce this risk.[32] The benefit of the addition of aspirin to warfarin for cardiovascular disease needs to be rigorously defined and justified.[33–35] Recent guidelines advise against this combination for older patients with AF and stable coronary disease.[36] Interventions to reduce falls are important as these patients experience higher rates of intracranial hemorrhage, 2.8% (95% CI 1.9 to 4.1).[37,38] Although use of lower INR targets has been suggested to offset bleeding risk, this strategy has not been shown to decrease hemorrhage.[31,36,39–41] An INR interval of 2.0 to 2.5 in theory seems like a rational compromise, but this degree of precision would be difficult to achieve in practice given warfarin's variable dose re-

sponse.[31,42] Because elderly patients are slower to normalize an elevated INR, more aggressive management of excessive anticoagulation in this age group seems prudent.[43] Vigilant monitoring in the early phase of therapy will help reduce bleeding during this risk-prone period. Further research of age-related vasculopathies that predispose to intracerebral hemorrhage is needed.[33,39,44–46] A reduction of the risk of extracranial hemorrhage is critically important as these events are associated with significant morbidity and often precipitate termination of therapy. Thresholds for discontinuation of warfarin may differ among physicians, patients, and families and thus warrant further study.[47] Our findings also highlight the need for continued research into the mechanisms of AF and precipitants of thrombus formation, insights that may elucidate novel pathways for stroke prevention without the attendant risk of hemorrhage.[48]

Conclusions

Published rates of major hemorrhage derived from younger noninception cohorts underestimate the bleeding that occurs in clinical practice. Higher rates of bleeding and short-term tolerability likely contribute to the reported underutilization of warfarin among elderly patients with AF. Stroke prevention among the highest-risk patients remains a challenge. Given the aging of the population, stroke prevention in AF is a pressing health concern.

Acknowledgments

The authors would like to thank Dr Robert Hughes, director of the Massachusetts General Hospital Anticoagulation Management Service, and the clinic nurses for their dedication to patient care.

Source of Funding

The Robert Wood Johnson Foundation Generalist Physician Faculty Scholars Program Grant No. 039174.

Disclosures

Dr Hylek has served on advisory boards for Bristol-Myers Squibb and has received research support from AstraZeneca and Bristol-Myers Squibb. The other authors report no conflicts.

References

1. Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. *Ann Intern Med*. 1999;131:492–501.
2. 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;154:1449–1457.
3. Bungard TJ, Ghali WA, Teo KK, McAlister FA, Tsuyuki RT. Why do patients with atrial fibrillation not receive warfarin? *Arch Intern Med*. 2000;160:41–46.
4. Gage BF, Boechler M, Doggette AL, Fortune G, Flaker GC, Rich MW, Radford MJ. Adverse outcomes and predictors of underuse of antithrombotic therapy in Medicare beneficiaries with chronic atrial fibrillation. *Stroke*. 2000;31:822–827.
5. Jencks SF, Cuerdon T, Burwen DR, Fleming B, Houck PM, Kussmaul AE, Nilasena DS, Ordin DL, Arday DR. Quality of medical care delivered to Medicare beneficiaries: a profile at state and national levels. *JAMA*. 2000;284:1670–1676.
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;348:2635–2645.
7. Waldo AL, Becker RC, Tapson VF, Colgan KJ. Hospitalized patients with atrial fibrillation and a high risk of stroke are not being provided with adequate anticoagulation. *J Am Coll Cardiol*. 2005;46:1729–1736.
8. Hylek EM, D'Antonio J, Evans-Molina C, Shea C, Henault LE, Regan S. Translating the results of randomized trials into clinical practice: the challenge of warfarin candidacy among hospitalized elderly patients with atrial fibrillation. *Stroke*. 2006;37:1075–1080
9. Thom T, Haase N, Rosamond W, Howard VJ, Rumsfeld J, Manolio T, Zheng ZJ, Flegal K, O'Donnell C, Kittner S, Lloyd-Jones D, Goff DC Jr, Hong Y, Adams R, Friday G, Furie K, Gorelick P, Kissela B, Marler J, Meigs J, Roger V, Sidney S, Sorlie P, Steinberger J, Wasserthiel-Smoller S, Wilson M, Wolf P; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. American Heart Association. Heart and Stroke Statistics–2006 Update [published corrections appears in *Circulation.* 2006;113:e696 and *Circulation*. 2006;114:e630]. *Circulation*. 2006;113:e85–e151.
10. Miyasaka Y, Barnes ME, Gersh BJ, Cha SS, Bailey KR, Abhayaratna WP, Seward JB, Tsang TSM. Secular trends in incidence of atrial fibrillation in Olmstead County, Minnesota, 1980–2000, and implications on the projections for future prevalence. *Circulation*. 2006;114:119–125.
11. 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. *Ann Intern Med*. 1996;124:970–979.
12. Go AS, Hylek EM, Chang Y, Phillips KA, Henault LE, Capra AM, Jensvold NG, Selby JV, Singer DE. Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? *JAMA*. 2003;290:2685–2692.
13. Palareti G, Leali N, Coccheri S, Poggi M, Manotti C, D'Angelo A, Pengo V, Erba N, Moia M, Ciavarella N. Bleeding complications of oral anticoagulant treatment: an inception-cohort, prospective collaborative study (ISCOAT). *Lancet*. 1996;348:423–428.
14. 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.
15. DiMarco JP, Flaker G, Waldo AL, Corley SD, Greene HL, Safford RE, Rosenfeld LE, Mitrani G, Nemeth M. Factors affecting bleeding risk during anticoagulant therapy in patients with atrial fibrillation: Observations from the AFFIRM Study. *Am Heart J*. 2005;149:650–656.
16. Landefeld CS, Goldman OL. Major bleeding in outpatients treated with warfarin: incidence and prediction by factors known at the start of outpatient therapy. *Am J Med*. 1989;87:144–152.
17. Fihn SD, McDonell M, Martin D. Risk factors for complications of chronic anticoagulation: a multicenter study. *Ann Intern Med*. 1993;118:511–520.
18. Steffensen F, Kristensen K, Ejlersen E, Dahlerup J, Sorensen H. Major haemorrhagic complications during oral anticoagulant therapy in a Danish population-based cohort. *J Intern Med*. 1997;242:497–503.
19. Petitti DB, Strom BL, Melmon KL. Duration of warfarin anticoagulant therapy and the probabilities of recurrent thromboembolism and hemorrhage. *Am J Med*. 1986;81:255–259.
20. Executive Steering Committee for the SPORTIF III Investigators. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation. *Lancet*. 2003;362:1691–1698.
21. Executive Steering Committee for the SPORTIF V Investigators. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation. *JAMA*. 2005;293:690–698.
22. Active Writing Group on behalf of the Active Investigators. Clopidogrel plus aspirin versus oral anticoagulation for atrial fibrillation in the Atrial fibrillation Clopidogrel Trial with Irbesartan for prevention of Vascular Events. *Lancet*. 2006;367:1903–1912.
23. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA*. 2001;285:2864–2870.
24. 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.
25. US Preventive Services Task Force. Aspirin for the primary prevention of cardiovascular events: recommendation and rationale. *Ann Intern Med*. 2002;136:157–160.
26. 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.
27. SPAF Investigators. Warfarin versus aspirin for prevention of thromboembolism in atrial fibrillation: Stroke Prevention in Atrial Fibrillation II Study. *Lancet*. 1994;343:687–691.

28. Feinstein A. Clinical biostatistics. XI. Sources of 'chronology bias' in cohort statistics. *Clin Pharmacol Ther*. 1971;12:864–879.
29. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol*. 2003;158:915–920.
30. Lin H, Wolf PA, Kelly-Hayes M, Beiser AS, Kase CS, Benjamin EJ, D'Agostino RB. Stroke severity in atrial fibrillation: the Framingham Study. *Stroke*. 1996;27:1760–1764.
31. Hylek EM, Go AS, Chang Y, Jensvold NG, Henault LE, Selby JV, Singer DE. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. *N Engl J Med*. 2003;349:1019–1026.
32. Arima H, Hart RG, Colman S, Chalmers J, Anderson C, Rodgers A, Woodward M, MacMahon S, Neal B, 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;36:2164–2169.
33. Hart RG, Tonarelli SB, Pearce LA. Avoiding central nervous system bleeding during antithrombotic therapy: recent data and ideas. *Stroke*. 2005;36:1588–1593.
34. Shireman TI, Howard PA, Kresowik TF, Ellerbeck EF. Combined anticoagulant-antiplatelet use and major bleeding events in elderly atrial fibrillation patients. *Stroke*. 2004;35:2362–2367.
35. Buresly K, Eisenberg M, Zhang X, Pilote L. Bleeding complications associated with combinations of aspirin, thienopyridine derivatives, and warfarin in elderly patients following acute myocardial infarction. *Arch Intern Med*. 2005;165:784–789.
36. 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 and the European Society of Cardiology Committee on Practice Guidelines. *Circulation*. 2006;114:e257–354.
37. 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;118:612–617.
38. Tinetti ME. Clinical practice: preventing falls in elderly persons. *N Engl J Med*. 2003;348:42–49.
39. Fang MC, Chang Y, Hylek EM, Rosand J, Greenberg SM, Go AS, Singer DE. Advanced age, anticoagulation intensity, and risk for intracranial hemorrhage among patients taking warfarin for atrial fibrillation. *Ann Intern Med*. 2004;141:745–752.
40. SPAF Investigators. Adjusted-dose warfarin versus low-intensity, fixed-dose warfarin plus aspirin for high-risk patients with atrial fibrillation: Stroke prevention in atrial fibrillation III randomised clinical trial. *Lancet*. 1996;348:633–638.
41. Kearon C, Ginsberg JS, Kovacs MJ, Anderson DR, Wells P, Julian JA, MacKinnon B, Weitz JI, Crowther MA, Dolan S, Turpie AG, Geerts W, Solymoss S, van Nguyen P, Demers C, Kahn SR, Kassis J, Rodger M, Hambleton J, Gent M. Comparison of low-intensity warfarin therapy with conventional-intensity for long-term prevention of recurrent venous thromboembolism. *N Engl J Med*. 2003;349:631–639.
42. Oden A, Fahlen M, Hart R Optimal INR for prevention of stroke and death in atrial fibrillation: a critical appraisal. *Thromb Res*. 2006;117:493–499.
43. Hylek E, Regan S, Go A, Hughes R, Singer D, Skates S. 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.
44. Rosand J, Hylek EM, O'Donnell HC, Greenberg SM. Warfarin-associated hemorrhage and cerebral amyloid angiopathy: a genetic and pathologic study. *Neurology*. 2000;55:947–951.
45. Smith EE, Rosand J, Knudsen KA, Hylek EM, Greenberg SM. Leukoaraiosis is associated with warfarin-related hemorrhage following ischemic stroke. *Neurology*. 2002;59:193–197.
46. Hylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. *Ann Intern Med*. 1994;120:897–902.
47. Farrell JJ, Friedman LS. Gastrointestinal bleeding in older people. *Gastroenterol Clin North Am*. 2000;29:1–36.
48. Aviles RJ, Martin DO, Apperson-Hansen C, Houghtaling PL, Rautaharju P, Kronmal RA, Tracy RP, Van Wagoner DR, Psaty BM, Lauer MS, Chung MK. Inflammation as a risk factor for atrial fibrillation. *Circulation*. 2003;108:3006–3010.

CLINICAL PERSPECTIVE

Warfarin is effective in the prevention of stroke in atrial fibrillation but is under used in clinical care. Concerns exist that published rates of hemorrhage may not reflect real-world practice. Few patients ≥80 years of age were enrolled in trials, and studies of prevalent use largely reflect a warfarin-tolerant subset. We sought to define the tolerability of warfarin among an elderly inception cohort with atrial fibrillation. Consecutive patients ≥65 years of age who started warfarin were identified and followed for 1 year. Study outcomes included major hemorrhage, time to termination of warfarin, and reason for discontinuation. Of 472 patients enrolled, 32% were ≥80 years of age, and 91% had ≥1 stroke risk factor. The cumulative incidence of major hemorrhage for patients ≥80 years of age was 13.1 per 100 person-years and 4.7 for those <80 years of age ($P=0.009$). Despite trial-level anticoagulation control, we found that the first 90 days of warfarin, age ≥80 years, and international normalized ratio (INR) ≥4.0 were each associated with increased risk of hemorrhage. Within the first year, 26% of patients ≥80 years of age stopped taking warfarin; perceived safety issues accounted for 81% of these. Rates of major hemorrhage and warfarin termination were highest among patients with $CHADS_2$ scores (an acronym for congestive heart failure, hypertension, age ≥75, diabetes mellitus, and prior stroke or transient ischemic attack) of ≥3. Rates of hemorrhage derived from younger noninception cohorts underestimate the bleeding that occurs in real-world practice. This finding, coupled with the short-term tolerability of warfarin, likely contributes to its underutilization. Stroke prevention among elderly patients with atrial fibrillation remains a challenging and pressing health concern.

Go to http://cme.ahajournals.org to take the CME quiz for this article.

REVIEW ARTICLE

Magnetic Resonance Imaging for Diagnosing Foot Osteomyelitis

A Meta-analysis

Alok Kapoor, MD; Stephanie Page, MD; Michael LaValley, PhD; Daniel R. Gale, MD; David T. Felson, MD, MPH

Background: Uncertainty exists regarding the optimal workup of patients with suspected osteomyelitis of the foot, many of whom have diabetes mellitus. We conducted a meta-analysis to determine the diagnostic test performance of magnetic resonance imaging (MRI) for osteomyelitis of the foot and compared this performance with that of technetium Tc 99m bone scanning, plain radiography, and white blood cell studies.

Methods: We searched MEDLINE (from 1966 to week 3 of June 2006) and EMBASE (from 1980 to week 3 of June 2006) for English-language studies in which adults suspected of having osteomyelitis of the foot or ankle were evaluated by MRI. We then extracted data using a standard form derived from the Cochrane Methods Group. To summarize the performance of diagnostic tests, we used the summary receiver operating characteristic curve analysis, which relies on the calculation of the diagnostic odds ratio (DOR). We also examined subsets of studies defined by the presence or absence of particular design flaws or populations.

Results: Sixteen studies met inclusion criteria. In all studies combined, the DOR for MRI was 42.1 (95% confidence interval, 14.8-119.9), and the specificity at a 90% sensitivity cut point was 82.5%. The DOR did not vary greatly among subsets of studies. In studies in which a direct comparison could be made with other technologies, the DOR for MRI was consistently better than that for bone scanning (7 studies—149.9 vs 3.6), plain radiography (9 studies—81.5 vs 3.3), and white blood cell studies (3 studies—120.3 vs 3.4).

Conclusions: We found that MRI performs well in the diagnosis of osteomyelitis of the foot and ankle and can be used to rule in or rule out the diagnosis. Magnetic resonance imaging performance was markedly superior to that of technetium Tc 99m bone scanning, plain radiography, and white blood cell studies.

Arch Intern Med. 2007;167:125-132

Author Affiliations: Division of General Internal Medicine and Clinical Epidemiology Research and Training Unit, Boston University (Drs Kapoor, Page, LaValley, and Felson), Boston, and Commonwealth Radiology Group, Salem (Dr Gale), Mass. Dr Page is now with the Department of Hospitalist Medicine, Mount Auburn Hospital, Cambridge, Mass.

OSTEOMYELITIS OF THE FOOT and ankle is the primary or secondary reason for 75 000 hospitalizations in the United States each year.[1] By far the most common group at risk is persons with diabetes mellitus. In terms of diagnostic evaluations, history and routine laboratory tests, including the erythrocyte sedimentation rate, are not particularly informative.[2,3] Although bone biopsy serves as a gold standard diagnostic test and is generally safe, the fear of introducing infection and the need for a surgical practitioner to perform the biopsy make development of diagnostic algorithms using noninvasive imaging strategies attractive.

Plain radiography is the traditional and often the initial modality used for evaluating bone infections in the foot. Radiographic changes are often not visible until 2 to 4 weeks after onset of infection, accounting in part for the low sensitivity of plain radiography.[4-6] The specificity of plain radiography tends to be higher than its sensitivity but can be compromised by posttraumatic reactions, nonspecific periosteal reactions as seen in chronic venous stasis, and most commonly Charcot osteoarthropathy. Charcot osteoarthropathy, or Charcot foot, is a disruption in foot architecture that results from microtrauma to an insensate foot. It is often indistinguishable from osteomyelitis.

Bone scanning with technetium Tc 99m [^{99m}Tc]–labeled diphosphonate can detect early changes of osteomyelitis but suffers from lack of specificity. White blood cell (WBC) scanning, usually with indium In 111, is more specific but lacks sensitivity.[5] In addition, WBC scanning requires the inconvenient and time-consuming process of drawing and incubating patient blood before reinjecting and obtaining images.

Diagnostic findings of pedal osteomyelitis on MRI include a focal area of de-

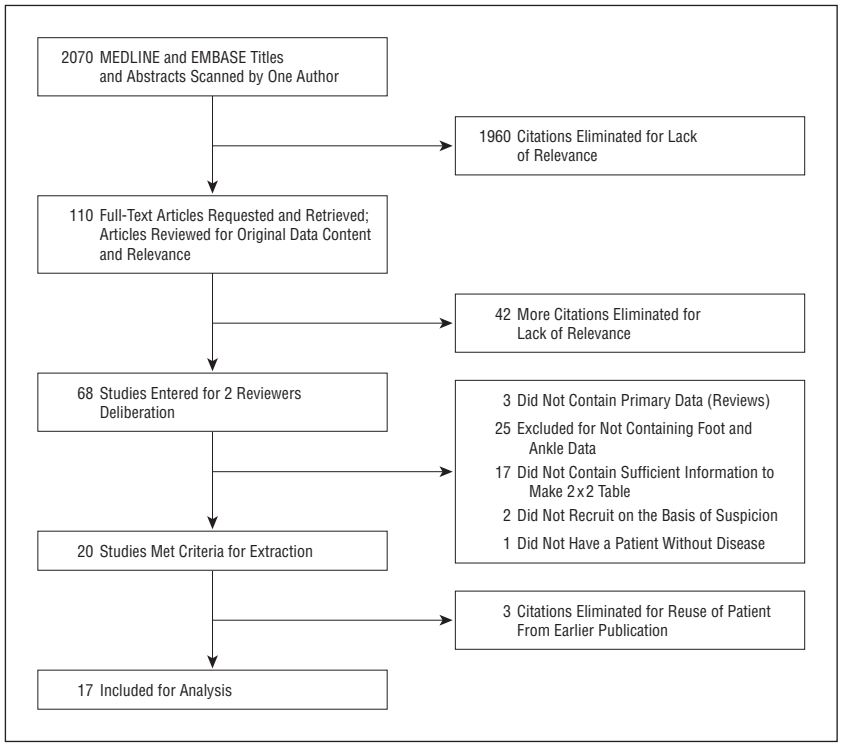


Figure 1. Flowchart of article selection.

creased marrow signal intensity on T1-weighted images and a focally increased signal intensity on fat-suppressed T2-weighted or short tau inversion recovery images.[7] The MRI changes seen in osteomyelitis may be confused with changes seen in bony infarcts, fractures, and Charcot foot.

Accurate estimates of MRI test performance for osteomyelitis of the foot are difficult to establish. Most studies have reported on small cohorts, combined persons with suspected osteomyelitis of the foot and those with suspected osteomyelitis of other body sites, included persons both with and without diabetes, and left unstated the prevalence of Charcot foot. Studies have drawn different conclusions about the value of MRI alone (or compared with other technologies) and have reported vastly different estimates of diagnostic specificity (0%-100%). Previous systematic reviews[8-10] were limited by the number of publications they analyzed or the lack of foot-specific information.

We conducted a comprehensive meta-analysis of the test performance of MRI for the diagnosis of osteomyelitis of the foot and ankle. We then conducted subset analyses to explore the reason for variability among included studies. We also compared the accuracy of MRI with ^{99m}Tc bone scanning, plain radiography, and WBC scanning.

METHODS

STUDY IDENTIFICATION

We searched MEDLINE (from 1966 to week 3 of June 2006) and EMBASE (from 1980 to week 3 of June 2006) for English-language articles. (The complete search strategy is available from the authors on request.) We also searched the bibliographies of included studies and asked specialists within the fields of surgery and radiology to recommend citations.

STUDY SELECTION

We included studies that evaluated the diagnostic test performance of MRI in adult patients suspected of having osteomyelitis of the foot or ankle or who had foot infection and were systematically examined for osteomyelitis. Specifically, studies were enrolled when information from the usual diagnostic performance 2 × 2 table (index diagnostic test result positive or negative vs true disease state present or absent) could be extracted about discrete foot and ankle cases, when 80% or more of the patients were 16 years or older, and when at least one site with the disease and one without were identified by the reference standard (eg, bone biopsy). Two authors (A.K. and S.P.) evaluated each study for inclusion, and a third author (D.T.F.) refereed ties.

STUDY EXTRACTION AND QUALITY ASSESSMENT

Once studies were selected, we used a data extraction instrument derived from the Cochrane Methods Group checklist on Systematic Review of Screening and Diagnostic Tests.[11] Two independent reviewers (A.K. and S.P.) extracted data that pertained to study population characteristics. The prevalence of diabetes in each study was noted. To understand better the quality of the sensitivity and specificity estimates reported in each study, we also extracted information about blinding and the type of reference standard used. Specifically, we calculated the frequency that a bone biopsy–based reference standard was used to determine or exclude osteomyelitis. For a positive disease determination, positive histologic analysis or culture results (however it was determined by the study authors) from a bone specimen were recorded. For the negative reference standard, we recorded the percentage of patients in whom osteomyelitis was excluded by negative histologic analysis results (however it was determined by the authors). If an individual study characteristic was not explicitly documented, no determination was made regarding the study status and a "not specified" label was assigned to the study for that characteristic. No attempt was made to contact study authors except to determine whether the same patients were enrolled more than once by authors with multiple publications.

COMPARISON OF DIAGNOSTIC IMAGING TESTS

We extracted data on the diagnostic performance of bone scan, plain radiography, and WBC studies if a 2 × 2 diagnostic performance table could be derived from a study that was already included for its MRI data.

DATA SYNTHESIS AND ANALYSIS

For each study, we constructed a 2 × 2 contingency table that consisted of true-positive, false-positive, false-negative, and true-negative results according to the reference standard used in each case. We then calculated the sensitivity and specificity in the usual fashion and the diagnostic odds ratio (DOR), as determined by the formula (true positive × true negative)/(false positive × false negative). We

Table 1. Characteristics of Included Studies

Source	Enrollment Criteria	No. of Sites in the Study (No. of Patients)*	Mean Age of Patients, y	Prevalence of Diabetes, %	Prospective Design	Consecutive Enrollment	MRI Assessors†
Craig et al,[14] 1997	Diabetic patients scheduled for partial amputation	57 (13)	57.0	100.0	Yes	NS	NS
Croll et al,[15] 1996	Patients admitted with nongangrenous diabetic foot infections	27 (27)	66.0	100.0	Yes	NS	No
Enderle et al,[16] 1999	Diabetic patients suspected of having chronic osteomyelitis from random surgery and medicine clinics	19 (19)	60.7	100.0	Yes	Yes	Yes
Ertugrul et al,[17] 2006	Diabetic patients with ulcers at Wagner grade ≥3‡	31 (31)	62.0	100.0	Yes	NS	NS
Horowitz et al,[18] 1993	Patients admitted with diabetic foot infections	47 (41)	54.4	100.0	Yes	NS	No
Kearney et al,[19] 1999	Diabetic outpatients suspected of having osteomyelitis	13 (13)	59.0	100.0	Yes	NS	NS
Ledermann et al,[20] 2002	Diabetic and nondiabetic patients suspected of osteomyelitis	84 (72)	NS	NS	NS	NS	Yes
Levine et al,[21] 1994	Diabetic patients with suspected osteomyelitis complicating soft tissue infection	29 (27)	51.6	100.0	No	NS	No
Lipman et al,[22] 1998	Patients with peripheral neuropathy and high clinical suspicion	20 (20)	46.0	85.0	Yes	Yes	Yes
Maas et al,[23] 2002	Patients with neuropathy and inflammation in addition to leprosy	18 (12)	63.0	0	No	Yes	Yes
Morrison et al,[24] 1998	Patients suspected of having osteomyelitis of the foot	73 (62)	56.0	84.9	No	NS	Yes
Nigro et al,[25] 1992	Patients with foot inflammation and possible osteomyelitis	47 (44)	55.0	70.5	NS	NS	NS
Remedios et al,[26] 1998	Diabetic patients with peripheral neuropathy, chronic foot ulcers, and signs of osteomyelitis	9 (9)	57.0	100.0	Yes	NS	NS
Seabold et al,[27] 1990	Patients highly suspected of having osteomyelitis in and around Charcot joint	12 (11)	50.6	91.7	No	No	No
Vesco et al,[28] 1999	Diabetic patients with foot ulcers	24 (24)	59.0	100.0	Yes	Yes	NS
Weinstein et al,[29] 1993	Diabetic patients admitted with suggestion of osteomyelitis, nonhealing ulcer, or soft tissue infection	75 (47)	49.6	100.0	Yes	Yes	Yes
Yuh et al,[30] 1989	Patients suspected of having osteomyelitis or nonhealing ulcer	44 (24)	58.2	100.0	NS	Yes	Yes

Abbreviations: MRI, magnetic resonance imaging; NS, not specified.

*Site refers to site in the body. If each patient had 1 site suggestive of osteomyelitis, the number of sites would equal the number of patients. Because the included studies typically did not document performance estimates at the patient level, we calculated the summary estimate of performance at the level of the site. Later, we examined the summary estimate in a subset of studies in which the number of multiple sites for the same patient was low or nil. In addition, the number of sites listed corresponds to that for the entire study. Individual imaging tests may not have been performed in all cases. Please see "Comparison of Diagnostic Imaging Tests" subsection in the "Results" section.

†Blinded to other tests and to reference standard.

‡Wagner grading system is a clinical tool for evaluating diabetic foot ulcers. Scoring ranges from 1 to 5 for progressively deeper ulcers and less salvageable feet; grade 3 lesions are associated with osteomyelitis and/or abscess. Stage 4 and 5 ulcers indicated gangrenous lesions.

then conducted a summary receiver operating characteristic curve analysis as our meta-analytic method. This method has been described before.[12,13] We repeated this analysis in 13 subsets that represented different study populations (eg, low or unspecified prevalence of Charcot foot) and the presence or absence of design flaws (eg, no blinding).

COMPARISON OF DIAGNOSTIC IMAGING TESTS

We compared the head-to-head test performance of MRI with 3 other imaging tests. Because we had collected data on all MRI diagnostic studies, we evaluated these other technologies compared with MRI, which was our focus in this study. We included only studies in which 1 of the 3 diagnostic modalities was compared with MRI. To make comparisons, we used the same summary receiver operating characteristic method mentioned earlier. In certain studies not every patient underwent each test being compared, perhaps because a diagnosis was reached when the patient underwent the first diagnostic test, making the next one unnecessary. To account for this bias, we also measured the performance of the subset of studies in which all (or nearly all) patients underwent both diagnostic tests being compared. All statistical procedures were performed using SAS statistical software, version 9.1.3 (SAS Institute Inc, Cary, NC).

RESULTS

Our search strategy yielded 2070 titles with and without abstracts. One author (A.K.) reviewed them and requested 110 articles for full-text review. After eliminating those that did not meet the inclusion criteria (**Figure 1**), we were left with the 17 studies described in **Table 1**.[14-31] One study[27] examined only patients with

Table 2. MRI Diagnostic Criteria, Frequency of Biopsy Use, and MRI Performance

Source	Signs on MRI Used to Determine Positive Result*	Biopsy Reference Standard, %		Prevalence of Osteomyelitis, %	MRI Performance	
		Positive†	Negative‡		Sensitivity, %	Specificity, %
Craig et al,[14] 1997	Primary, half of patients getting T1 with gadolinium, and soft tissue mass or ulcer	100.0	100.0	36.8	90.5	65.0
Croll et al,[15] 1996	NS	100.0	55.6	33.3	88.9	100.0
Enderle et al,[16] 1999	Increased uptake on STIR with ulcer or soft tissue mass and T1 with gadolinium§	100.0	100.0	73.7	100.0	75.0
Ertugrul et al,[17] 2006	Decreased T1, turboinversion recovery magnitude, and T1 with gadolinium	100.0	100.0	74.2	78.2	60.0
Horowitz et al,[18] 1993	Increased TR/TE or increased T2 and some gadolinium, with or without cortical disruption	100.0	NS	31.9	100.0	100.0
Kearney et al,[19] 1999	NS	NS	0	69.2	100.0	50.0
Ledermann et al,[20] 2002	Primary, T1 with gadolinium, and many secondary signs	100.0	100.0	63.1	90.6	83.9
Levine et al,[21] 1994	Standard primary	100.0	31.3	44.8	76.9	100.0
Lipman et al,[22] 1998	Primary and many secondary	100.0	20.0	75.0	77.3	40.0
Maas et al,[23] 2002	Standard primary, T1 with gadolinium, and many secondary‖	50.0	0	88.9	100.0	50.0
Morrison et al,[24] 1998	Standard primary, T1 with gadolinium for most patients, and many secondary signs‖	100.0	0	58.9	91.3	83.2
Nigro et al,[25] 1992	NS	92.3	NS	55.3	100.0	95.2
Remedios et al,[26] 1998	Standard primary	100.0	20.0	44.4	100.0	80.0
Seabold et al,[27] 1990	Standard primary	100.0	0	36.4	100.0	0¶
Vesco et al,[28] 1999	Standard primary, T1 with gadolinium, and ulcer, sinus tract, or soft tissue mass	0	0	54.2	100.0	81.8
Weinstein et al,[29] 1993	Standard primary	100.0	55.2	61.3	100.0	79.3
Yuh et al,[30] 1989	Standard primary	100.0	21.1	61.4	100.0	89.5

Abbreviations: MRI, magnetic resonance imaging; NS, not specified; STIR, short tau inversion recovery; TR/TE, repetition time/echo time.

*Standard primary refers to the presence of focally decreased marrow signal on T1-weighted images and focally increased marrow signal on T2-weighted images or STIR images. T1 with gadolinium refers to use of enhancement with gadolinium contrast in fat suppressed T1-weighted images. Secondary signs include cortical disruption and adjacent cutaneous ulcer or soft tissue mass plus the presence of a sinus tract and in some cases adjacent soft tissue inflammation or edema.

†The frequency of the use of positive bone histologic analysis results or bone culture (however it was determined by the study authors) as the reference standard to confirm disease was calculated.

‡The frequency of the use of negative bone histologic analysis results (however it was determined by the study authors) as the reference standard to exclude disease was calculated.

§T1 with gadolinium was recorded but not included in determination of positive scan result.

‖In these cases, authors did not provide a single composite sensitivity and specificity, so only primary signs were analyzed.

¶All of the study subjects had Charcot joint; this may explain the low specificity.

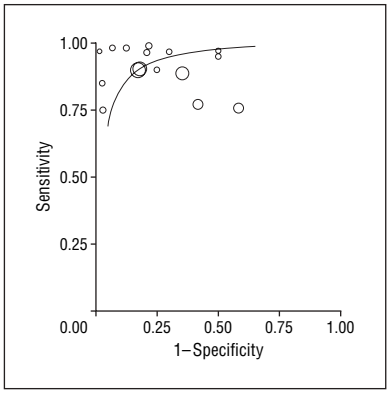


Figure 2. Summary receiver operating characteristic curve of magnetic resonance imaging performance in 16 studies. Bubble size represents sample size.

suspected osteomyelitis in or around a Charcot joint. Although this was not an a priori exclusion criterion, we believed the study was not consistent with the intent of our analysis. We chose to provide a summary estimate of the performance of MRI in a typical patient population at risk rather than one in which the population is artificially enriched with problem cases, and so the study was eliminated from analysis.

Eleven of the 16 studies involved almost exclusively diabetic patients. Nine of 16 recruited patients prospectively (ie, study authors enrolled patients before any imaging tests were recorded). Most studies did not specify or standardize the exact reason for diagnostic suspicion of osteomyelitis. In many cases, it was implied by the presence of a complicated or infected foot ulcer. Indeed, foot ulcer was required or uniformly present in 6 studies. In most instances, the number of cases with Charcot disease was not reported.

Most studies judged an MRI scan to be positive by the same criterion: a lesion in the bone that showed focally decreased marrow signal intensity in T1-weighted images and

Table 3. Diagnostic Performance of 4 Technologies in Studies That Compared MRI With Another Imaging Test

Source	MRI		Technetium Tc 99m Bone Scan		Plain Radiography		WBC Scan	
	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %
Croll et al,[15] 1996	88.9	100.0	50.0	50.0	22.2	94.4	33.3	69.2
Enderle et al,[16] 1999	100.0	75.0	83.3	75.0	71.4	80.0		
Horowitz et al,[18] 1993	100.0	100.0			43.8	100.0		
Kearney et al,[19] 1999	100.0	50.0	88.9	100.0	66.7	100.0		
Levine et al,[21] 1994	76.9	100.0	100.0	25.0	60.0	81.3	80.0	28.6
Lipman et al,[22] 1998	77.3	40.0			73.3	40.0		
Nigro et al,[25] 1992	100.0	95.2	90.5	33.3	69.6	33.3		
Remedios et al,[26] 1998	100.0	81.8	100.0	0	38.5	100.0	90.9	84.6
Yuh et al,[30] 1989	100.0	89.5	94.4	18.2	75.0	60.0		

Abbreviations: MRI, magnetic resonance imaging; WBC, white blood cell.

a focally increased signal intensity in fat-suppressed T2-weighted or short tau inversion recovery images. Eight studies[14,16,18,20,22-24,28] evaluated other diagnostic signs that were sometimes termed *secondary signs,* including cortical disruption, adjacent cutaneous ulcer, soft tissue mass, presence of a sinus tract, and, in some cases, adjacent soft tissue inflammation or edema. See **Table 2** for further details.[14-31]

The prevalence of criterion standard–defined osteomyelitis averaged approximately 50%, with a range of 32% to 89%. Most authors reported results according to the number of sites with potential osteomyelitis or number of at-risk bones imaged. We calculated a ratio of the number of patients to number of sites and compared MRI performance in studies with low and high ratios. Magnetic resonance imaging sensitivity was usually high and ranged from 77% to 100%; MRI specificity ranged from 40% to 100%.

Among all studies, the DOR for MRI was 42.1 (95% confidence interval [CI], 14.8-119.9). The specificity at a clinically relevant cut point of 90% sensitivity was 82.5%. We present the curve for diagnostic performance in **Figure 2**. We found no substantial or statistically significant differences in estimates of MRI diagnostic test performance among subsets of studies (available from the authors on request). The number of studies in certain subgroups was small (eg, subgroup with Charcot prevalence >10%), with small numbers of patients represented in each. Small numbers in subsets prohibit robust conclusions. We therefore only discuss the subsets for which 8 or more studies were available for analysis. Studies that did not use bone histologic analysis to exclude disease tended to have higher performance (DOR, 67.4; 95% CI, 18.3-248.0). Studies published in 1998 or afterward reported lower performance (DOR, 25.3; 95% CI, 5.5-116.8). Most of the later studies had a prospective design and documented assessment of MRI blinded to other results.

COMPARISON OF DIAGNOSTIC IMAGING TESTS

We compared the diagnostic performance of 4 technologies in studies that compared MRI with another imaging test (**Table 3**). We found 7 studies that directly compared MRI with ^{99m}Tc bone scanning, all using the triple-phase technique. Magnetic resonance imaging performance was markedly superior (DOR, 149.9; 95% CI, 54.6-411.3) vs bone scan (DOR, 3.6; 95% CI, 1.0-13.3) (**Figure 3**). At the 90% sensitivity cut point, the specificity for MRI was 98% compared with 28.5% for technetium. Similarly, in 9 studies that compared plain radiography with MRI, MRI outperformed plain radiography (DOR, 81.5; 95% CI, 14.2-466.1 compared with DOR, 3.3; 95% CI, 2.2-5.0). In 3 studies in which MRI was compared with WBC study, the DOR for MRI was 120.3 (95% CI, 61.8-234.3) compared with 3.4 (95% CI, 0.2-62.2) for WBC studies.

COMMENT

This meta-analysis demonstrated that MRI performs well in the diagnosis of osteomyelitis of the foot and ankle in adults. Good diagnostic performance was consistent across a subset of studies of different designs and different patients. Moreover, MRI outperformed technetium, plain radiography, and WBC studies.

Although the performance of MRI was strong, our review revealed many flaws in the published literature concerning imaging tests for osteomyelitis in the foot or ankle. Few studies prospectively followed up a cohort of patients in which assessment of MRI was blinded to other imaging tests and reference standard results, and few verified the diagnosis in all cases with a biopsy. Although the estimate of performance did not change substantially within study subsets, the relatively small number of studies did not permit exploring the combined effect of multiple design issues.

In addition, the frequency of Charcot foot was not typically documented in our studies. Performance estimates could vary significantly among studies of varying prevalence of Charcot foot. In the 13 studies in which prevalence was not documented, the prevalence of Charcot foot was probably low. However, in diabetic patients with coexistent diabetic foot infection, prevalence is uncertain. Although it may be uncommon in the general diabetic

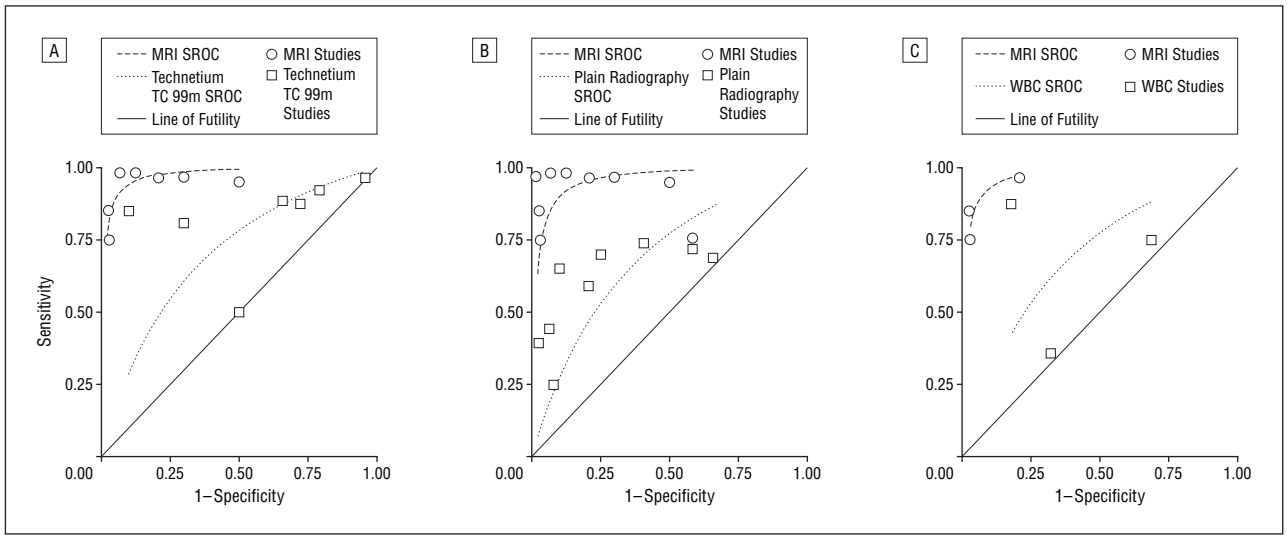


Figure 3. Head-to-head performance of magnetic resonance imaging (MRI) and technetium Tc 99m bone scanning (A), MRI and plain radiography (B), and MRI and white blood cell (WBC) scanning (C). SROC indicates summary receiver operating characteristic curve.

Table 4. Posttest Probability Stratified by Imaging Test Result Across a Spectrum of Pretest Probabilities

Pretest Probability, %	Posttest Probability, %			
	MRI Positive*	MRI Negative	Technetium Tc 99m Positive†	Technetium Tc 99m Negative
10.0	36.2	1.3	11.2	3.4
25.0	63.0	3.8	27.4	9.6
50.0	83.6	10.7	53.1	24.2
75.0	93.9	26.5	77.2	49.0
90.0	97.9	51.9	91.1	74.2

Abbreviation: MRI, magnetic resonance imaging.

*Using the all-studies estimate, which was lower than that calculated from studies that also had technetium Tc 99m data. The 90% sensitivity and 82.5% specificity translate to a positive likelihood ratio of 5.1 and a negative likelihood ratio of 0.12.

†Using a higher sensitivity threshold of 95%, which decreases the negative likelihood ratio to 0.32; the specificity at this threshold is 16% and the positive likelihood ratio is 1.13.

population, Charcot foot is likely much more prevalent among patients with peripheral neuropathy.[32] Given that the management strategies for Charcot foot and osteomyelitis are vastly different (offloading and contact casting vs long-term antibiotics), making an accurate diagnosis is essential.

This meta-analysis has 2 major implications. First, this study confirms that MRI is a strong test to aid in both confirming and excluding osteomyelitis of the foot. Using the clinically relevant cut point of 90% sensitivity, the positive likelihood ratio is 5.1 and the negative likelihood ratio is 0.12. Assuming a pretest probability of 50% (not far from the 55% calculated from all studies), a patient with a positive MRI would have an 84% chance of having the diagnosis (**Table 4**). If any other features or examination findings favor the diagnosis of osteomyelitis, such as substantial depth of ulcer or positive probe to bone, the addition of a positive MRI virtually clinches the diagnosis. A negative MRI study in our baseline hypothetical patient results in a posttest probability of 11%. Combined with absence of substantial ulcer depth or a negative probe to bone, MRI effectively rules out osteomyelitis.

The second major implication is that there should be a diminished use of ^{99m}Tc bone scanning in the diagnosis of osteomyelitis of the foot. Although bone scanning has been proposed for ruling out the disease (given its purported high sensitivity), the lack of adequate specificity creates many false-positive results.[4,6,33] Using a hypothetical prevalence of osteomyelitis of 25%, for every 100 patients subjected to bone scanning, 12 of 13 with a negative result would have the diagnosis correctly excluded (negative predictive value, 91%), but only 24 of 87 with a positive result would be correctly identified as having osteomyelitis (positive predictive value, 27%). A diagnostic algorithm that includes bone scanning would thereby result in the ordering of numerous second imaging tests or biopsies. At a lower prevalence of osteomyelitis (eg, 5%-15%), ^{99m}Tc scanning may successfully rule out disease (Table 4). However, clinicians often underestimate the prevalence of osteomyelitis, particularly in patients with a diabetic foot infection; this finding suggests that an assumption of low prevalence may be risky.[34] In addition, MRI permits detection of deep collections of pus or necrotic tissue and visualization of foot anatomy, which helps the surgeon plan surgery when indicated. The Infectious Disease Society of America has already recognized that MRI is the preferred advanced imaging test for suspected osteomyelitis but recommends performing serial plain radiography before ordering an MRI.[35] We are unaware of any study that has formally evaluated serial plain radiography vs early MRI. Such an investigation and/or cost-effectiveness analysis would likely clarify better the place for MRI in the diagnostic algorithm of os-

teomyelitis of the foot.[35] Clinicians should, of course, consider history, physical examination findings, and imaging test results before deciding on therapeutic interventions.

In this meta-analysis, we chose to focus on osteomyelitis of the foot and ankle, because disease of the foot and ankle is a distinct entity that affects a particular patient population, that is, patients with diabetes and/or peripheral neuropathy. We did not analyze non–English-language articles. We are unaware of any evidence of bias in English language studies that assessed technology.

We did not exclude studies on the basis of date of publication or advent of innovation or variation in interpretation of MRI. Although MRI evaluation of osteomyelitis has evolved, with gadolinium now often used, subset analysis based on gadolinium use did not reveal any substantial variation in performance. Secondary diagnostic signs (such as cortical breaks) appeared to be incorporated into the diagnostic algorithm for osteomyelitis more frequently in recent publications. Ahmadi et al[31] recently published a retrospective analysis of additional criteria for use in assessing osteomyelitis superimposed on Charcot foot, but this work is still largely untested.

As for the comparison of diagnostic tests, we focused on comparing MRI with plain radiography and radionuclide scanning, making 3 discrete head-to-head comparisons with MRI. A recent review by the Health Technology Assessment group supports this approach, suggesting that heterogeneity of diagnostic test comparisons will be less of a problem in head-to-head comparisons.[36] We did not analyze other imaging modalities, such as combined bone scanning and WBC study, computed tomography, immunoglobulin tagged tracer scanning, or positron emission tomography, because 2 or fewer studies directly compared them to MRI. We did not compare the performance of biopsy with MRI. A biopsy affords information regarding the exact pathogen responsible for infection, something that imaging tests cannot do.

According to 2006 figures, Medicare reimburses $288 for a 3-phase bone scan and $416 for a lower-extremity MRI without contrast ($451 with contrast).[37,38] We calculated these values on the basis of how our center bills Medicare, which is the sum of the reimbursement to our facility when providing the service to an outpatient and the reimbursement to the radiologist interpreting the film (the professional component alone). For an inpatient, Medicare reimburses the relevant diagnosis-related code; therefore, unique MRI payment information is not available. Given the small difference in cost (which is approximated by Medicare reimbursement) between MRI and ^{99m}Tc bone scan and the large difference in diagnostic performance between these technologies, MRI would be more cost-effective except when the probability of disease was low. Local availability and cost of each test must also be considered when selecting the appropriate test. Formal decision modeling is needed to fully characterize the place of MRI in the diagnostic algorithm of osteomyelitis of the foot and ankle.

In summary, MRI has a strong performance in the diagnosis of osteomyelitis of the foot and ankle in adults. It outperforms 3-phase ^{99m}Tc bone scanning and plain radiography. The role of bone scanning is probably eclipsed by that of MRI except in cases in which MRI is contraindicated or the probability of disease is low.

Accepted for Publication: September 15, 2006.

Correspondence: Alok Kapoor, MD, Division of General Internal Medicine, Boston University, 91 E Concord St, MAT 200, Second Floor, Boston, MA 02118 (alok.kapoor@bmc.org).

Author Contributions: Dr Kapoor 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:* Kapoor and Felson. *Acquisition of data:* Kapoor. *Analysis and interpretation of data:* Kapoor, Page, LaValley, Gale, and Felson. *Drafting of the manuscript:* Kapoor, Page, and Felson. *Critical revision of the manuscript for important intellectual content:* Kapoor, Page, LaValley, and Gale. *Statistical analysis:* LaValley. *Obtained funding:* Felson. *Administrative, technical, and material support:* Kapoor, Page, Gale, and Felson. *Study supervision:* Felson.

Financial Disclosure: None reported.

Funding/Support: This study was supported by National Research Service Award T-32 HP 10028-08 and by grant AR47785 from the National Institutes of Health.

Acknowledgment: We thank Gary Gibbons, MD, Charles Foster, MD, and Jorge Medina, MD, for their consultation on this project. Special thanks also to Louise Falzon, MLIS, and Joseph Harzbecker, MLS, for the guidance in preparation of search strategies.

REFERENCES

1. Agency for Healthcare Research Quality. Health Care Utilization Project (HCUP). http://hcup.ahrq.gov/HCUPnet.asp. Accessed November 9, 2005.
2. Jude EB, Selby PL, Mawer EB, Burgess J, Boulton AJM. Inflammatory and bone turnover markers in Charcot arthropathy and osteomyelitis of the feet in diabetic patients. *Diabetologia.* 2002;45(suppl 2):A341-A342.
3. Eneroth M, Larsson J, Apelqvist J. Deep foot infections in patients with diabetes and foot ulcer: an entity with different characteristics, treatments, and prognosis. *J Diabetes Complications.* 1999;13:254-263.
4. Black ER, Bordley DR, Tape TG, Panzer RJ. *Diagnostic Strategies for Common Medical Problems.* 2nd ed. East Peoria, Ill: Versa Press; 1999.
5. Jeffcoate WJ, Lipsky BA. Controversies in diagnosing and managing osteomyelitis of the foot in diabetes. *Clin Infect Dis.* 2004;39(suppl 2):S115-S122.
6. Becker W. Imaging osteomyelitis and the diabetic foot. *Q J Nucl Med.* 1999;43:9-20.
7. Karchevsky M, Schweitzer ME, Morrison WB, Parellada JA. MRI findings of septic arthritis and associated osteomyelitis in adults. *AJR Am J Roentgenol.* 2004;182:119-122.
8. Eckman MH, Greenfield S, Mackey WC, et al. Foot infections in diabetic patients: decision and cost-effectiveness analyses. *JAMA.* 1995;273:712-720.
9. Matowe L, Gilbert FJ. How to synthesize evidence for imaging guidelines. *Clin Radiol.* 2004;59:63-68.
10. Termaat MF, Raijmakers PGHM, Scholten HJ, Bakker FC, Patka P, Haarman HJTM. The accuracy of diagnostic imaging for the assessment of chronic osteomyelitis: a systematic review and meta-analysis. *J Bone Joint Surg Am.* 2005;87:2464-2471.
11. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med.* 2003;138:W1-W12.
12. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Med Decis Making.* 1993;13:313-321.
13. Moses LE, Shapiro D, Littenberg B. Combining in-

dependent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med*. 1993; 12:1293-1316.
14. Craig JG, Amin MB, Wu K, et al. Osteomyelitis of the diabetic foot: MR imaging—pathologic correlation. *Radiology*. 1997;203:849-855.
15. Croll SD, Nicholas GG, Osborne MA, Wasser TE, Jones S. Role of magnetic resonance imaging in the diagnosis of osteomyelitis in diabetic foot infections. *J Vasc Surg*. 1996;24:266-270.
16. Enderle MD, Coerper S, Schweizer HP, et al. Correlation of imaging techniques to histopathology in patients with diabetic foot syndrome and clinical suspicion of chronic osteomyelitis: the role of high-resolution ultrasound. *Diabetes Care*. 1999; 22:294-299.
17. Ertugrul MB, Baktiroglu S, Salman S, et al. The diagnosis of osteomyelitis of the foot in diabetes: microbiological examination vs. magnetic resonance imaging and labelled leucocyte scanning. *Diabet Med*. 2006;23:649-653.
18. Horowitz JD, Durham JR, Nease DB, Lukens ML, Wright JG, Smead WL. Prospective evaluation of magnetic resonance imaging in the management of acute diabetic foot infections. *Ann Vasc Surg*. 1993;7:44-50.
19. Kearney T, Pointin K, Cunningham D, Gedroyc W, Robinson S, Elkeles RS. The detection of pedal osteomyelitis in diabetic patients. *Pract Diabetes Int*. 1999;16:98-100.
20. Ledermann HP, Schweitzer ME, Morrison WB. Nonenhancing tissue on MR imaging of pedal infection: characterization of necrotic tissue and associated limitations for diagnosis of osteomyelitis and abscess. *AJR Am J Roentgenol*. 2002; 178:215-222.
21. Levine SE, Neagle CE, Esterhai JL, Wright DG, Dalinka MK. Magnetic resonance imaging for the diagnosis of osteomyelitis in the diabetic patient with a foot ulcer. *Foot Ankle Int*. 1994;15:151-156.
22. Lipman BT, Collier BD, Carrera GF, et al. Detection of osteomyelitis in the neuropathic foot: nuclear medicine, MRI and conventional radiography. *Clin Nucl Med*. 1998;23:77-82.
23. Maas M, Slim EJ, Heoksma AF, et al. MR imaging of neuropathic feet in leprosy patients with suspected osteomyelitis. *Int J Lepr Other Mycobact Dis*. 2002;70:97-103.
24. Morrison WB, Schweitzer ME, Batte WG, Radack DP, Russel KM. Osteomyelitis of the foot: relative importance of primary and secondary MR imaging signs. *Radiology*. 1998;207:625-632.
25. Nigro ND, Bartynski WS, Grossman SJ, Kruljac S. Clinical impact of magnetic resonance imaging in foot osteomyelitis. *J Am Podiatr Med Assoc*. 1992; 82:603-615.
26. Remedios D, Valabhji J, Oelbaum R, Sharp P, Mitchell R. 99mTc-nanocolloid scintigraphy for assessing osteomyelitis in diabetic neuropathic feet. *Clin Radiol*. 1998;53:120-125.
27. Seabold JE, Flickinger FW, Kao SC, et al. Indium-111-leukocyte/technetium-99m-MDP bone and magnetic resonance imaging: difficulty of diagnosing osteomyelitis in patients with neuropathic osteoarthropathy. *J Nucl Med*. 1990;31: 549-556.
28. Vesco L, Boulahdour H, Hamissa S, et al. The value of combined radionuclide and magnetic resonance imaging in the diagnosis and conservative management of minimal or localized osteomyelitis of the foot in diabetic patients. *Metabolism*. 1999;48:922-927.
29. Weinstein D, Wang A, Chambers R, Stewart CA, Motz HA. Evaluation of magnetic resonance imaging in the diagnosis of osteomyelitis in diabetic foot infections. *Foot Ankle*. 1993;14:18-22.
30. Yuh WT, Corson JD, Baraniewski HM, et al. Osteomyelitis of the foot in diabetic patients: evaluation with plain film, 99mTc-MDP bone scintigraphy, and MR imaging. *AJR Am J Roentgenol*. 1989;152:795-800.
31. Ahmadi ME, Morrison WB, Carrino JA, Schweitzer ME, Raikin SM, Ledermann HP. Neuropathic arthropathy of the foot with and without superimposed osteomyelitis: MR imaging characteristics. *Radiology*. 2006;238:622-631.
32. Horden LD. UpToDate. http://www.uptodate.com/. Accessed November 9, 2005.
33. Mushlin AI, Littenberg B. Diagnosing pedal osteomyelitis: testing choices and their consequences. *J Gen Intern Med*. 1994;9:1-7.
34. Newman LG, Waller J, Palestro CJ, et al. Unsuspected osteomyelitis in diabetic foot ulcers: diagnosis and monitoring by leukocyte scanning with indium In 111 oxyquinoline. *JAMA*. 1991;266: 1246-1251.
35. Lipsky BA, Berendt AR, Deery G, et al; Infectious Diseases Society of America. Diagnosis and treatment of diabetic foot infections. *Clin Infect Dis*. 2004;39:885-910. http://www.journals.uchicago.edu/CID/journal/issues/v39n7/34365/34365.html. Accessed August 22, 2006.
36. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol Assess*. 2005;9(12): 1-113, iii.
37. Centers for Medicare & Medicaid. Hospital Outpatient PPS Addendum B. July 2006. http://www.cms.hhs.gov/HospitalOutpatientPPS/AU/list.asp#TopOfPage. Accessed August 24, 2006.
38. Centers for Medicare & Medicaid. Physician fee schedule relative value files: RVU06A. July 2006. http://www.cms.hhs.gov/PhysicianFeeSched/PFSRVF/list.asp#TopOfPage. Accessed August 24, 2006.

Depression Is a Risk Factor for Rehospitalization in Medical Inpatients

Anand Kartha, M.D., M.Sc.; David Anthony, M.D., M.Sc.; Christopher S. Manasseh, M.D.; Jeffrey L. Greenwald, M.D.; Veerapa K. Chetty, Ph.D.; James F. Burgess, Jr., Ph.D.; Larry Culpepper, M.D., M.P.H.; and Brian W. Jack, M.D.

Background: Rehospitalization occurs in approximately 20% of medical inpatients within 90 days of discharge. Rehospitalization accounts for considerable morbidity, mortality, and costs. Identification of risk factors could lead to interventions to reduce rehospitalization. The objective of the study was to determine if physical and mental health, substance abuse, and social support are risk factors for rehospitalization.

Method: This was a prospective cohort study in an inner-city population conducted from September 2002 to September 2004. Participants included 144 adult inpatients with at least 1 hospital admission in the past 6 months. Measurements included age, length of stay, number of admissions in the past year, and medical comorbidity as well as measures of depression, alcohol and drug abuse, social support, and health-related quality of life. The outcome studied was the rehospitalization status of participants within 90 days of the index hospitalization.

Results: The mean age of the subjects was 54.8 years; 48% were black and 78% spoke English as a primary language. Subjects were admitted a mean of 2.5 times in the year before the index admission. Sixty-four patients (44%) were subsequently rehospitalized within 90 days after the index admission. In bivariate analysis, rehospitalized patients had more prior admissions (median of 3.0 vs. 2.0 admissions, $p = .002$), greater medical comorbidity (mean Charlson Comorbidity Index score of 2.6 vs. 2.0, $p = .04$), and poorer physical functional status (mean SF-12 physical component score of 31.5 vs. 36.2, $p = .03$). A logistic regression model, including prior admissions in the last year, comorbidity, physical functional status, and depression, showed that depression tripled the odds of rehospitalization (odds ratio = 3.3, 95% CI = 1.2 to 9.3). This model had fair accuracy in identifying patients at greatest risk for rehospitalization (c statistic = 0.72).

Conclusions: Hospitalized patients with a history of prior hospitalization within 6 months who screen positive for depression are 3 times more likely to be rehospitalized within 90 days in this relatively high-risk population. Screening during hospitalization for depressive symptoms may identify those at risk for rehospitalization.

(Prim Care Companion J Clin Psychiatry 2007;9:256–262)

Received Oct. 6, 2006; accepted March 9, 2007. *From the Section of General Internal Medicine, Boston University School of Medicine, Boston, Mass. (Drs. Kartha and Greenwald); the Department of Family Medicine, Brown University School of Medicine, Providence, R.I. (Dr. Anthony); the Department of Health Services, Boston University School of Public Health, Boston, Mass. (Dr. Burgess); and the Department of Family Medicine, Boston University School of Medicine, Boston, Mass. (Drs. Manasseh, Chetty, Culpepper, and Jack).*

This project was supported by grants numbered 1UC1HS014289-01 and 1U18HS015905-01 from the Agency for Healthcare Research and Quality (Dr. Jack) and grant numbered T32-HP-10028-06 from the U.S. Department of Health and Human Services (Drs. Kartha and Anthony).

Dr. Culpepper has served as a consultant to Wyeth, Pfizer, Forest, and Neurogen and has served on the speakers or advisory boards of Wyeth, Forest, and Pfizer. Drs. Kartha, Anthony, Manasseh, Greenwald, Chetty, Burgess, and Jack report no other financial affiliations relevant to the subject of this article.

Corresponding author and reprints: Brian W. Jack, M.D., Department of Family Medicine, Dowling 5, Boston Medical Center, 1 BMC Place, Boston, MA 02118 (e-mail: brian.jack@bmc.org).

Rehospitalization accounts for considerable morbidity, mortality, and costs and may also be a marker of poor quality health care. Studies of patients discharged from adult inpatient services find 90-day rehospitalization rates between 19% and 23%.[1] Rehospitalization can place patients at risk of nosocomial infection, iatrogenic illness, and medical errors.[2] By one estimate, rehospitalizations account for as much as half of all hospitalizations.[3] In 2003, national health care expenditures in the United States on all hospitalizations were $516 billion,[4] so rehospitalization might account for over $250 billion of U.S. national health care expenditures annually. The Agency for Healthcare Research and Quality found that in 1999, 4 states spent $1.9 billion in 6 months on rehospitalization.[5] Therefore, even a small reduction in rehospitalization could have an important impact on health care services provision and quality.

Some successes in preventing rehospitalization in specific disorders such as congestive heart failure have been demonstrated, but strategies to prevent rehospitalization have had limited success in adult inpatients with other diagnoses.[6] For an intervention aimed at reducing rehospitalization to be practical and cost-effective, there needs to be a reliable means of targeting those patients who are at high risk. Known risk factors for rehospitalization include advanced age, specific diagnoses (such as congestive heart failure), history of prior admission, length of

stay, and some measure of the severity of the illness.[7,8] However, these factors that are readily available from administrative databases explain only a small amount of the variance in rehospitalization. Predictors of rehospitalization in urban, underserved populations are even less well defined. Efforts to reduce rehospitalization in these populations deserve particular attention because there is evidence that people of lower socioeconomic status experience higher rehospitalization rates.[9]

We designed a study to assess if measurable behavioral and functional factors help to explain some of the variability in predicting rehospitalization. We hypothesized that depressive symptoms, alcohol and drug abuse, lack of social support, and poor health-related quality of life contribute to an increased risk of rehospitalization in a recurrently admitted, inner-city population.

METHOD

Study Setting

This study was conducted on an adult inpatient service that admits from a consortium of 15 community health centers (CHCs) comprising the Boston HealthNet to Boston Medical Center (BMC), an inner-city, academic medical center located in Boston, Mass. The Boston HealthNet inpatient service is staffed by family physicians and general internists from the CHCs and family physicians from the Department of Family Medicine at BMC. The study was conducted from September 2002 to September 2004. The institutional review board of Boston University Medical Center approved the study.

Study Design

This was a prospective cohort study. The index admission was the hospitalization during which the patient was enrolled in the study. Trained staff members screened inpatients for eligibility. Patients were eligible if they (1) were admitted to the Boston HealthNet medical inpatient service, (2) were over 18 years of age, (3) had a history of at least 1 hospitalization in the 6-month period before the index admission, (4) had an assigned primary care physician (PCP) at 1 of the CHCs, (5) were willing to be rehospitalized if indicated, and (6) were admitted from a noninstitutionalized setting. Patients were excluded if (1) they expressed that they were unwilling or unable to keep appointments with their PCP, (2) rehospitalization was planned (e.g., for diagnostic test or procedure) at the time of enrollment, or (3) they were previously enrolled. Subjects were approached for study enrollment at each admission even if they declined participation at a previous admission.

Data Collection

After obtaining informed consent, research staff members collected demographic data and information about known risk factors for rehospitalization, including age, length of stay of the index admission, and number of admissions in the year before the index admission, and calculated the Charlson Comorbidity Index score[10] from review of the hospital record. Research staff then conducted a face-to-face structured interview with the subject that included administering the following validated instruments.

Nine-item Patient Health Questionnaire (PHQ-9). The PHQ-9[11] uses a 4-point Likert scale and a standard scoring algorithm to diagnose major depression; the screen has been used in numerous studies. The PHQ-9 was developed as a 1-stage, self-report version of the depression section of the previously validated Primary Care Evaluation of Mental Disorders instrument.[12] The PHQ-9 demonstrates a sensitivity of 88% and a specificity of 88% for major depression (as compared with mental health professional's diagnoses as the gold standard).[11]

Alcohol Use Disorders Identification Test (AUDIT). The AUDIT,[13] which may be self-administered or administered by an interviewer, consists of 10 questions and takes approximately 3 minutes to complete. The first 8 questions refer to alcohol-related issues over the prior year, and the final 2 include information about previous years. The AUDIT has been extensively researched and validated, demonstrating high correlation with other frequently used instruments such as the Michigan Alcoholism Screening Test.[14] Using the traditional cutoff of 8, the test has sensitivity for alcohol abuse or dependence (as defined by DSM-III-R criteria), ranging from 78% to 100%. Specificity values have been lower, ranging from 25% to 96%. Of note, the AUDIT has been validated among adults at an inner-city medical clinic serving a population similar to that of BMC. In that population, the instrument was shown to be 96% sensitive for current alcohol abuse or dependence and to accurately distinguish between current and former alcohol-related problems.[14] The AUDIT has been used in numerous research projects, including a recent study of a substance abuse treatment system[14] accessed through the emergency department at BMC. Given extensive use in populations such as ours and ease of administration, we chose the AUDIT over other screening instruments.[13-15]

10-item Drug Abuse Screening Test (DAST-10). The DAST-10[16] identifies drug use problems using a 10-item screen, with scores greater than 3 suggesting a moderate level of drug-related problems worthy of further investigation. The original DAST 28-item instrument demonstrated excellent content and construct validity and internal consistency ranging from .86 to .95, with an 85% overall accuracy in diagnosing drug abuse or dependence (according to DSM-III criteria), as compared with the Diagnostic Interview Schedule. Both 20-item and 10-item versions of the DAST have been developed. The DAST-20 correlates almost perfectly with the original DAST, and the DAST-10 correlates highly with the 20-item version

(r = .98).[17] The DAST-10, due to its similarity in structure, has been frequently used with the AUDIT to screen for drug and alcohol–related problems.[16]

Norbeck Social Support Questionnaire (NSSQ). The NSSQ[18] is used to measure multiple dimensions of social support, with the mean (SD) score for the normative population equal to 201.9 (95.9) and higher scores suggesting greater social support. The NSSQ has good internal reliability (correlations ranging from .69 to .98) and test-retest reliability (correlations ranging from .86 to .92).[18] The test yields 3 scores (total functional support, total network support, and total loss). The total functional support score was used in this study. The NSSQ has been used in studies of low-income populations and in numerous studies to identify correlations between social support and rehospitalization among patients with ischemic heart disease.[18,19]

Short Form-12 Health Survey (SF-12). The SF-12[20] was developed as a valid abbreviation of the widely used Short Form-36 questionnaire that was designed to accurately measure the physical and mental components of health-related quality of life. Administration of the survey yields 2 scores: the physical component summary (PCS) and the mental component summary (MCS). These scores are scaled on a range of 0 to 100 designed to produce mean scores of 50 and standard deviations of 10 in a representative sample of the U.S. population; higher scores suggest greater functional status. The 2 summary scores offer excellent reliability as measured by the internal consistency method (PCS = 89% and MCS = 86%).[20]

Independent Variables

Age, length of stay, number of prior admissions, Charlson Comorbidity Index, NSSQ functional scale, and SF-12 PCS and MCS were characterized as continuous variables. The PHQ-9 (major depression), AUDIT (score ≥ 6), and DAST-10 (score ≥ 3) were dichotomized using standardized scoring systems.

Data

Dependent variables. The outcome studied was the rehospitalization status within 90 days of the index hospitalization. Rehospitalization refers to the first hospitalization that occurred for any reason within 90 days after discharge from the index admission. Rehospitalization was determined by review of the hospital database that is available for all patients. To determine if subjects were rehospitalized at another hospital, we contacted all subjects by telephone. If we were unable to contact subjects after 5 telephone calls, then we contacted the alternate contacts, whose information was obtained during study enrollment.

Statistical analyses. Analysis included (1) comparison of subjects who agreed to enroll in the study versus those who refused participation, (2) bivariate comparison of baseline characteristics of rehospitalized subjects versus those not rehospitalized, and (3) bivariate comparison of risk factors for rehospitalization comparing those subjects rehospitalized and not rehospitalized. In these analyses, t test, χ^2 test, Fisher exact test, and Wilcoxon rank sum test were used where appropriate, with a p value ≤ .05 used as level of significance.

The final analysis included a series of multivariate logistic regression analyses to determine the odds of rehospitalization after controlling for all independent variables. We used a 3-stage process to develop our logistic models. Model 1 includes risk factors known to be associated with rehospitalization (age, length of index hospital stay, number of hospital admissions in the past year, and Charlson Comorbidity Index score). Model 2 includes all of the psychosocial and functional scales that were hypothesized to contribute to rehospitalization, which included the PHQ-9 (major depression), PHQ-9 (anxiety), AUDIT (alcohol abuse), DAST-10 (drug abuse), NSSQ (social support), and SF-12 PCS and MCS (physical and mental components of health-related quality of life). Finally, model 3 incorporates independent variables found to be related to rehospitalization at the p = .05 level of significance in either the bivariate analysis or in model 1 or model 2. This model includes the number of hospital admissions in the prior year and Charlson Comorbidity Index, SF-12 PCS, and PHQ-9 (major depression) scores. Model 3 is the preferred model, as it provides the most conservative assessment of risk factor performance.

Model performance was assessed by the concordance statistic (c statistic) for discriminative ability (refers to the ability to distinguish high-risk subjects from low-risk subjects). A value of 0.5 indicates no discriminatory power, whereas a c statistic of 1.0 indicates perfect discrimination. Model calibration (refers to whether the predicted probabilities agree with the observed probabilities) was assessed by the Hosmer and Lemeshow goodness-of-fit test. Lack of fit in our models was considered statistically significant if the p value was less than .05. All analyses were done using SAS version 8.02 software (SAS Institute, Inc., Cary, N.C.).

RESULTS

Sample Characteristics

Of 204 eligible subjects who were approached and asked to participate in the study, 144 (71%) were enrolled and 60 (29%) refused. The age, gender, length of index hospital stay, and number of prior admissions in the past year did not differ among enrolled subjects when compared with those who chose not to participate (Table 1). Eligible subjects who previously refused were reapproached at each admission. Enrolled subjects were admitted a mean of 2.5 times in the year before the index admission.

Table 2 shows baseline characteristics of enrolled subjects by rehospitalization status within 90 days of the

Table 1. Characteristics of Eligible Subjects Enrolled in the Study Compared With Those Who Refused Participation (N = 204)

Characteristic	Enrolled (N = 144)	Not Enrolled (N = 60)	p Value[a]
Age, mean ± SD, y	54.8 ± 15.8	54.6 ± 17.5	.95
Gender (male), N (%)	64 (44)	30 (50)	.47
Length of index hospital stay, median (range), d	3 (0–65)	2 (0–44)	.44
No. of hospital admissions in past year, median (range)	2 (0–11)	1 (1–10)	.36

[a]Statistics were calculated as follows: t test for age, χ^2 test for gender, and Wilcoxon rank sum test for length of stay and hospital admissions.

Table 2. Baseline Characteristics of Subjects Grouped by Rehospitalization Status Within 90 Days of the Index Hospitalization (N = 144)

Characteristic	Rehospitalized (N = 64)	Not Rehospitalized (N = 80)	p Value[a]
Age, mean ± SD, y	55.0 ± 16.3	54.7 ± 15.4	.93
Gender (male), N (%)	29 (45)	36 (45)	.97
Race, N (%)			.17
Black	27 (42)	43 (54)	
White	29 (45)	24 (30)	
Other	8 (13)	13 (16)	
Marital status, N (%)			.10
Married	19 (30)	11 (14)	
Divorced	8 (13)	17 (21)	
Single	24 (38)	33 (41)	
Other	13 (20)	19 (24)	
Primary language, N (%)			.61
English	52 (81)	60 (75)	
Spanish	7 (11)	11 (14)	
Other	4 (6)	8 (10)	
Admission diagnosis, N (%)[b]			.64
Metabolic[c]	9 (15)	2 (3)	
Arrhythmia	6 (10)	3 (4)	
Asthma	5 (8)	13 (19)	
Chest pain	5 (8)	6 (9)	
Congestive heart failure	4 (7)	5 (7)	
Alcohol	4 (7)	4 (6)	
Pancreatitis	3 (5)	0 (0)	
Cellulitis	2 (3)	6 (9)	
Other	23 (38)	30 (44)	
Insurance, N (%)[b]			.28
Commercial	8 (14)	7 (10)	
Medicare	12 (20)	16 (23)	
Medicaid	11 (18)	15 (21)	
Other	29 (48)	32 (46)	

[a]Statistics were calculated as follows: t test for age, χ^2 test for gender and marital status, and Fisher exact test for primary language, admission diagnosis, and insurance.
[b]Information available for 130 subjects.
[c]Metabolic diagnoses include diabetes (1), hypoglycemia (1), hyperglycemia (1), ketoacidosis (3), renal failure (1), hyponatremia (2), and dehydration (2).

index hospitalization. For all 144 enrolled subjects, the mean age was 54.8 years, 49% were black, 45% were male, and 78% spoke English as a primary language. Nearly all subjects were insured through commercial insurance, Medicare, or Medicaid. The remainder were either self-pay or covered by the Massachusetts system that provides care to those who otherwise lack coverage.

Of 144 subjects enrolled, 64 (44%) had at least 1 rehospitalization in the 90 days after discharge from the index admission (median = 1.0, mean = 1.8, range, 1–6). While there were no statistically significant differences, rehospitalized subjects tended to be white, married, and English speaking compared with those not rehospitalized. Nearly half of rehospitalized subjects were originally hospitalized for metabolic diseases (hypoglycemia, hyperglycemia, ketoacidosis, renal failure, hyponatremia, and dehydration), arrhythmias, asthma, chest pain, and congestive heart failure. In subjects who were not rehospitalized, the most common diagnoses included asthma, chest pain, cellulitis, and congestive heart failure.

The group of patients rehospitalized within 90 days was admitted frequently (median of 2 admissions in the past 6 months) and had a high degree of medical and psychosocial risk: 27% met criteria for major depression on the PHQ-9, 18% were at risk of hazardous drinking (AUDIT score ≥ 6), and 6% were at risk of drug-related problems (DAST-10 score ≥ 3). The rehospitalized sample had poor social support (median score of 71 on the NSSQ, while a score of 201.9 was average for a normative population) and reported poor physical health (mean score of 34.2 on the SF-12 PCS, while the mean score of the U.S. population is 50) and poor mental health (mean score of 41.8 on the SF-12 MCS, while the mean score of the U.S. population is 50).

Bivariate Analyses

Table 3 shows results of unadjusted bivariate analyses of risk factors for rehospitalization grouped by rehospitalization status. Rehospitalized subjects had a greater number of admissions in the prior year (median of 3.0 vs. 2.0, p = .002), greater medical comorbidity (mean Charlson Comorbidity Index score of 2.6 vs. 2.0, p = .04), and poorer health-related quality of life as measured by the SF-12 PCS (mean score of 31.5 vs. 36.2, p = .03). Rehospitalized subjects also had a greater length of the index admission (mean of 5.1 vs. 4.4 days), were more at risk of hazardous drinking (24.6% vs. 16.0%), or met criteria for major depression (35.0% vs. 21.1%), although these variables were not statistically significant.

Predictive Models

Table 4 shows the adjusted odds ratios from the logistic regression models of risk factors for rehospitalization. In model 1, only the number of hospital admissions in the prior year was significantly related to rehospitalization (OR = 1.27, 95% CI = 1.05 to 1.54). In model 2, major depression was a significant risk factor (OR = 4.6, 95% CI = 1.06 to 20.3). Model 3 included the 4 variables found to be significant in the bivariate analysis or in models 1 or 2. These variables were (1) number of hospital admissions

Table 3. Risk Factors for Rehospitalization Grouped by Rehospitalization Status

Risk Factor[a]	Rehospitalized (N = 64)	Not Rehospitalized (N = 80)	p Value[b]
Length of index hospital stay, median (range), d	3 (0–65)	3 (0–29)	.88
No. of hospital admissions in past year, median (range)	3 (1–11)	2 (1–11)	.002*
Charlson Comorbidity Index score, mean ± SD (median, range)[c]	2.6 ± 2.0 (2, 0–8)	2.0 ± 1.9 (2, 0–8)	.04*
PHQ-9 score, % with major depression	35	21	.07
AUDIT score (range, 0–10), % at high risk for alcohol abuse (score ≥ 6)	25	16	.21
DAST-10 score (range, 0–10), % at high risk for drug abuse (score ≥ 3)	7	5	.70
NSSQ functional support score, median (range)	75 (0–241)	65 (0–273)	.78
SF-12 physical component summary score (range, 0–100), mean ± SD	31.5 ± 10.0	36.2 ± 10.4	.03*
SF-12 mental component summary score (range, 0–100), mean ± SD	40.4 ± 13.3	42.9 ± 12.6	.36

[a]Data may not be available for all 144 subjects for each analysis.
[b]Statistics were conducted as follows: t test for age and SF-12 scores; χ^2 test for PHQ-9, AUDIT, and DAST-10 scores; and Wilcoxon rank sum test for length of stay, hospital admissions, and Charlson Comorbidity Index and NSSQ scores.
[c]Reflects the cumulative increased likelihood of 1-year mortality; the higher the score, the more severe the burden of comorbidity; a 35% increase in risk of dying is reflected in a 1-point increase in weights.
*Statistically significant, p ≤ .05.
Abbreviations: AUDIT = Alcohol Use Disorders Identification Test, DAST-10 = 10-item Drug Abuse Screening Test, NSSQ = Norbeck Social Support Questionnaire, PHQ-9 = 9-item Patient Health Questionnaire, SF-12 = Short Form-12 Health Survey.

Table 4. Adjusted Odds Ratios of Risk Factors for Rehospitalization

		Statistic	
Risk Factor[a]	Adjusted Odds Ratio (95% CI)	p	c
Model 1 (general risk factors) (N = 143)			0.66
Age	1.00 (0.98 to 1.02)	.97	
Length of index hospital stay	1.01 (0.96 to 1.06)	.70	
No. of hospital admissions in prior year	1.27 (1.05 to 1.54)	.02*	
Charlson Comorbidity Index score	1.19 (0.99 to 1.43)	.06	
Model 2 (psychosocial and functional status risk factors)			0.70
PHQ-9 score (major depression)	4.64 (1.06 to 20.3)	.04*	
AUDIT score (alcohol abuse)	1.80 (0.50 to 6.38)	.37	
DAST-10 score (drug abuse)	0.95 (0.10 to 9.14)	.97	
NSSQ score (social support)	1.00 (0.99 to 1.01	.67	
SF-12 score (physical functional status)	0.97 (0.92 to 1.02)	.27	
SF-12 score (mental functional status)	1.02 (0.97 to 1.08)	.39	
Model 3[b]			0.72
No. of hospital admissions in prior year	1.21 (0.97 to 1.50)	.08	
Charlson Comorbidity Index score	1.16 (0.91 to 1.47)	.22	
SF-12 score (physical functional status)	0.97 (0.93 to 1.02)	.26	
PHQ-9 score (major depression)	3.34 (1.20 to 9.25)	.02*	

[a]Data may not be available for all 144 subjects for individual analysis.
[b]Includes risk factors found to be significant at the p = .05 level in preceding bivariate and logistic regression analyses.
*Statistically significant, p ≤ .05.
Abbreviations: AUDIT = Alcohol Use Disorders Identification Test, DAST-10 = 10-item Drug Abuse Screening Test, NSSQ = Norbeck Social Support Questionnaire, PHQ-9 = 9-item Patient Health Questionnaire, SF-12 = Short Form-12 Health Survey.

in the prior year, (2) Charlson Comorbidity Index (medical comorbidity), (3) SF-12 PCS (health-related quality of life, physical component summary), and (4) PHQ-9. The odds of rehospitalization were again significantly greater in subjects who screened positive for major depression (OR = 3.3, 95% CI = 1.20 to 9.25). For models 1, 2, and 3, the c statistics were 0.66, 0.70, and 0.72, respectively, and the p values for the Hosmer and Lemeshow goodness-of-fit test were 0.58, 0.15, and 0.35, respectively.

DISCUSSION

In patients who had a history of hospitalizations in the 6 months prior to the index hospitalization, we found that a positive screen for major depression at admission tripled the odds of being rehospitalized within 90 days of discharge among adult inpatients in an urban academic medical center. While we await further study in randomized intervention trials, clinicians should consider screening for and treating depression in their frequently readmitted inpatients in an effort to reduce medical rehospitalization.

Our findings are supported by the few studies that have examined the association between depression and rehospitalization. A study of medical and surgical inpatients found that depressed patients spent twice as many days rehospitalized over a 4-year period.[21] A Swiss study in elderly inpatients found that patients with depressive symptoms had higher inpatient service utilization over a

6-month period.[22] Similarly, a case-control study of elderly patients in a Medicare managed care plan found increased odds of unplanned rehospitalization within 30 days if patients had a history of depression (OR = 3.2, 95% CI = 1.4 to 7.9).[23] Taken together, our study adds to the literature indicating that depression is a powerful risk factor for rehospitalization. Depression is also associated with worse outcomes after myocardial infarction.[24] In a prospective registry of myocardial infarction, depressive symptoms were measured in 1873 patients with the PHQ during hospitalization and 1 month after discharge. In this study, all categories of depression were associated with higher rehospitalization, more frequent angina, more physical limitations, and worse quality of life.[24]

Depression might lead to rehospitalization through a variety of mechanisms. The neuroendocrine changes seen in depression can worsen physical illness. Depression may impair health-related quality of life leading to lower thresholds for admission. Somatization may be misinterpreted, thereby complicating diagnosis and management.[19] Medication nonadherence among patients with mental health diagnoses could lead to rehospitalization.[25] Finally, depressed patients might have a reduced social support network leading to increasing stress and worsening of symptoms, thereby lengthening time to recovery and necessitating rehospitalization.[26]

An array of clinical and administrative factors have been associated with rehospitalization including age, length of stay, number of prior hospital admissions, comorbid medical illness, admitting diagnosis, male sex, white race, Medicaid coverage, single marital status, and laboratory data such as glycosylated hemoglobin.[27] However, these factors tend to explain very little of the variation in rehospitalizations.[6] Therefore, in this study, we explored the contributions of mental health, substance abuse, social support, and perceived health-related quality of life, which we proposed to be related to rehospitalization and which have been suggested in the medical literature.[28–30] Although much of the variability in rehospitalization remains unexplained, these findings support our approach to combine psychosocial and functional determinants of rehospitalization to better understand these complex interactions.

We were successful in identifying a group at high risk of rehospitalization (overall 44% were rehospitalized within 90 days of the index admission). Three factors might explain this finding: (1) the population enrolled had a high degree of medical and psychosocial comorbidities (27% screened positive for major depression, 18% for alcohol abuse, and 6% for risk of drug abuse; many had poor social support and physical health); (2) we only enrolled subjects who had been hospitalized in the last 6 months, a powerful risk factor for rehospitalization[3,6]; and (3) we reapproached subjects who had declined enrollment at prior admissions at each hospitalization, resulting in a study group with a mean of 2.5 hospitalizations in the year before the index hospitalization. As a result of these 3 factors, we studied subjects at very high risk of rehospitalization. While our inclusion criteria might somewhat limit the generalizability of our results, it is relevant to many hospitals that provide care to similar inner-city populations who are frequently hospitalized.

Finally, while our data suggest that more prior admissions, greater length of stay, poorer physical functioning, anxiety disorders, and alcohol use are related to rehospitalizations, we could not confirm our hypotheses regarding these factors in multivariate analysis. Interestingly, we did not find a relationship between social support and rehospitalization. It is possible that our subjects used their supports to facilitate health-seeking behavior leading to rehospitalization. Also, we used the functional support score of the NSSQ in our analyses. If the effect of social support is primarily from aspects of social support we did not analyze (network properties and recent losses), it is possible that we may not have fully accounted for the variance in rehospitalization explained by social support.

A major strength of this study is its prospective design that permits us to describe both the magnitude and direction of causal relationships between factors. This is in favorable contrast to the retrospective designs that are used by a number of prior studies. Limitations of this study include the modest enrollment rate of eligible subjects; however, the fact that age, gender, length of stay, and number of prior admissions did not differ between groups may suggest that subjects were not entirely dissimilar. Finally, we did not consider the quality of care received during the index admission, which has been associated with rehospitalization.[31]

Despite these limitations, we conclude that hospitalized adults with a history of recent previous hospitalizations who screen positive for major depression are at a 3 times greater risk of being rehospitalized within 90 days of discharge. Future randomized controlled trials designed to treat patients screening positive for major depression in an attempt to reduce rehospitalization should be undertaken. While we await further study in randomized intervention trials, clinicians should consider screening for and treating depression in their frequently readmitted inpatients before discharge in an effort to reduce rehospitalization.

REFERENCES

1. Smith DM, Giobbie-Hurder A, Weinberger M, et al. Predicting nonelective hospital readmissions: a multi-site study. Department of Veterans Affairs Cooperative Study Group on Primary Care and Readmissions. J Clin Epidemiol 2000;53:1113–1118
2. Kohn LT, Corrigan JM, Donaldson MS, et al. To Err is Human: Building a Safer Health System. Washington, DC: National Academy Press; 2000
3. Hasan M. Readmission of patients to hospital: still ill defined and poorly understood. Int J Qual Health Care 2001;13:177–179
4. Centers for Medicare & Medicaid Services. National Health

Expenditures Aggregate Amounts and Average Annual Percent Change, by Type of Expenditure: Selected Calendar Years 1980–2003: Office of the Actuary, National Health Statistics Group; 2005
5. Friedman B, Basu J. The rate and cost of hospital readmissions for preventable conditions. Med Care Res Rev 2004;61:225–240
6. Soeken KL, Prescott PA, Herron DG, et al. Predictors of hospital readmission: a meta-analysis. Eval Health Prof 1991;14:262–281
7. Phillips RS, Safran C, Cleary PD, et al. Predicting emergency readmissions for patients discharged from the medical service of a teaching hospital. J Gen Intern Med 1987;2:400–405
8. Librero J, Peiro S, Ordinana R. Chronic comorbidity and outcomes of hospital care: length of stay, mortality, and readmission at 30 and 365 days. J Clin Epidemiol 1999;52:171–179
9. Weissman JS, Stern RS, Epstein AM. The impact of patient socioeconomic status and other social factors on readmission: a prospective study in four Massachusetts hospitals. Inquiry 1994;31:163–172
10. Charlson ME PP, Ales KL, McKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chron Dis 1987;40:373–383
11. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Int Med 2001;16:606–613
12. Spitzer RL, Kroenke K, Williams JB, et al. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999;282:1737–1744
13. Allen JP, Litten RZ, Fertig JB, et al. A review of research on the Alcohol Use Disorders Identification Test (AUDIT). Alcohol Clin Exp Res 1997; 21:613–619
14. Bernstein E, Bernstein J, Levenson S. Project ASSERT: an ED-based intervention to increase access to primary care, preventive services and the substance abuse treatment system. Ann Emerg Med 1997;30:181–189
15. Fiellin DA, Reid MC, O'Connor PG. Outpatient management of patients with alcohol problems. Ann Int Med 2000;133:815–827
16. Gavin DR, Ross HE, Skinner HA. Diagnostic validity of the drug abuse screening test in the assessment of DSM-III drug disorders. Br J Addict 1989;84:301–307
17. Maisto SA, Carey KB, Gordon CM, et al. Use of the AUDIT and the DAST-10 to identify alcohol and drug use disorders among adults with a severe and persistent mental illness. Psychol Assess 2000;12:186–192
18. Norbeck JS, Lindsey AM, Carrieri VL. Further development of the Norbeck Social Support Questionnaire: normative data and validity testing. Nurs Res 1983;32:4–9
19. Gigliotti E. A confirmation of the factor structure of the Norbeck Social Support Questionnaire. Nurs Res 2002;51:276–284
20. Ware JE, Kosinski M, Keller SD. A 12-item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220–233
21. Saravay SM, Pollack S, Steinberg MD, et al. Four-year follow-up of the influence of psychological comorbidity on medical rehospitalization. Am J Psychiatry 1996;153:397–403
22. Bula CJ, Wietlisbach V, Burnand B, et al. Depressive symptoms as a predictor of 6-month outcomes and services utilization in elderly medical inpatients. Arch Intern Med 2001;161:2609–2615
23. Marcantonio ER, McKean S, Goldfinger M, et al. Factors associated with unplanned hospital readmission among patients 65 years of age and older in a Medicare managed care plan. Am J Med 1999;107:13–17
24. Parashar S, Rumsfeld JS, Spertus JA, et al. Time courses of depression and outcomes of myocardial infarction. Arch Intern Med 2006;166: 2035–2043
25. Haywood TW, Kravitz HM, Grossman LS, et al. Predicting the "revolving door" phenomenon among patients with schizophrenic, schizoaffective, and affective disorders. Am J Psychiatry 1995;152: 856–861
26. Tse WS, Bond AJ. The impact of depression on social skills. J Nerv Ment Dis 2004;192:260–268
27. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Arch Intern Med 2000;160: 1074–1081
28. Brennan PL, Kagay CR, Geppert JJ, et al. Elderly Medicare inpatients with substance use disorders: characteristics and predictors of hospital readmissions over a four-year interval. J Stud Alcohol 2000;61:891–895
29. Pearson S, Stewart S, Rubenach S. Is health-related quality of life among older, chronically ill patients associated with unplanned readmission to hospital? Aust N Z J Med 1999;29:701–706
30. Stewart MJ, Hirth AM, Klassen G, et al. Stress, coping, and social support as psychosocial factors in readmissions for ischaemic heart disease. Int J Nurs Stud 1997;34:151–163
31. Ashton CM, Del Junco DJ, Souchek J, et al. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. Med Care 1997;35:1044–1059

Willingness of Minorities to Participate in Biomedical Studies: Confirmatory Findings from a Follow-Up Study Using the Tuskegee Legacy Project Questionnaire

Ralph V. Katz, DMD, MPH, PhD; B. Lee Green, PhD; Nancy R. Kressin, PhD; Cristina Claudio, PhD; Min Qi Wang, PhD; and Stefanie L. Russell, DDS, MPH, PhD

Financial support: The three-city research subject study was supported by NIDCR/NIH grant U54 DE 14257, the NYU Oral Cancer Research on Adolescent and Adult Health Promotion), an Oral Health Disparities Research Center.

Objectives: The purposes of this analysis were to compare the self-reported willingness of blacks, Puerto-Rican Hispanics and whites to participate as research subjects in biomedical studies, and to determine the reliability of the Tuskegee Legacy Project Questionnaire (TLP).

Methods: The TLP Questionnaire, initially used in a four-city study in 1999–2000, was administered in a follow-up study within a random-digit-dial telephone survey to a stratified random sample of adults in three different U.S. cities: Baltimore, MD; New York City; and San Juan, PR. The questionnaire, a 60-item instrument, contains two validated scales: the Likelihood of Participation (LOP) Scale and the Guinea Pig Fear Factor (GPFF) Scale.

Results: Adjusting for age, sex, education, income and city, the LOP Scale was not statistically significantly different for the racial/ethnic groups (ANCOVA, p=87). The GPFF Scale was statistically significantly higher for blacks and Hispanics as compared to whites (adjusted ANCOVA, p<0.001).

Conclusions: The of the findings from the current three-city study, as well as from our prior four-city study, are remarkably similar and reinforce the conclusion that blacks and Hispanics self-report that, despite having a higher fear of participation, they are just as likely as whites to participate in biomedical research.

Key words: research ■ minorities ■ Tuskegee Syphilis Study ■ race/ethnicity

 From the Department of Epidemiology & Health Promotion, NYU College of Dentistry, New York University, New York, NY (Katz, professor and chair; Russell, assistant professor); Office of Institutional Diversity, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL (Green, executive director); Department of General Internal Medicine, Boston University Medical School, and Center for Health Quality, Outcomes & Economic Research, Department of Veterans Affairs, Bedford VA (Kressin, associate professor); Community Dentistry Section, School of Dentistry, University of Puerto Rico, San Juan, PR (Claudio, professor); and Department of Public and Community Health, School of Public Health, University of Maryland, College Park, MD (Wang, professor). Send correspondence and reprint requests for *J Natl Med Assoc.* 2007;99:1052–1060 to: Ralph Katz, Professor and Chair, Department of Epidemiology & Health Promotion, NYU College of Dentistry, New York University; phone: (212) 998-9550; fax: (212) 995-4436; e-mail: ralph.katz@nyu.edu

INTRODUCTION

The U.S. Public Health Service (USPHS) Syphilis Study at Tuskegee (1932–1972) is arguably the most infamous biomedical research study in U.S. history.[1-5] There is widespread belief that the "legacy" of this unethical research event is that the black community has a greater reluctance to participate in clinical research studies as a result of the abuses foisted on the 399 African-American sharecroppers in Macon County, AL, who were the subjects in this 40-year USPHS study of the effects of untreated syphilis in the Negro male.[6] While a considerable amount has been written about the long-lasting effects of the USPHS Tuskegee Syphilis Study on the black community, most of this work has been from legal, historical, healthcare access or ethical perspectives.[7-20]

Between 1997–2003, the early literature on the issue of willingness of blacks to participate in biomedical studies, as compared to whites, was, understandably and typically dominated by qualitative studies that largely explored the parameters and range of issues to be studied, rather than definitively investigating the issue in depth.[21-28] Only four of those early published studies presented quantified data comparing blacks to whites on willingness to participate in biomedical research, often as related to the USPHS Tuskegee Syphilis Study.[25-28] A recent literature review article on this topic pointed out the limitations in these four early exploratory quantitative studies, e.g., all four were conducted in a single city, three only reported the findings for selected subsamples of their total number of subjects, and three only used a single question on willingness to participate as their

measure of this complex decision.[29] The one study which did report on total sample findings focused only on cancer research participation in elderly subjects and reported no difference in willingness to participate between blacks and whites.[28] Thus, the early literature provides no body of findings which could be generalized broadly. Our recently published large-scale survey on adults in four U.S. cities (Birmingham, AL; Tuskegee, AL; Hartford, CT; and San Antonio, TX), which was the first survey to use the 60-item Tuskegee Legacy Project (TLP) Questionnaire, found that blacks self-reported that they were just as likely as whites to participate in biomedical research despite having a higher fear of participation.[30]

The primary specific aim of this analysis was to compare the self-reported willingness of blacks, Puerto-Rican Hispanics and whites residing in New York City, Baltimore and San Juan to participate as research subjects in biomedical studies, as measured by the Likelihood of Participation (LOP) Scale and the Guinea Pig Fear Factor (GPFF) Scale as components of the TLP Questionnaire. The primary contrast of interest in this study is between blacks and non-Hispanic whites, with a secondary interest in clarifying if these associations generalize to Puerto-Rican Hispanics. An additional aim of this second study to use the TLP Questionnaire was to determine the reliability of both the LOP and GPFF scales of the TLP Questionnaire instrument across similar ethnic/racial groups in differing U.S. cities.

METHODS

Overview

This three-city research subject study was designed to administer the TLP Questionnaire via random-digit-dial (RDD) telephone interviews to a total of 900 subjects (300 blacks, 300 non-Hispanic whites and 300 Puerto-Rican Hispanics) aged ≥18 years in three cities: New York City, Baltimore and San Juan. The choice of

Table 1. Questions from the TLP Questionnaire on willingness to participate, and key questions that formed GPFF and LOP scales in the three-city research subject study

Q16. How likely are you to agree to become a participant in any kind of medical study at the present time?* [responses to Q16: VL SL NQS SUL VUL]

Q17. Would you feel the same no matter who was running the study? I'm going to read you a list of people who might run a study. For instance, how likely would you be to participate in a medical research study if it were run by:

a. your own doctor*
b. a university medical school/hospital*
c. the government*
d. a nonprofit foundation*
e. a tobacco company*
f. a drug company*
g. an insurance company* [responses to Q17a-g: VL SL NQS SUL VUL]

Q18. Each medical research study is different, so people who participate might have to do different things in different studies. How likely are you to participate in a medical study if you had to do the following:

a. give blood*
b. take IV injections*
c. do exercises*
d. be interviewed in person*
e. be interviewed by telephone*
f. have diet limited or restricted*
g. take medicine by mouth*
h. undergo major surgery*
i. undergo minor surgery* [responses to Ql8a-i: VL SL NQS SUL VUL]

LOP Scale comprised of Q16 + Q17a-g + Q18a-i [as marked in italicized boldface above with single asterisk]

Q19. There are lots of things that might make people NOT WANT to participate in medical research studies. How much would the following interfere with your taking part in a medical research study?

i) any fear you have of getting AIDS**
ii) any fear of being a 'guinea pig'**
iii) any fear of results not being private or confidential**
iv) any fear of having to pay for the research treatments**
v) lack of trust in research** [responses to Ql9i-v: totally a great deal some a little not at all]

GPFF Scale comprised of #19i–v [as marked in boldface above with double asterisks]

these three cities was based upon obtaining the desired sample size for the three ethnic/racial groups within the broader parameters set by the goals of the projects within the NYU Oral Cancer RAAHP (Research on Adolescent and Adult Health Promotion) Center, a U54 Oral Health Disparities Research Center funded by the National Institute of Dental Craniofacial Research (NIDCR) at the National Institutes of Health (NIH). The data collection phase was conducted in the four-month period of September to December 2003. This study was approved by the institutional review board of New York University.

The primary research instrument was the TLP Questionnaire, a 60-item instrument, which was slightly modified for this study by the elimination of a few questions that had proved redundant in prior use. The TLP Questionnaire addresses a range of issues related to the recruitment of minorities into biomedical studies. Details on the history and development of the TLP Questionnaire as well as the justifications of the methodologic decisions in the analysis of the TLP Questionnaire have been published elsewhere.[20,34] The TLP Questionnaire contains two identified conceptual domains of interest (the LOP domain and the GPFF domain) which had been validated as scales via standardized psychometric analysis techniques using data from our prior study.[30] As in the first study, these two scales are referred to as the LOP Scale and the GPFF Scale.

The Random-Digit-Dial Process

ORC Macro, a U.S.-based international opinion research corporation, conducted a RDD survey using a computer-assisted telephone interviewing (CATI) system for the data collection. The survey sample for this study was drawn from the total noninstitutionalized adult populations (ages ≥18) residing in telephone-equipped dwelling units in three target cities: New York City, Baltimore and San Juan. The study provided for a disproportionally allocated, stratified, random-digit sample of telephone-equipped residential households in the three targeted cities, which were sampled independently. The telephone survey followed a 10-attempt dialing protocol, in which up to 10 attempts were made unless a final disposition was obtained. Experienced, supervised personnel conducted the interviews using Computers for Marketing Corp.'s CATI software package.

Key Variables from the TLP Questionnaire

Table 1 consists of the key questions from the TLP Questionnaire, which formed the basis for this analysis. It shows both the precise wording of the four key questions and their subparts as well as the elements of those questions that were used to create the LOP Scale and the GPFF Scale. The LOP Scale was comprised of 17 variables contained within questions 16, 17a–g and 18a–i of the TLP Questionnaire, while the GPFF Scale was comprised of five variables, all contained within question 19i–v of the TLP Questionnaire (Table 1). The LOP and the GPFF scales were calculated by summing the response values for the constituent questions where VL = 5, SL = 4, NQS = 3, SUL = 2 and VUL = 1, and then each was converted, proportionally, to a 100-point scoring scale with the top score indicating higher likelihood of participation or higher fear of participation, respectively, for the LOP Scale and GPFF Scale.

The variable of age was calculated from the date of birth variable on the TLP Questionnaire. The level of education and level of income variables were collected in an ordinal listings of nine ascending categories of educational level and of 10 ascending categories of income level. They were then each collapsed into three categories for the demographic table and the multivariate analyses. To acknowledge and account for cultural differences among the cities (i.e., above and beyond simple demographic differences), the variable of "city" was included as a separate covariate in all multivariate analyses.

Table 2. Distribution of the 1,162 subjects by age, sex, education, income within racial/ethnic groups for the three-city research subject study (unweighted)

Race/Ethnic Group	Mean Age (SD)	% Female	Education Level	Income Level
Blacks[1,2] (n=356)	47.2 (15.5)	67.4%	< High-school grad = 18.1% High-school grad/+ = 54.0% College grad/+ = 28.0%	<$20,000 = 33.5% $20–$74,999 = 57.8% ≥$75,000 = 8.7%
Whites[1,3] (n=493)	48.4 (17.1)	63.3%	< High-school grad = 11.8% High-school grad/+ = 42.2% College grad/+ = 45.9%	<$20,000 = 20.8% $20–$74,999 = 56.5% ≥$75,000 = 23.7%
Hispanics[2,3] (n=313)	44.3 (15.8)	68.4%	< High-school grad = 21.9% High-school grad/+ = 41.2% College grad/+ = 37.0%	<$20,000 = 42.3% $20-74,999 = 49.7% ≥$75,000 = 8.0%

Statistically significant contrasts: 1 for Blacks versus Whites contrast: differed on education and income (p≤0.05); 2 for Blacks versus Hispanics contrast: differed on age and education (p≤0.05); 3 for Hispanics versus Whites contrast: differed on age, education and income (p≤0.05)

Statistical Analysis

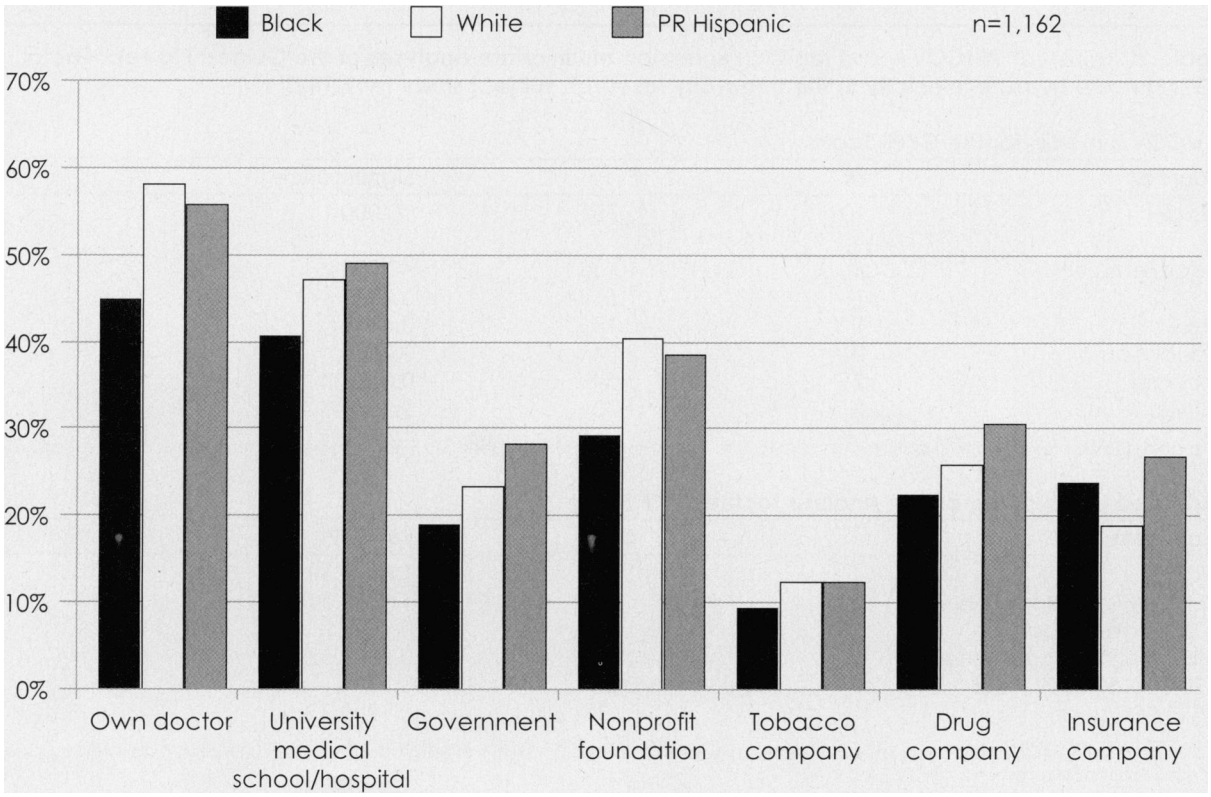
ANCOVA multivariate analyses and logistic regression analyses, which accounted for the multistage sampling techniques used in the RDD survey, were performed. ANCOVA multivariate analysis was used to determine whether the LOP Scale or the GPFF Scale scores differed across the racial/ethnic groups adjusting for key variables. The final ANCOVA multivariate analyses resulted from a two-step process. Step 1 consisted of a bivariate analysis of each independent variable (race/ethnicity, age, sex, education, income and city) by each dependent variable with alpha set at 0.05. Step 2 consisted of an ANCOVA multivariate analysis for the study sample as a whole with race/ethnicity as the independent variable with the model for any of the two dependent variables (GPFF and LOP), including only those covariates that achieved statistical significance in Step 1. Finally, for each dependent variable (GPFF and LOP) for which statistically significant findings were observed, pairwise comparisons, using the post hoc Bonferroni criterion, were conducted to explore two-way differences (i.e., blacks versus whites, blacks versus Hispanics and Hispanics versus whites).

Whenever an ANCOVA-adjusted analysis showed a statistically significant difference for LOP or GPFF scales across the racial/ethnic groups, a second step logistic regression analysis was planned to be performed across racial/ethnic groups, and an odds ratio (OR) was calculated to measure of the magnitude of this observed difference. As the LOP and GPFF scales are both continuous variables, a series of correlation analyses seeking the maximum correlation point between the respective scale score and its individual constituent items was used to determine the best dichotomization point for that scale, as required for conducting the logistic regression analysis. As a result of this maximum correlation analysis, the median score was selected as the most appropriate dichotomization cut-off point for the logistic regression analysis.

RESULTS

In this study, the TLP Questionnaire was administered to 1,162 adults (356 African Americans, 313 Puerto-Rican Hispanics and 493 non-Hispanic whites) in three cities: San Juan, Baltimore and New York City, with response rates by city, of 52%, 51% and 44%, respectively. The overall completion rate (# of completed interviews/# of initiated interviews) was 82.6%. The majority of African Americans came from New York City (54.5%) and Baltimore (41.9%), while the majority of Puerto-Rican Hispanics came from San Juan (49.8%) and New York City (47.9%); non-Hispanic whites main-

Figure 1. Percentage willing to participate in biomedical studies dependent upon "who" was running the study for blacks, whites and Puerto-Rican (PR) Hispanics in the three-city research subject study



ly came from New York City (63.7%) and Baltimore (33.3%). Table 2 shows the age, sex, education and income distribution of the 1,162 subjects within the three racial/ethnic groups.

The unadjusted analysis of question 16, a direct general "gestalt-type" inquiry on the subject's overall willingness to participate in biomedical research, revealed no statistically significant differences among blacks, Puerto-Rican Hispanics and whites (28.0%, 31.0% and 33.1%, respectively; p=0.29) regarding the percentage of each racial/ethnic group's willingness to participate (i.e., the combination of the VL + SL responses). These data show that the vast majority of whites (67%) as well as blacks (72%) and Puerto-Rican Hispanics (69%) self-reported that they did not want to participate in biomedical research projects.

Figure 1 shows the unadjusted findings on willingness to participate in biomedical studies based on questions 17a–g, which addressed the influence of who was conducting the study (Q17). The data reveal a large range in percent willing to participate depending upon who was conducting the study, with two "who" categories (your own doctor and university medical school/hospital) at the high end of the rankings with 40–58% willing and one at the low end (tobacco companies) with 10% willing. Blacks indicated they were less willing to participate than whites on six of the seven prompts in question 17 on who was conducting the study. Interestingly, the three racial/ethnic groups, while showing some differences in response to any one "who" probe, exhibited—on the whole—very similar ratings in regards to the relative ranking of who was to be trusted, as can be readily seen in Figure 1.

In parallel fashion, Figure 2 shows the unadjusted findings on willingness to participate in biomedical studies based on the question that addressed the influence of "what one is asked to do in the biomedical study" (questions 18a–i). Again, a large range is exhibited depending upon "what one is asked to do" in biomedical studies and, again, the three racial/ethic groups demonstrate very similar ratings across the nine specific probes (i.e., they appear to more or less travel together "up and down" the scale of willingness to participate). Only in the two "what asked to do" categories involving blood did blacks indicate the lowest willingness to participate (i.e., for giving blood and having an IV).

Unadjusted mean GPFF Scale and LOP Scale scores for each racial/ethnic group are shown in Figure 3. While the mean GPFF Scale scores for blacks (59.8, SD ± 27.9) and for Puerto-Rican Hispanics (60.4, SD ± 26.2) were virtually equal, whites had a lower mean GPFF score (50.6, SD ± 26.8). The two-way contrasts for mean GPFF Scale scores were statistically significantly for both the blacks versus whites, and the Puerto-Rican Hispanics versus whites, contrasts ($p<0.05$). For the LOP Scale, the observed mean LOP scores for blacks, whites and Puerto-Rican Hispanics were 41.8 (± 21.1), 42.0 (± 21.2) and 49.6 (± 20.8), respectively, with no statistically significant difference between blacks and whites, but each of the contrasts between Puerto-Rican

Table 3. Adjusted* ANCOVA and logistic regression multivariate analyses of the Guinea Pig Fear Factor (GPFF) scale by race/ethnicity in the three-city research subject study (n=1,162)

ANCOVA model for the GPFF Scale

Source	df	F	Significance
Model	10	5.75	<0.0001
Race/ethnicity	2	10.53	<0.0001[a]
Age	1	1.61	0.2045
Sex	1	0.18	0.6741
Education	2	2.84	0.0588
Income	2	3.54	0.0293
City	2	1.70	0.1840

a a post hoc test revealed that both blacks and Puerto-Rican Hispanics had higher GPFF Scale scores as compared to whites ($p \leq 0.01$)

Adjusted Logistic Regression Analysis for the GPFF Scale

Variables	OR[a]	95% CI[b]
Blacks[1]	1.65	1.01–2.69
Puerto-Rican Hispanics[1]	2.68	1.58–4.54
≥ College graduate[2]	1.89	0.91–3.91
≥ High-school graduate[2]	1.71	0.86–3.38
$20,000–$74,999/year[3]	0.63	0.41–1.11
>$75,000/year[3]	0.37	0.19–0.71

a OR: odds ratio; b CI: confidence interval Reference groups: 1 Whites; 2 < High-school graduate; 3 <$20,000/year; * adjusted for age, sex, education, income and city

Hispanics and the other two racial/ethnic groups being statistically significant.

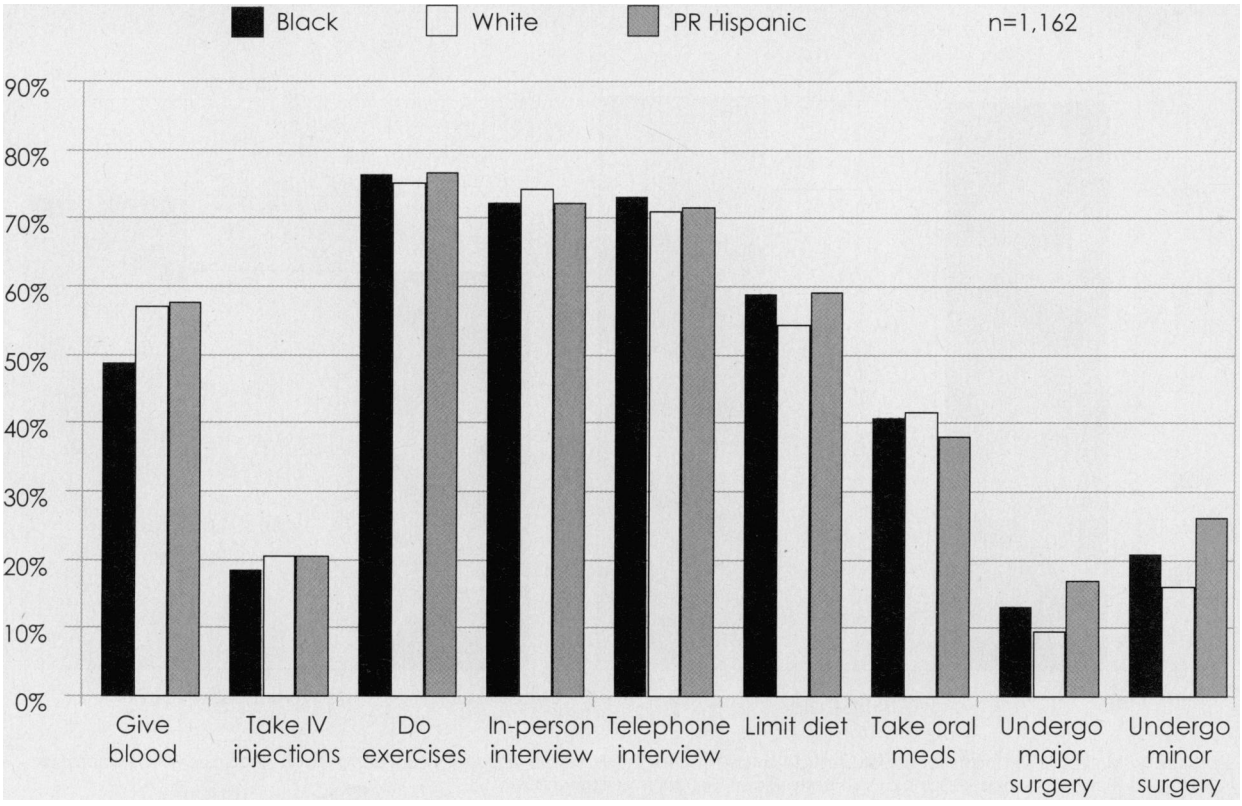
Table 3 shows both the adjusted ANCOVA and logistic regression multivariate analyses for the GPFF Scale using race/ethnicity as the independent variable. The ANCOVA model shown in the top half of Table 3 resulted from a two-step process that adjusted for age, sex, education, income and city. The adjusted results for the GPFF Scale show that the race/ethnic factor was statistically significant ($p<0.001$), as were the variables of education and income. A post hoc test of adjusted GPFF means, using the Bonferroni criterion, revealed that both blacks and Puerto-Rican Hispanics had a significantly higher GPFF Scale score as compared to whites ($P<0.0001$). Conversely, the adjusted results for the LOP Scale across the racial/ethnic groups was not statistically significant ($p=0.87$).

Given that the ANCOVA adjusted analysis showed a statistically significant difference for the GPFF Scale across the racial/ethnic groups, the bottom half of Table 3 shows the logistic regression multivariate analysis for the GPFF Scale adjusted for race, age, sex, education and income. The findings revealed that, controlling for important differences in education and income, the OR of have a GPFF Scale score above the median (indicating more fear) for blacks, as compared to whites was 1.65 (95% CI: 1.01–2.69) and for Puerto-Rican Hispanics, as compared to whites, the odds ratio was 2.65 (95% CI: 1.58–4.54).

DISCUSSION

Taking the findings from these two complex measures (LOP Scale and GPFF Scale) together for this study sample of 1,162 subjects, the final conclusion is that while both blacks and Puerto-Rican Hispanics are more likely to report a higher level of fear related to participation in biomedical studies than are whites, they are nevertheless just as likely as whites to be willing to participate in biomedical research studies, as measured by the LOP Scale. While the three racial/ethnic groups did not differ on the self-reported willingness to participate, Hispanics reported slightly (with borderline significance) higher likelihood of willingness to participate in biomedical research than did blacks. Based on these findings using the LOP, the recruitment of minority subjects for biomedical studies appears to be a fully attainable goal for most types of biomedical studies, in addition to being desirable to ensure diversity in study populations. It should be noted that the findings from this current three-city, follow-up study as presented in this report address the broad issue of willingness to participate in biomedical studies in minority popula-

Figure 2. Percentage willing to participate in biomedical studies dependent upon "what one is asked to do" for blacks, whites and Puerto-Rican (PR) Hispanics in the three-city research subject study



tions, and not do not specifically address the subissue of whether general awareness of or specific knowledge about the Tuskegee Syphilis Study as an separate independent variable had a direct influence on that willingness to participate.

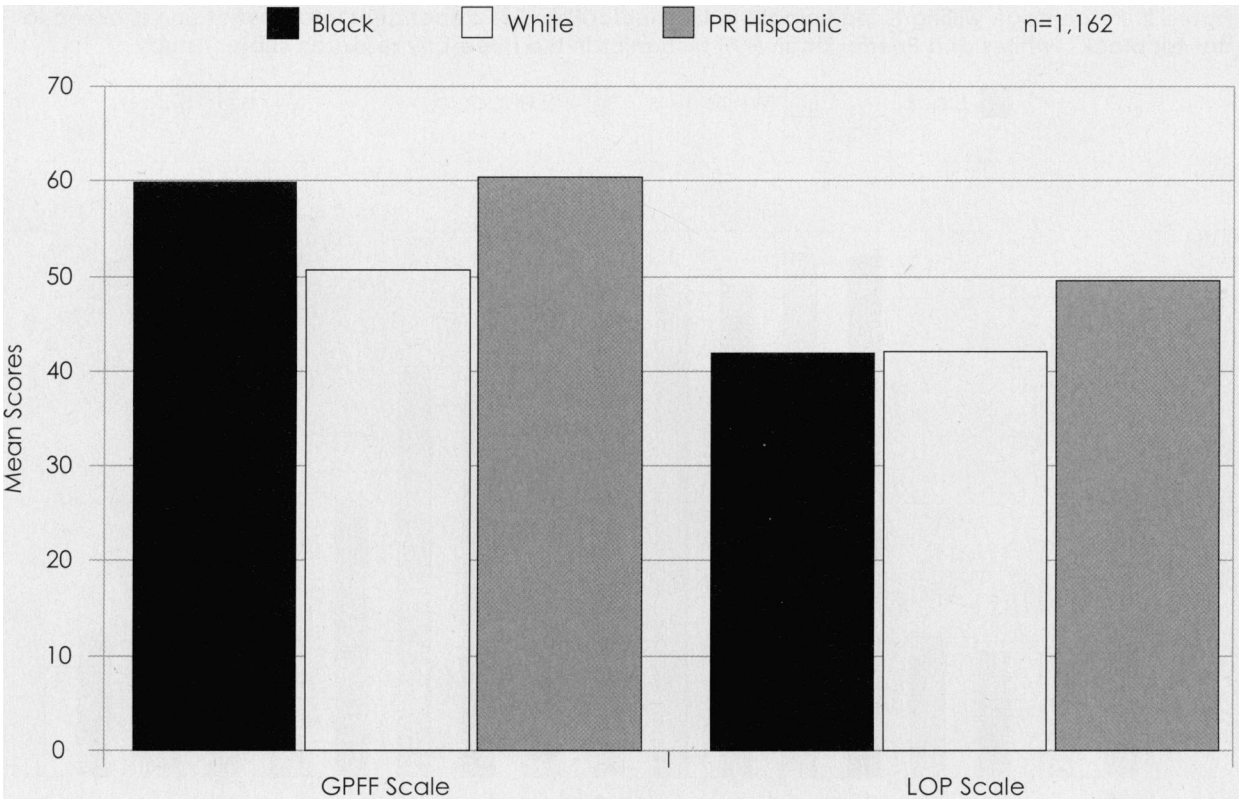
The findings of this three-city research subject study using the TLP Questionnaire agree very closely with the findings of our prior TLP Study, which administered the same TLP Questionnaire to blacks, whites and Hispanics in four other U.S. cities three years earlier (i.e., between 1999–2000).[30] In that prior four-city study, the major findings were extremely similar to those found in this three-city research subject study: 1) <33% of any of the three racial/ethnic groups indicated that they were likely to participate in biomedical studies; 2) there were no statistically significant differences on the LOP Scale among blacks, whites and Hispanics; 3) the mean LOP and GPFF Scale scores for each of the three racial/ethnic groups were nearly identical to those in this current study; and 4) the odds of blacks, as compared to whites, having a higher GPFF Scale score were 1.8 (versus a nearly identical OR of 1.7 in this current study). The findings from this current three-city research subject study also are in very close agreement with the findings from our prior four city study in regards to the pattern of responses shown in Figures 1 and 2 in this report (i.e., the percentage indicating willingness to participate dependent upon "who was conducting the study" and "what one was asked to do in a study" were nearly identical in these two studies).

Given the findings between these two sets of cities, the consistency of the LOP and GPFF scales has been demonstrated across a range of U.S. cities (i.e., two northern cities, one mid-Atlantic city, two southern cities, one southwestern city and one city in Puerto Rico).

The findings of this study on willingness to participate also agrees with those of Brown and Topcu[28] in that neither study found a statistically significant difference between blacks and whites on self-reported willingness to participate as subjects in biomedical studies, even though the Brown and Topcu study was limited to older adults and was only asking about participation in cancer studies. Moreover, our findings on the GPFF Scale are in general agreement with prior studies that did present findings that clearly indicated a higher level of distrust in biomedical research among their black subjects as compared to their white subjects.[26,27,31]

Interestingly, recently published articles which have

Figure 3. Guinea Pig Fear Factor (GPFF) Scale and Likelihood of Participation (LOP) Scale scores for blacks, whites and Puerto-Rican (PR) Hispanics in the three-city research subject study



Statistically significant contrasts for unadjusted data (p≤0.05): for GPFF: blacks versus whites; Puerto-Rican Hispanics versus whites, for LOP: Puerto-Rican Hispanics versus blacks; Puerto-Rican Hispanics versus whites

directly evaluated actual enrollment rates of minorities into biomedical research studies have found that minorities (largely blacks and Hispanics) do enroll, proportionally, in clinical research at expected and targeted rates when a reasonable effort is made to enroll minority participants. A report on the enrollment of minorities into the national Women's Health Initiative Study (WHIS) stated that "the WHI achieved 93% of is targeted minority goal" and noted that "recruitment yields for [black and Hispanic] minority groups surpassed that of white women."[32] A recent review of 20 studies which reported enrollment rates by race and ethnicity for >70,000 individuals involving a wide range of biomedical studies (ranging from interview studies to drug treatment and surgical trials) reported that they "found very small differences in the willingness of minorities, most of whom were blacks and Hispanics in the United States, to participate in health research compared to non-Hispanic whites" and concluded that "racial and ethnic minorities in the United States are as willing as non-Hispanic whites to participate in health research."[33] A recent article has even summated recruitment strategies to enroll minority subjects into studies and succinctly identified barriers to research participation by minorities (and also suggests a comprehensive conceptual model describing how individuals make rational decisions about participation in biomedical research studies).[34]

The TLP Questionnaire was developed to address and understand a wide range of issues related to the recruitment and retention of blacks and other minorities in biomedical research studies. Attainment of this goal is critical in order to ensure that the findings from biomedical studies provide health data on the diverse populations of the United States, and to assist biomedical researchers to achieve compliance with 1994 NIH Guidelines for the Inclusion of Women and Minorities in clinical studies.[35]

CONCLUSION

The findings from this three-city research subject study provides independent evidence that there was: 1) no difference in self-reported willingness to participate in biomedical research, as measured by the LOP Scale in the TLP Questionnaire, among blacks, Puerto-Rican Hispanics and whites; and 2) a statistically significant difference across the three racial/ethnic groups as regards the GPFF, with the odds of having a higher fear of participation in biomedical research being statistically higher in both blacks and Puerto-Rican Hispanics, as compared to whites. In addition, comparison of the finding from this three-city study on both the LOP and GPFF scales with the four-city study conducted three years prior provides strong evidence that there is a consistency of these scales within the TLP Questionnaire over time and across differing U.S. populations.

The combination of these two main findings, from both the current three-city study and the prior four-city study, leads to the conclusion that blacks and Hispanics self-report that despite having a higher fear of participation they are just as likely as whites to participate in biomedical research. These findings, consistent in >2,200 subjects across seven U.S. cities, begin to provide a body of findings that, for the first time, can be generalized to broader U.S. populations. Further, these findings are largely in concert with the few similar early exploratory studies on self-reported participation in the literature, and with the emerging literature that has assessed actual enrollment rates in biomedical studies by race and ethnicity.

ACKNOWLEDGEMENTS

The authors wish to thank the 31 student researchers from Oakwood College, Spelman College, Morehouse College, Vassar College, Central Connecticut State University, Simsbury High School, Howard University College of Dentistry, University of Puerto Rico School of Dentistry, University of Connecticut School of Dental Medicine, and New York University College of Dentistry, who over the years contributed to the development and pilot testing of the Tuskegee Legacy Project Questionnaire.

REFERENCES

1. Jones, JH. Bad Blood: the Tuskegee syphilis experiment. New York, NY: Free Press; 1981.

2. Fletcher JC. Tuskegee Syphilis Study Legacy Committee Report of 1996, CDC Workshop at Tuskegee University. In: Reverby SM, ed. Tuskegee's Truths: Rethinking the Tuskegee Syphilis Study, University of North Carolina Press; 2000:559-566.

3. *New York Times*, June 17, 1997, p.1,10.

4. Hunninghake DB, Darby CA, Probstfield JL. Recruitment experience in clinical trials: literature summary and annotated bibliography. *Control Clin Trials.* 1987;8(4 suppl):6S-30S.

5. Swanson MS, Ward AJ. Recruiting minorities into clinical trials: toward a participant-friendly system. *J Natl Cancer Inst.* 1995;87(23):1747-1759.

6. 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(11):1498-1505.

7. Kressin NR, Meterko M, Wilson, NJ. Participation in Biomedical Research: Are There Racial Disparities within VA? *J Natl Med Assoc.* 2000;92:62-69.

8. Shavers-Hornaday VL, Lynch CF. Why are African Americans under- represented in medical research studies? Impediments to Participation. *Ethn Health.* 1997;2:31-45.

9. Gamble, VN. A legacy of distrust: African Americans and medical research. *Am J Prev Med.* 1993;9:(suppl)35-38.

10. Caplan AL. Twenty years after. The Legacy of the Tuskegee Syphilis Study. When Evil Intrudes. *Hastings Cent Rep.* 1992;22(6):29-32.

11. Benedek TG. The Tuskegee study of syphilis: analysis of moral versus methodologic aspects. *J Chron Dis.* 1978;31:35-50.

12. Corbie-Smith GM. Minority Recruitment and Participation in Health Research. *N C Med J.* 2004;6(6):385-387.

13. Gamble VN. The Tuskegee Syphilis Study and women's health. *J Am Med Womens Assoc.* 1997;52(4):195-196.

14. Matthews AK, Sellergren SA, Manfredi C, et al. Factors influencing medical information seeking among African American cancer patients. *J Health Commun.* 2002;7(3):205-219.

15. Pressel DM. Nuremberg and Tuskegee: lessons for contemporary American medicine. *J Natl Med Assoc.* 2003;95(12):1216-1225.

16. Editorial. Race, ethnic group, and clinical research. *BMJ.* 2003;327:763-764.

17. Bouleware LE, Cooper LA, Ratner LE, et al. Race and Trust in the health care system. *Public Health Rep.* 2003;118(4):358-365.

18. White RM. Misinformation and misbeliefs in the Tuskegee Study of Untreated Syphilis fuel mistrust in the healthcare system. *J Natl Med Assoc.* 2005;97(11):1566-1573.

19. Fairchild AL and Bayer R, Uses and Abuses of Tuskegee. *Science.* 1999; 284(5416):919-921.

20. Katz, RV, Kegeles SS, Green BL, et al. The Tuskegee Legacy Project: history, preliminary scientific findings, and unanticipated societal benefits. *Dent Clin North Am.* 2003;47:1-19.

21. Corbie-Smith G, Thomas SB, Williams MV, et al. Attitudes and beliefs of African Americans toward participation in medical research. *J Gen Intern Med.* 1999;14:537-546.

22. Green BL, Partridge EE, Fouad MN, et al. African-Americans attitudes regarding cancer clinical trials and research studies: results from focus group methodology. *Ethn Dis.* 2000;10(1):76-86.

23. Freimuth VS, Quinn SC, Thomas SB, et al. African Americans' views on research and the Tuskegee Syphilis Study. *Soc Sci Med.* 2001; 52:797-808.

24. Bates BR, Harris TM. The Tuskegee Study of untreated syphilis and public perceptions of biomedical research: a focus group study. *J Natl Med Assoc.* 2004;96(8):1051-1064.

25. Green BL, Maisiak R, Wang MQ, et al. Participation in health education, health promotion, and health research by African Americans: effects of the Tuskegee Syphilis Experiment. *J Health Educ.* 1997;28(4):196-201.

26. Sengupta S, Strauss RP, DeVellis R, et al. Factors affecting African-American participation in AIDS research. *J Acquir Immune Defic Syndr.* 2000;24(3):275-284.

27. Shavers VL, Lynch CF, Burmeister LF. Knowledge of the Tuskegee Study and its impact on willingness to participate in medical research studies. *J Natl Med Assoc.* 2000;92(12):563-572.

28. Brown DR, Topcu M. Willingness to participate in clinical treatment research among older African Americans and Whites. *Gerontologist.* 2003;43(1):62-72.

29. Green BL, McCallum JM, Arekere DM, et al. The relationship between awareness and knowledge of the USPHS Syphilis Study at Tuskegee and willingness to participate in biomedical research: a Literature Review. *J Health Care Poor Underserved.* 2006;17:716-733.

30. 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.

31. Brandon DT, Issac LA, LaVeist TA. The legacy of Tuskegee and trust in medical care: is Tuskegee responsible for race differences in mistrust of medical care? *J Natl Med Assoc.* 2005;97(7):951-956.

32. Fouad, MN, Corbie-Smith G, Curb D, et al. Special Populations recruitment for the Women's Health Initiative: successes and limitations. *Control Clin Trials.* 2004;25:335-352.

33. Wendler D, Kington R, Madans J, et al. Are Racial and Ethnic Minorities Less Willing to Participate in Health Research? *PLoS Med.* 2006;3(2) e19:1-10.

34. Farmer DF, Jackson SA, Camacho F, et al. Attitudes of African American and Low Socioeconomic Status White Women toward Medical Research. *J Health Care Poor Underserved.* 2007;18(1)85-99

35. NIH Guidelines on Inclusion of Women and Minorities as Subjects in Clinical Research, NIH Guide, 23:(11)March 18, 1994. http://grants.nih.gov/grants/guide/notice-files/not94-100.html. ■

We Welcome Your Comments

The *Journal of the National Medical Association* welcomes your Letters to the Editor about articles that appear in the *JNMA* or issues relevant to minority healthcare. Address correspondence to EditorJNMA@nmanet.org.

AIDS Care, September 2007; 19(8): 1039–1047

Routledge
Taylor & Francis Group

Factors associated with discontinuation of antiretroviral therapy in HIV-infected patients with alcohol problems

T. W. KIM[1], A. PALEPU[2], D. M. CHENG[1,3], H. LIBMAN[4], R. SAITZ[1,5], & J. H. SAMET[1,6]

[1]*Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University, School of Medicine, Boston, MA,* [2]*Centre for Health Evaluation and Outcome Sciences, St. Paul's Hospital, Department of Medicine, University of British Columbia, Vancouver, British Columbia,* [3]*Department of Biostatistics, Boston University School of Public Health,* [4]*Beth Israel Deaconess Hospital, Harvard Medical School, Boston, MA,* [5]*Youth Alcohol Prevention Center, Department of Epidemiology, and* [6]*Department of Social and Behavioral Sciences, Boston University, School of Public Health, Boston, MA*

Abstract

Although mortality rates among HIV-infected populations have declined with the advent of combination antiretroviral therapy (ART), patients with substance use disorders have benefited less from these therapies. While adherence to ART has been well studied, less is known about factors associated with discontinuation of ART. The aim of this study is to investigate predictors of discontinuation of ART in HIV-infected patients with alcohol problems, focusing on their substance use and depressive symptoms. The study cohort ($n=266$) was prospectively assessed with biannual standardised interviews between 2001 and 2005. Four predictor variables (cocaine, heroin, heavy alcohol use and substantial depressive symptoms) were assessed six months prior to the outcome (ART discontinuation). Longitudinal logistic regression models examined the association between predictor variables and ART discontinuation adjusting for age, gender, race/ethnicity, homelessness, CD4, HIV RNA and HIV Symptom Index. Subjects were 77% male; 43% black; 22% homeless; 45% used cocaine; 20% used heroin; 29% had heavy alcohol use; and 40% had substantial depressive symptoms. Discontinuation occurred in 135 (17%) of the observations ($n=743$). In bivariate analyses, cocaine use, heroin use and depressive symptoms were significantly associated with ART discontinuation but heavy alcohol use was not. In the multivariable model, substantial depressive symptoms (adjusted odds ratio (AOR) = 1.66; 95% confidence interval (CI): 1.04, 2.65) but not cocaine (AOR = 1.28; 95%CI: 0.76, 2.16) or heroin use (AOR = 1.27 95%CI: 0.66, 2.44), remained significantly associated with ART discontinuation. Among HIV-infected adults with alcohol problems, depressive symptoms, but not substance use, predicted subsequent ART discontinuation. Recognition and treatment of depressive symptoms in this population may result in better maintenance of ART and its associated clinical benefits.

Introduction

Increasingly, researchers and health policy-makers are recognising the difficulty in translating medical treatment advances to vulnerable populations. Although mortality from HIV infection has declined with the advent of combination antiretroviral therapy (ART), patients with substance use disorders have benefited less (Lucas et al., 2006). Since patients with intermittent use of ART have higher mortality rates (Hogg et al., 2002; Riley et al., 2005), sustaining these drug regimens over long periods of time is one of the major challenges in the care of HIV-infected patients.

Studies of predictors of ART discontinuation among populations with HIV and addictions are limited. The impact of injection drug use (IDU) is often evaluated by analysing subjects' HIV transmission risk behavior (Chen et al., 2003; D'Arminio et al., 2005; Mocroft et al., 2005) and may assume that the impact of substance use is unaffected by its cessation. Evaluating IDU as a risk category does not account for the spectrum of use that may occur at different stages of a patient's addiction trajectory (McLellan et al., 2000). The relationship between the use of specific drugs of abuse and discontinuation of ART also has received little attention (Li et al., 2005; Yuan et al., 2006) despite the spectrum of biological and behavioural effects caused by individual stimulants and depressants. Also, there has been limited examination of the impact of alcohol use on ART discontinuation despite the high prevalence of alcohol use disorders among HIV-infected persons (Conigliaro et al., 2003; Galvan et al., 2002; Samet et al., 2004) and its association with suboptimal adherence (Braithwaite

Correspondence: Theresa W. Kim, Section General Internal Medicine, Boston University School of Medicine, 91 E. Concord Street, Ste. 200, Boston, MA 0248. E-mail: theresa.kim@bmc.org

ISSN 0954-0121 print/ISSN 1360-0451 online
DOI: 10.1080/09540120701294245

et al., 2005; Samet et al., 2004) and HIV disease progression (Samet et al., 2003). Furthermore, studies examining addictions and ART use frequently do not include information about depressive symptoms (Kerr et al., 2005; Riley et al., 2005; Yuan et al., 2006). This is important because of the high co-morbidity of substance use and mood disorders in the general population (Grant et al., 2004), which is even greater among HIV-infected patients (Galvan et al., 2003).

The aim of this study is to investigate predictors of discontinuation of ART in HIV-infected patients with alcohol problems, focusing on their substance use and depressive symptoms. We hypothesised that active substance use (cocaine, heroin and heavy alcohol use) and depressive symptoms would each be associated with a greater likelihood of ART discontinuation.

Methods

Study design and population

This study examines predictors of discontinuation of ART among subjects in the HIV-Longitudinal Interrelationships of Viruses and Ethanol (HIV-LIVE) study. This was a prospective, observational cohort study of HIV-infected patients with past or current alcohol problems. Subjects were prospectively followed with biannual standardised interviews between August 2001 and March 2005.

The 400 subjects of the HIV-LIVE study were recruited from different sources including: (1) a previous cohort study (Samet et al., 2003) of people with HIV and alcohol problems ($n = 154$; 38%); (2) the Diagnostic Evaluation Unit (Samet et al., 1995), an intake clinic for HIV-infected patients at Boston Medical Center ($n = 88$; 22%); (3) the HIV Primary Care and Specialty Clinics at Beth Israel Deaconess Medical Center ($n = 31$; 8%); and (4) additional healthcare centres, homeless shelters, drug treatment programs, subject referrals and flyers ($n = 127$; 32%).

Eligibility criteria for the HIV-LIVE study included: (1) documented HIV antibody by ELISA confirmed by Western Blot; (2) two or more affirmative responses to the CAGE alcohol screening questionnaire (Buchsbaum et al., 1991) or co-investigator physician diagnosis of alcohol abuse or dependence; (3) ability to speak English or Spanish; and (4) at least one contact person who was likely to know their whereabouts. Exclusion criteria included: (1) scoring <21 on the 30-item Mini-Mental State Examination (Smith et al., 2006); (2) trained interviewer assessment that the patient was incapable of comprehending informed consent or answering the interview questions; or (3) plans to move from the Boston area in the subsequent two years.

Subjects were included in the analyses if they reported being on ART at any study interview (hereafter referred to as T_N) and then returned six months later for the next scheduled interview (T_{N+1}).

Data collection

After enrolment, research associates interviewed subjects using a standardised instrument at baseline and biannually thereafter. For interviews conducted in Spanish, standardised scales in Spanish were used when available; the remainder of the questionnaire was translated from English into Spanish, back translated to check for accuracy and then corrected. CD4 cell counts and HIV RNA viral load measurements conducted as part of clinical care were recorded if within four months of the research interview. If they were not available through clinical care, blood samples were obtained and tested for these parameters.

All subjects provided written informed consent prior to enrolment. The Institutional Review Boards of Boston University Medical Center and Beth Israel Deaconess Medical Center approved this study. Additional privacy protection was secured by the issuance of a Certificate of Confidentiality by the Department of Health and Human Services to protect subjects from release of their research data even under a court order or subpoena.

Dependent variable

We defined discontinuation of ART as either: (1) being off ART medications at T_{N+1} (This was assessed with the question: 'Are you currently taking antiretroviral medications for HIV?'); or (2) stopping all ART medications for ≥ 30 days between T_N and T_{N+1}. This second criterion was included since some subjects who were taking ART medications on the day of the follow-up research interview (T_{N+1}) may have stopped them for an extended period of time since the last interview. This criterion was assessed with the questions: 'Did you stop taking all of your HIV medications for more than a day during the past six months (not including switching or changing medications)?' and, for those answering affirmatively, asking about the number of days without ART.

Main independent variables

The main independent variables were: (1) heavy alcohol consumption over the previous month using the National Institute on Alcohol Abuse and

Alcoholism definition, which is more than 14 standard drinks per week (or more than four drinks in a day) for men or more than seven drinks per week (or more than three drinks in a day) for women; past month alcohol consumption was assessed using a validated calendar method (Sobell, 1992); (2) any cocaine use in the past six months using the Composite International Diagnostic Interview (CIDI), Short Form (Kessler et al., 1998); (3) any heroin use during the past six months using the CIDI; Short Form and (4) substantial depressive symptoms, defined as a score of ≥ 23 on the 20-item Center for Epidemiologic Studies-Depression scale (CES-D) (Radloff, 1977). This score indicates substantial depressive symptoms in patients with chronic medical conditions (Golub et al., 2004). All main independent variables were modelled as dichotomous variables.

To reduce the likelihood that we were measuring depressive symptoms rather than symptoms related to HIV infection, secondary analyses were performed using a modified version of the CES-D (Ickovics et al., 2001) removing somatic complaints (fatigue, anorexia, lack of energy, insomnia and poor concentration).

Additional secondary analyses were performed to assess the association between a current diagnosis of alcohol dependence or drug dependence (in models that did not include substance use) and ART discontinuation. Current alcohol dependence was determined with the CIDI (Robins et al., 1988) at the baseline (recruitment) HIV-LIVE research interview and the CIDI, Short Form at subsequent interviews. Current drug dependence was assessed with the CIDI, Short Form at all research interviews.

Other independent variables

The following covariates were considered for inclusion in regression models based on clinical considerations and evidence from previous studies of ART use in patients with addictions: age; gender; race/ethnicity; homelessness (any night spent in a shelter or on the street in past six months) (Kertesz et al., 2005); time in jail (any in past six months); and HIV symptoms. We measured HIV-related symptoms using the 20-item HIV Symptom Index, which has been used to monitor symptoms in the AIDS Adult Clinical Trials Group (Justice et al., 2001). We observed a strong correlation between the CES-D and the HIV Symptom Index ($r = 0.47$); therefore, we used a modified HIV Symptom Index described in other studies (Kilbourne et al., 2002) excluding questions that overlap with the CES-D questionnaire (e.g. questions about sadness, anxiety, sleep problems, fatigue and memory loss). Laboratory parameters included CD4 cell count, HIV $\log_{10}$ RNA and hepatitis C antibody status.

Statistical analysis

We used generalised estimation equations (GEE) (Zeger et al., 1988) and logistic regression methods to examine factors associated with ART discontinuation. Predictor variables were assessed at time T_N approximately six months prior to the outcome at T_{N+1}. Only observations where the subject was on ART at T_N contributed to analyses. Since subjects could contribute multiple pairs of observations, we used the GEE approach to adjust for the correlation between repeated measures from the same subject over time using a working independence correlation matrix. The empirical standard errors from the GEE approach were used for all analyses.

To avoid the potential for colinearity, we assessed the correlation between each pair of independent variables and none of the variables included in regression models were highly correlated (i.e. no correlation was greater than 0.40). Covariates were selected for inclusion in the multivariable models if their p-values were <0.10 in unadjusted analyses. To avoid linearity assumptions, continuous covariates were categorised based on descriptive statistics (e.g. median or tertiles) from the study sample. All predictor variables except for gender, age and race/ethnicity were allowed to vary with time. Health insurance status was not included in the models since 99% of all subjects had access to private, Medicaid or other publicly funded health insurance. For descriptive purposes, we calculated the proportion of subjects who ever discontinued ART during the study using Kaplan-Meier analyses. Subjects were censored if they did not complete a follow-up interview.

Analyses were conducted using two-sided tests and an alpha level of 0.05 was used to determine statistical significance. All analyses were carried out using SAS version 8.2 (SAS Institute, Cary, NC).

Results

Among the 400 HIV-LIVE subjects, 303 (76%) were on ART at least once during the study and therefore eligible for analysis. Thirty-seven of these 303 subjects (12%) were not included in this analysis because of a missed study visit at the subsequent follow-up. No significant differences were found between the analytic sample (266/303; 88%) and those with incomplete follow-up (37/303; 12%) in terms of age, race/ethnicity, housing status, heavy alcohol consumption or cocaine or heroin use. A higher proportion of subjects with incomplete follow-up had substantial depressive symptoms

Table I. Baseline characteristics of HIV-infected subjects with past or current alcohol problems on antiretroviral drugs ($n=266$).

Characteristic	Total sample *N* (%)
Female	60 (23)
Race/ethnicity	
Black	115 (43)
Hispanic	64 (23)
White	90 (34)
Non-high school graduate	88 (33)
Homeless[a, b]	59 (22)
Jail[a]	45 (17)
Injection drug use, lifetime	146 (55)
Hepatitis C Ab positive	155 (58)
Substantial depressive symptoms[c]	117 (40)
Heavy alcohol use[d]	77 (29)
Cocaine use[a]	120 (45)
Heroin use[a]	101 (25)
Age, mean (SD) (years)	43.6 (7.2)
Income, median $US (interquartile range)[e]	7500 (7500, 25000)
CD4 count (cells/μl), median (interquartile range)	378 (232, 606)
HIV $\log_{10}$ viral load, median	2.1 (UDT[f], 3.9)

[a]In the past six months.
[b] ≥one night in a shelter or on the street.
[c]Center for Epidemiologic Studies Depression (CES-D) scale ≥ 23.
[d]More than 14 standard drinks per week (or more than four drinks in a day) for men or more than 7 drinks per week (or more than three drinks in a day) for women, in the past month.
[e]Highest in past five years.
[f]UDT = undetectable.

(CES-D ≥23) (27/37; 73%) compared to those in the analytic cohort (117/266; 44%) ($p=0.001$).

Table I shows demographic and clinical characteristics of the study sample ($n=266$) at baseline. The cohort was predominantly male (77%) and non-white (either black or Hispanic—43 and 23% respectively). Study participants were mostly poor (median income $7,500) and a substantial minority (22%) was homeless. More than half (55%) of the cohort reported past injection drug use; a similar proportion (58%) tested positive for hepatitis C antibody. Almost half (45%) reported past six-month cocaine use and a quarter reported heroin use. Past month heavy alcohol consumption was reported by 29% of subjects. A significant proportion (40%) had substantial depressive symptoms (CES-D ≥23). Twenty-three percent of the sample discontinued ART at least once during the study period.

Subjects were followed every six months for up to six visits, and the median number of observations per subject was three (interquartile range 2–4). For this analysis, the proportion of subjects who completed one, two, three, four, five and six observations during the three-year follow-up period was (# interviews/subjects): 20.3%; 22.6%; 20.3%; 14.3%; 15.0%; and 7.1%, respectively. Overall, the 266 subjects contributed a total of 743 observations to the longitudinal analyses. Discontinuation of ART was reported in 18% (135/743) of the observations; 12.4% (92/743) were not on ART at T_{N+1}, and 5.8% (43/743) had stopped ART for at least 30 days since T_N. The median number of days without ART in the previous six months for this last group was 60 (interquartile range 44–94).

We obtained information about reasons for ART discontinuation in 62% (84/135) of the cases. Subjects could specify multiple reasons for stopping ART from a checklist of responses. In about three-fourths of the cases of discontinuation (62/84; 74%), reasons other than 'MD told me to' were cited. The most common reason was 'just took self off' or 'personal problems' (41/84; 49%), followed by 'MD told me to' (21/84; 25%), 'substance abuse' (14/84; 17%), 'meds made me sick' (10/84; 12%) and issues related to lost medications or lack of refills (8/84; 10%).

As shown in Table II, factors associated with discontinuation of ART in the unadjusted analyses were: cocaine use (unadjusted OR = 1.86; 95%CI: 1.18, 2.94); heroin use (OR = 2.12; 95%CI: 1.25, 3.60); and depressive symptoms (OR = 2.26; 95%CI: 1.49, 3.43) but not heavy alcohol use. Other factors associated with discontinuation were: age less than 40; female gender; homelessness; higher viral load; and more HIV-related symptoms.

Cocaine, heroin and heavy alcohol use were not significant predictors of ART discontination in a multivariable model that adjusted for age, gender, homelessness, CD4 count, HIV viral load and HIV Symptom Index (Table II). Similarly, diagnoses of alcohol dependence (adjusted odds ratio (AOR) = 1.18; 95%CI: 0.71, 1.96) and drug dependence (AOR = 1.54; 95%CI: 0.88, 2.70) were not associated with ART discontinuation.

Subjects with substantial depressive symptoms were more likely to discontinue ART (AOR = 1.66; 95%CI: 1.04, 2.65). Similar results were found when analysing depressive symptoms with the modified CES-D (removing questions about somatic symptoms) (AOR = 1.82; 95%CI: 1.07, 3.11). Other factors independently associated with ART discontinuation were age <40, higher HIV viral load and higher HIV Symptom Index score (8–15).

Discussion

In this sample of HIV-infected adults with past or current alcohol problems, patients with substantial depressive symptoms had almost twice the likelihood of discontinuing ART six months later. Although HIV-infected patients with depression are more likely to have HIV-related symptoms (Kilbourne

Table II. Predictors of discontinuation of antiretroviral therapy in unadjusted and adjusted analyses.

		OR (95%CI)	
Variable	% Discontinued ART	Unadjusted	Adjusted[a]
Heavy alcohol use[b]			
Yes	19	1.23 (0.77, 1.95)	1.01 (0.63, 1.63)
No	16	1	1
Cocaine use[c]			
Yes	23	1.86 (1.18, 2.94)**	1.28 (0.76, 2.16)
No	14	1	1
Heroin use[c]			
Yes	27	2.12 (1.25, 3.60)*	1.27 (0.66, 2.44)
No	17	1	1
Depressive symptoms[d]			
Yes	24	2.26 (1.49, 3.43)***	1.66 (1.04, 2.65)*
No	12	1	1
Age (years)			
<40	24	1.96 (1.20, 3.22)**	1.82 (1.11, 2.94)*
≥40	14	1	1
Gender			
Female	25	1.93 (1.09, 3.42)*	1.69 (0.91, 3.12)
Male	15	1	1
Race/ethnicity			
Non-white	18	1.55 (0.92, 2.60)	–
White	13	1	
Homeless			
Yes	26	1.95 (1.16, 3.29)*	1.34 (0.75, 2.40)
No	15	1	1
Jail			
Yes	24	1.62 (0.84, 3.11)	–
No	16	1	
CD4 cell count			
<374[e]	23	1.59 (0.99, 2.52)	1.07 (0.63, 1.84)
≥374	15	1	1
HIV $\log_{10}$ RNA			
≥2.1[e]	27	2.73 (1.79, 4.18)***	2.28 (1.47, 3.53)**
<2.1	12	1	1
HIV Symptom Index[f]			
8–15	25	2.74 (1.59, 4.75)***	1.94 (1.05, 3.60)**
4–7	16	1.88 (1.12, 3.16)*	1.49 (0.85, 2.62)
0–3	10	1	1
Hepatitis C Ab positive			
Yes	17	1.04 (0.62, 1.61)	–
No	17	1	

*$p < 0.05$.
**$p < 0.001$.
***$p < 0.0001$.
[a]Analyses based on 266 subjects contributing 743 observations. Multivariable model adjusted for heavy alcohol use, cocaine use, heroin use, depressive symptoms, age, gender, homelessness, CD4, HIV $\log_{10}$ RNA and HIV Symptom Index score.
[b]More than 14 standard drinks per week (or more than four drinks in a day) for men, or more than seven drinks per week (or more than three drinks in a day) for women, in the past month.
[c]In the past six months.
[d]Center for Epidemiologic Studies Depression (CES-D) scale in which ≥23 indicates substantial depressive symptoms.
[e]Study sample median.
[f]Modified HIV Symptom Index, in which questions that overlap with the CES-D questionnaire (e.g. questions about sadness, anxiety, sleep problems, fatigue and memory loss) were excluded. Variable was categorised by tertiles for the study sample.

et al., 2002) and patients with HIV symptoms are more likely to discontinue ART, (Ahdieh et al., 2005; Mocroft et al., 2005), we found that the association between depression and ART discontinuation was independent of HIV symptoms. We did not detect an association between heroin, cocaine or heavy alcohol use and subsequent ART discontinuation in multivariable analyses that adjusted for depressive symptoms, clinical indicators of HIV disease severity and housing status.

To our knowledge, no other studies have examined the relationship between heavy alcohol use and ART discontinuation despite the former's high prevalence among HIV-infected drug users (Miguez et al., 2003) and its association with suboptimal adherence (Braithwaite et al., 2005; Samet et al., 2004). This study also differs from previous studies examining substance use and ART utilization in a number of respects: the current study (1) used validated instruments to assess substance use and substance use disorders rather than relying on physician detection in clinical care; (2) prospectively assessed predictor variables at least six months before the outcome suggesting a temporal relationship between depressive symptoms and discontinuation; (3) examined a measure of depressive symptoms highlighting the clinical importance of co-occurring affective disorders with substance use disorders; and (4) included an indicator of HIV quality of life (HIV Symptom Index) often missing in studies of antiretroviral medication use.

Our data are consistent with prior studies examining HIV-infected women (Ahdieh et al., 2005), gay men (Li et al., 2005) and homeless persons (Moss et al., 2004) in which depressive symptoms were found to be associated with discontinuation of ART. Based upon our results, this observation appears to apply to HIV-infected persons with past or current alcohol problems as well. Depression is common in HIV-infected patients (Bing et al., 2001; Fairfield et al., 2001). However, despite being prevalent and associated with higher mortality in this population (Cook et al., 2004: Ickovics et al., 2001), it is underdiagnosed in clinical practice. Asch et al. (2003) found that clinical depression was recognised in only about half of patients; patients with less education were more at risk for having depression underrecognised. Psychopharmacologic and psychotherapeutic treatments are associated with better ART utilization (Cook et al., 2006).

Since mood disorders commonly occur with substance use disorders (Conway et al., 2006; Grant et al., 2004), screening and management of depression in HIV-infected adults with alcohol problems are likely to be particularly important to maximise the benefits of ART. A growing body of evidence suggests that shared neurobiological pathways are involved in addiction and depression (Brady et al., 2005). Even previous alcohol dependence is associated with a fourfold higher risk of depression (Hasin & Grant, 2002). Co-occurring mental health and substance use disorders have worse psychiatric and addiction outcomes than either alone (Compton et al., 2003). This paper's findings suggest that depression may be associated with worse HIV outcomes in this population by means of its association with ART discontinuation.

We did not find an association between cocaine, heroin or heavy alcohol use and ART discontinuation in adjusted analyses. Since eligibility required being prescribed ART, this study examined a specific subgroup of substance users. It may be that physicians offered ART to certain patients with substance abuse based upon unmeasured factors thought to be predictive of better ART adherence. Among IDUs in Vancouver, patient's self-assessment of ability to meet challenges related to ART management (e.g. refilling prescriptions on time, continuing ART when using drugs or in withdrawal etc.) was associated with sustaining ART use rather than frequency of heroin or cocaine use (Kerr et al., 2005). These findings suggest that substance use, per se, may not be reasonable justification for withholding ART on the basis of fears of medication treatment interruption. Instead an assessment of a patient's ability to sustain ART use should involve a careful evaluation for depressive symptoms, self-efficacy, unstable housing (Moss et al., 2004) and HIV-related somatic symptoms.

Similar to other studies (Ahdieh et al., 2001; Kerr et al., 2004; Mocroft et al., 2005), we found that more HIV symptoms predicted subsequent discontinuation of ART. This study adds to the growing literature demonstrating their clinical importance. HIV-related symptoms are common; patients reported an average of 12 in the previous year in one study (Heath et al., 2002). HIV-related symptoms are among the strongest predictors of health-related quality of life (Hays et al., 2000), hospitalisation and mortality (Kilbourne et al., 2002). Our study extends that literature by demonstrating their importance among people with alcohol problems.

The finding that a higher HIV viral load is associated with a greater likelihood of subsequent discontinuation is also consistent with other studies (Ahdieh et al., 2001; Li et al., 2005). Since higher viral loads are often found in patients with lower CD4 counts, this finding may be the consequence of patients with more advanced HIV disease having greater difficulty tolerating ART. However, since our analyses were adjusted for CD4 count and HIV symptoms, a higher viral load (which was measured at the first interview) may have indicated that the

subject was already non-adherent prior to discontinuing therapy.

The following limitations should be considered when interpreting this study's findings. First, we did not have information about why patients discontinued ART in over a third of the cases. This is important as the dynamics of physician-recommended discontinuation because of, for example, intolerable medication side effects, may be different than those of patient-initiated discontinuation without physician consultation. However, the information that was available to us in terms of reasons for ART discontinuation was consistent with overall findings, namely that 49% reported discontinuing for personal reasons versus 17% for substance abuse. Second, we did not have information about provider or clinic characteristics that may have influenced the likelihood of ART discontinuation. Third, we did not have clinical or immunological information on the consequences of discontinuing ART as we have defined it in this study. Some subjects may have discontinued ART for 30 days and then restarted ART without negative clinical consequences such as acquisition of viral resistance mutations. However, prior studies have found higher mortality risk with intermittent use of ART (Riley et al., 2005). Also, because of the observational design of this study, conclusions about causal relationships cannot be inferred. However, information on predictor variables was prospectively assessed before the outcome, which supports a temporal association between depressive symptoms and ART discontinuation. Another study limitation, given the importance of depressive symptoms with regards to ART discontinuation, is that we did not have additional information on psychiatric diagnosis or antidepressant utilization. Finally, although in this analysis heavy alcohol use did not impact ART discontinuation, our sample did not include people without current or past alcohol problems, in whom heavy alcohol use could have an impact on ART utilization. The generalisability of this study's findings may be limited to HIV-infected persons with past or current alcohol problems who are prescribed ART.

In this study of HIV-infected adults with past or current alcohol problems, depressive symptoms predicted subsequent ART discontinuation. Their recognition and treatment may be important in maintaining ART and its associated clinical benefits in this patient population.

Acknowledgements

The authors appreciate the contributions of the staff researchers on the project and data management assistance including Vincent Faber at DM-STAT, Inc., Medford, MA. Support for this study came from the following grants from the National Institute on 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-AA15674 (Impact of Alcohol Use on HIV Infection—In the United States and Russia). This research was conducted in part in the General Clinical Research Center at Boston University School of Medicine, USPHS Grant M01 RR00533. Support for Theresa W. Kim came from the National Institute on Drug Abuse of the NIH: R25-DA13582 (Clinical Addiction Research and Education [CARE] Program). Anita Palepu is supported by a Michael Smith Foundation for Health Research Senior Scholar Award.

References

Ahdieh, G.L., Silverberg, M.J., Palacio, H., Minkoff, H., Anastos, K., Young, M.A., et al. (2001). Discontinuation of potent antiretroviral therapy: Predictive value of, and impact on, CD4 cell counts and HIV RNA levels. *AIDS*, *15*, 2101–2108.

Ahdieh, G.L., Tarwater, P.M., Schneider, M.F., Anastos, K., Cohen, M., Khalsa, A., et al. (2005). Factors and temporal trends associated with highly active antiretroviral therapy discontinuation in the Women's Interagency HIV Study. *Journal of Acquired Immune Deficiency Syndromes*, *38*, 500–503.

Asch, S.M., Kilbourne, A.M., Gifford, A.L., Burnam, M.A., Turner, B., Shapiro, M.F., et al. (2003). Underdiagnosis of depression in HIV: Who are we missing? *Journal of General Internal Medicine*, *18*, 450–460.

Bing, E.G., Burnam, M.A., Longshore, D., Fleishman, J.A., Sherbourne, C.D., London, A.S., et al. (2001). Psychiatric disorders and drug use among human immunodeficiency virus-infected adults in the US. *Archives of General Psychiatry*, *58*, 721–728.

Brady, K.T., & Sinha, R. (2005). Co-occurring mental and substance use disorders: The neurobiological effects of chronic stress. *American Journal of Psychiatry*, *162*, 1483–1493.

Braithwaite, R.S., McGinnis, K.A., Conigliaro, J., Maisto, S.A., Crystal, S., Day, N., et al. (2005). A temporal and dose-response association between alcohol consumption and medication adherence among veterans in care. *Alcoholism: Clinical and Experimental Research*, *29*, 1190–1197.

Buchsbaum, D.G., Buchanan, R.G., Centor, R.M., Schnoll, S.H., & Lawton, M.J. (1991). Screening for alcohol abuse using CAGE scores and likelihood ratios. *Annals of Internal Medicine*, *115*, 774–777.

Chen, R.Y., Westfall, A.O., Mugavero, M.J., Cloud, G.A., Raper, J.L., Chatham, A.G., et al. (2003). Duration of highly active antiretroviral therapy regimens. *Clinical Infectious Disease*, *37*, 714–722.

Compton, W.M., Cottler, L.B., Jacobs, J.L., Ben-Abdallah, A., & Spitznagel, E.L. (2003). The role of psychiatric disorders in predicting drug dependence treatment outcomes. *American Journal Psychiatry*, *160*, 890–895.

Conigliaro, J., Gordon, A.J., McGinnis, K.A., Rabeneck, L., & Justice, A.C. (2003). How harmful is hazardous alcohol use and abuse in HIV infection: Do healthcare providers know who is at risk? *Journal of Acquired Immune Deficiency Syndromes*, *33*, 521–525.

Conway, K.P., Compton, W., Stinson, F.S., & Grant, B.F. (2006). Lifetime comorbidity of DSM-IV mood and anxiety disorders and specific drug use disorders: Results from the National Epidemiologic Survey on alcohol and related conditions. *Journal of Clinical Psychiatry, 67*, 247–257.

Cook, J.A., Grey, D., Burke, J., Cohen, M.H., Gurtman, A.C., Richardson, J.L., et al. (2004). Depressive symptoms and AIDS-related mortality among a multisite cohort of HIV-positive women. *American Journal of Public Health, 94*, 1133–1140.

Cook, J.A., Grey, D., Burke-Miller, J., Cohen, M.H., Anastos, K., Ghandi, M., et al. (2006). Effects of treated and untreated depressive symptoms on highly active antiretroviral therapy use in a US multi-site cohort of HIV-positive women. *AIDS Care, 18*, 93–100.

D'Arminio, M.A., Cozzi-Lepri, A., Phillips, A., De, L.A., Murri, R., Mussini, C., et al. (2005). Interruption of highly active antiretroviral therapy in HIV clinical practice: Results from the Italian cohort of antiretroviral-naive patients. *Journal of Acquired Immune Deficiency Syndromes, 38*, 407–416.

Fairfield, K.M., Libman, H., Davis, R.B., Eisenberg, D., Beckett, A., & Phillips, R.S. (2001). Brief communication. Detecting depression: Providing high quality primary care for HIV-infected patients. *American Journal of Medical Quality, 16*, 71–74.

Galvan, F.H., Bing, E.G., Fleishman, J.A., London, A.S., Caetano, R., Burnam, M.A., et al. (2002). The prevalence of alcohol consumption and heavy drinking among people with HIV in the US: Results from the HIV Cost and Services Utilization Study. *Journal of Studies on Alcohol, 63*, 179–186.

Galvan, F.H., Burnam, M.A., & Bing, E.G. (2003). Co-occurring psychiatric symptoms and drug dependence or heavy drinking among HIV-positive people. *Journal of Psychoactive Drugs, 35*(Suppl. 1), S153–S160.

Golub, E.T., Latka, M., Hagan, H., Havens, J.R., Hudson, S.M., Kapadia, F., et al. (2004). Screening for depressive symptoms among HCV-infected injection drug users: Examination of the utility of the CES-D and the Beck Depression Inventory. *Journal of Urban Health, 81*, 278–290.

Grant, B.F., Stinson, F.S., Dawson, D.A., Chou, S.P., Dufour, M.C., Compton, W., et al. (2004). 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. *Archives of General Psychiatry, 61*, 807–816.

Hasin, D.S., & Grant, B.F. (2002). Major depression in 6,050 former drinkers: Association with past alcohol dependence. *Archives of General Psychiatry, 59*, 794–800.

Hays, R.D., Cunningham, W.E., Sherbourne, C.D., Wilson, I.B., Wu, A.W., Cleary, P., et al. (2000). Health-related quality of life in patients with human immunodeficiency virus infection in the US: Results from the HIV Cost and Services Utilization Study. *American Journal of Medicine, 108*, 714–722.

Heath, K.V., Singer, J., O'Shaughnessy, M.V., Montaner, J.S., & Hogg, R.S. (2002). Intentional nonadherence due to adverse symptoms associated with antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes, 31*, 211.

Hogg, R.S., Heath, K., Bangsberg, D., Yip, B., Press, N., O'Shaughnessy, M.V., et al. (2002). Intermittent use of triple-combination therapy is predictive of mortality at baseline and after one year of follow-up. *AIDS, 16*, 1051–1058.

Ickovics, J.R., Hamburger, M.E., Vlahov, D., Schoenbaum, E.E., Schuman, P., Boland, R.J., et al. (2001). Mortality, CD4 cell count decline and depressive symptoms among HIV-seropositive women: Longitudinal analysis from the HIV Epidemiology Research Study. *Journal of American Medical Association, 285*, 1466–1474.

Justice, A.C., Holmes, W., Gifford, A.L., Rabeneck, L., Zackin, R., Sinclair, G., et al. (2001). Development and validation of a self-completed HIV symptom index. *Journal of Clinical Epidemiology, 54*(Suppl. 1), S77–S90.

Kerr, T., Marshall, A., Walsh, J., Palepu, A., Tyndall, M., Montaner, J., et al. (2005). Determinants of HAART discontinuation among injection drug users. *AIDS Care, 17*, 539–549.

Kerr, T., Palepu, A., Barness, G., Walsh, J., Hogg, R., Montaner, J., et al. (2004). Psychosocial determinants of adherence to highly active antiretroviral therapy among injection drug users in Vancouver. *Antiviral Therapy, 9*, 407–414.

Kertesz, S.G., Larson, M.J., Horton, N.J., Winter, M., Saitz, R., & Samet, J.H. (2005). Homeless chronicity and health-related quality of life trajectories among adults with addictions. *Medical Care, 43*, 574–585.

Kessler, R.C., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H.U. (1998). The World Health Organization composite international diagnostic interview short form (CIDI-SF). *International Journal of Methods in Psychiatric Research, 7*, 171–185.

Kilbourne, A.M., Justice, A.C., Rollman, B.L., McGinnis, K.A., Rabeneck, L., Weissman, S., et al. (2002). Clinical importance of HIV and depressive symptoms among veterans with HIV infection. *Journal of General Internal Medicine, 17*, 512–520.

Li, X., Margolick, J.B., Conover, C.S., Badri, S., Riddler, S.A., Witt, M.D., et al. (2005). Interruption and discontinuation of highly active antiretroviral therapy in the multicenter AIDS cohort study. *Journal of Acquired Immune Deficiency Syndromes, 38*, 320–328.

Lucas, G.M., Griswold, M., Gebo, K.A., Keruly, J., Chaisson, R.E., & Moore, R.D. (2006). Illicit drug use and HIV-1 disease progression: A longitudinal study in the era of highly active antiretroviral therapy. *American Journal of Epidemiology, 163*, 412–420.

McLellan, A.T., Lewis, D.C., O'Brien, C.P., & Kleber, H.D. (2000). Drug dependence, a chronic medical illness: Implications for treatment, insurance and outcomes evaluation. *Journal of American Medical Association, 284*, 1689–1695.

Miguez, M.J., Shor-Posner, G., Morales, G., Rodriguez, A., & Burbano, X. (2003). HIV treatment in drug abusers: Impact of alcohol use. *Addiction Biology, 8*, 33–37.

Mocroft, A., Phillips, A.N., Soriano, V., Rockstroh, J., Blaxhult, A., Katlama, C., et al. (2005). Reasons for stopping antiretrovirals used in an initial highly active antiretroviral regimen: Increased incidence of stopping due to toxicity or patient/physician choice in patients with hepatitis C co-infection. *AIDS Research and Human Retroviruses, 21*, 527–536.

Moss, A.R., Hahn, J.A., Perry, S., Charlebois, E.D., Guzman, D., Clark, R.A., et al. (2004). Adherence to highly active antiretroviral therapy in the homeless population in San Francisco: A prospective study. *Clinical Infectious Disease, 39*, 1190–1198.

Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Riley, E.D., Bangsberg, D.R., Guzman, D., Perry, S., & Moss, A.R. (2005). Antiretroviral therapy, hepatitis C virus and AIDS mortality among San Francisco's homeless and marginally housed. *Journal of Acquired Immune Deficiency Syndromes, 38*, 191–195.

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*, 1069–1077.

Samet, J.H., Horton, N.J., Meli, S., Freedberg, K.A., & Palepu, A. (2004). Alcohol consumption and antiretroviral adherence among HIV-infected persons with alcohol problems. *Alcoholism: Clinical and Experimental Research, 28*, 572–577.

Samet, J.H., Horton, N.J., Traphagen, E.T., Lyon, S.M., & Freedberg, K.A. (2003). Alcohol consumption and HIV disease progression: Are they related? *Alcoholism: Clinical and Experimental Research*, *27*, 862–867.

Samet, J.H., Libman, H., Labelle, C., Steger, K., Lewis, R., Craven, D.E., et al. (1995). A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. *Archives of Internal Medicine*, *155*, 1629–1633.

Samet, J.H., Phillips, S.J., Horton, N.J., Traphagen, E.T., & Freedberg, K.A. (2004). Detecting alcohol problems in HIV-infected patients: Use of the CAGE questionnaire. *AIDS Research and Human Retroviruses*, *20*, 151–155.

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

Sobell, L.C., & Sobell, M.B. (1992). Timeline follow-back: a technique for assessing self-reported alcohol consumption. In R.Z. Litten, & J.P. Allen (Eds.), *Measuring Alcohol Consumption: Psychosocial and Biochemical Methods* (pp. 41–69). Totowa, NJ: Humana Press, Inc.

Yuan, Y., L'italien, G., Mukherjee, J., & Iloeje, U.H. (2006). Determinants of discontinuation of initial highly active antiretroviral therapy regimens in a US HIV-infected patient cohort. *HIV Medicine*, *7*, 156–162.

Zeger, S.L., Liang, K.Y., & Albert, P.S. (1988). Models for longitudinal data: A generalised estimating equation approach. *Biometrics*, *44*, 1049–1060.

DOI: 10.1111/j.1475-6773.2006.00630.x

Substance Abuse

Primary Care Quality and Addiction Severity: A Prospective Cohort Study

Theresa W. Kim, Jeffrey H. Samet, Debbie M. Cheng, Michael R. Winter, Dana Gelb Safran, and Richard Saitz

Background. Alcohol and drug use disorders are chronic diseases that require ongoing management of physical, psychiatric, and social consequences. While specific addiction-focused interventions in primary care are efficacious, the influence of overall primary care quality (PCQ) on addiction outcomes has not been studied. The aim of this study was to prospectively examine if higher PCQ is associated with lower addiction severity among patients with substance use disorders.

Study Population. Subjects with alcohol, cocaine, and/or heroin use disorders who initiated primary care after being discharged from an urban residential detoxification program.

Measurements. We used the Primary Care Assessment Survey (PCAS), a well-validated, patient-completed survey that measures defining attributes of primary care named by the Institute of Medicine. Nine summary scales cover two broad areas of PCQ: the patient–physician relationship (communication, interpersonal treatment, thoroughness of the physical exam, whole-person knowledge, preventive counseling, and trust) and structural/organizational features of care (organizational access, financial access, and visit-based continuity). Each of the three addiction outcomes (alcohol addiction severity (ASI-alc), drug addiction severity (ASI-drug), and any drug or heavy alcohol use) were derived from the Addiction Severity Index and assessed 6–18 months after PCAS administration. Separate longitudinal regression models included a single PCAS scale as the main predictor variable as well as variables known to be associated with addiction outcomes.

Main Results. Eight of the nine PCAS scales were associated with lower alcohol addiction severity at follow-up ($p \leq .05$). Two measures of relationship quality (communication and whole- person knowledge of the patient) were associated with the largest decreases in ASI-alc (-0.06). More whole-person knowledge, organizational access, and visit-based continuity predicted lower drug addiction severity (ASI-drug: -0.02). Two PCAS scales (trust and whole-person knowledge of the patient) were associated with lower likelihood of subsequent substance use (adjusted odds ratio, [AOR] = 0.76, 95 percent confidence interval [95% CI] = 0.60, 0.96 and AOR = 0.66, 95 percent CI = 0.52, 0.85, respectively).

Conclusion. Core features of PCQ, particularly those reflecting the quality of the physician–patient relationship, were associated with positive addiction outcomes. Our findings suggest that the provision of patient-centered, comprehensive care from

a primary care clinician may be an important treatment component for substance use disorders.

Key Words. Substance abuse, primary care, quality of care, physician–patient relationship

---

Primary care is being asked to expand its role in the identification and management of addictive disorders (National Council for Community Behaviorial Healthcare 2003; New Freedom Commission on Mental Health 2003). Training sessions for primary care physicians and clinical reminders have been developed to increase the use of efficacious counseling tools (Saitz et al. 2003; Saitz, Sullivan, and Samet 2000) and practical screening instruments to identify addiction disorders (Maisto and Saitz 2003). Pharmacological therapies are now available in primary care for the treatment of opioid dependence (e.g., buprenorphine) (Fiellin and O'Connor 2002) and alcohol use disorders (e.g., naltrexone and acamprosate). Various forms of integration between primary care and addictions treatment have been demonstrated to be associated with better addiction outcomes (Willenbring and Olson 1999; Weisner et al. 2001; Friedmann et al. 2003).

While there is substantial literature promoting greater access to primary care for individuals with addictions, little is known about whether the quality of primary care should be considered. In other chronic disorders such as diabetes mellitus (Campbell, Roland, and Wilkin 2001), depression (O'Malley, Forrest and Miranda 2003), and tobacco use (Safran et al. 1998), primary care quality (PCQ) is an important determinant of outcomes. The quality of primary care delivered to individuals with substance use disorders may be important for addiction outcomes as well.

How might higher quality primary care lead to better addiction outcomes? The Institute of Medicine (IOM) and others (Institute of Medicine

---

Address correspondence to Theresa W. Kim, M.D., Clinical Addiction Research and Education (CARE) Unit, Section General Internal Medicine, 91 E. Concord Street, Suite 200, Boston University Medical Center, Boston, MA 02118. Jeffrey H. Samet, M.A., M.D., M.P.H., Debbie M.Cheng, Sc.D., and Richard Saitz, M.D., M.P.H., are with the Clinical Addiction Research and Education (CARE) Unit, Boston University Medical Center, Boston, MA. Dr. Samet is also with Department of Social and Behavioral Sciences, Boston University School of Public Health. Dr. Cheng is also with the Department of Biostatistics, Boston University School of Public Health. Michael R. Winter, M.P.H., is with the Data Coordinating Center, Boston, MA. Dana Gelb Safran, Sc.D., is with The Health Institute, Institute for Clinical Research and Health Policy Studies, Boston, MA.

1996; Starfield et al. 1998; Safran 2003; Kroenke 2004) define primary care as "sustained partnerships" with essential attributes of comprehensiveness, coordination, and accessibility. Alcohol and drug problems, often unrecognized in medical encounters (Saitz et al. 1997; Weisner and Matzger 2003), may be missed less frequently in comprehensive and supportive primary care relationships. Adherence to treatment recommendations for addictive disorders may be higher in the presence of such relationships between patients and their primary care physicians. As alcohol and drug dependence are often chronic illnesses with relapsing and remitting courses (McLellan et al. 2000), effective chronic disease management strategies to enhance treatment adherence and monitor disease progression may ameliorate addiction severity. Additionally, coordination of addiction, medical, psychiatric, and social services care by primary care physicians may be especially valuable for patients with comorbid physical and/or psychiatric illnesses (Johnson et al. 1995; Mertens et al. 2003; De, Samet, and Saitz 2004).

We used data from the Health Evaluation and Linkage to Primary care (HELP) study, a randomized trial of an intervention to link adults entering an urban detoxification unit to primary care. Previous analyses showed that randomization to the intervention group was associated with greater likelihood of initiating primary care but was not associated with better addiction outcomes (Samet et al. 2003). An additional analysis demonstrated that attending at least two primary care visits, regardless of randomization arm, was associated with lower addiction severity and risk of subsequent substance use (Saitz et al. 2005). The current analysis takes this line of examination a step further and asks, among individuals with addictions who attend primary care, is the quality of primary care associated with addiction outcomes, and if so, which attributes of PCQ are important?

METHODS

Study Design

This is a prospective cohort study of patients with substance use disorders who initiated primary care during the HELP study (Samet et al. 2003). People eligible for the HELP study were adult inpatients of a residential detoxification unit who spoke Spanish or English, reported alcohol, heroin, or cocaine as their first or second drug of choice, and either resided in proximity to the primary care clinic to which they would be referred, or were homeless. Patients were excluded from the parent study if they intended to continue

existing primary care relationships, suffered significant dementia, had specific plans to leave the Boston area that would prevent research participation, could not provide contact information for tracking purposes, or were pregnant. All participants were randomly assigned to receive either standard medical care referral by clinical addictions staff on an as needed basis (usual care) or enhanced referral (the trial intervention) to primary care in the Boston area and surrounding communities. Enhanced referral included a multidisciplinary assessment (conducted in the detoxification unit at study entry), referral to a particular physician based upon patient preference, letter of introduction to the primary care physician, and follow-up phone calls to promote adherence to the initial primary care appointment (which occurred after discharge from the detoxification unit).

After the baseline research assessment at the detoxification unit, biannual research interviews were conducted over the 24-month study period. Except for demographics, all assessments, including measures of addiction severity and health care utilization, were repeated at each interview. Alcohol breath tests were performed at follow-up interviews to encourage honest reporting (Welte et al. 1998).

The current study included HELP participants who initiated primary care, *regardless* of randomization group, and then completed at least one follow-up interview after initiating primary care. In total, participants in the current study completed at least three separate research interviews: (1) an assessment at the detoxification unit where study recruitment took place (i.e., the "baseline" interview); (2) a subsequent interview in which the participant reported on the quality of their primary care relationship (the "PCQ assessment" interview); and (3) a follow-up interview in which addiction outcomes were assessed (the "follow-up addiction" interview).

The Institutional Review Board of Boston University Medical Center approved this study. Additional privacy protection was secured by the issuance of a Certificate of Confidentiality by the Department of Health and Human Services.

Main Independent Variable

The Primary Care Assessment Survey (PCAS; Safran et al. 1998) is a well-validated, patient-completed survey that measures defining attributes of primary care named by the IOM and others (Institute of Medicine 1996; Starfield et al. 1998; Safran 2003; Kroenke 2004). The PCAS is a well-validated measure of PCQ that has been used to monitor the performance of primary care

Table 1: Summary of Item Content for Primary Care Assessment Survey Scales

Defining Characteristic	*Description*
Physician–patient interaction	
Communication	Thoroughness of primary physician's questions about symptoms, attention to what patient says, clarity of explanations and instructions, and advice and help in making decisions about care
Interpersonal treatment	Primary physician's patience, friendliness, caring, respect, and time spent with patient
Thoroughness of physical exam	Primary physician's physical examination thoroughness
Whole-person knowledge	Primary physician's knowledge of patient's medical history; responsibilities at work, home, school; health concerns, values, and beliefs
Preventive counseling	Whether physician has discussed the following with patient: smoking, alcohol use, seatbelt use, diet, exercise, stress, safe sex
Trust	Assessment of physician's integrity, competence, and role as patient's agent
Structural feature of care	
Organizational access	Ability to get through to physician's office by telephone, to get a medical appointment when sick, to obtain information by telephone, punctuality of appointments, convenience of office location, and convenience of office hours
Financial access	Assessment of amount of money patient pays for physician visits, medication, and other prescribed treatments
Visit-based continuity	How often patient sees primary care physician (not an assistant or partner) for routine check-ups and for appointments when sick

delivered by large health plans and delivery systems over time (Murphy et al. 2001; Safran et al. 2002). Higher PCQ, as measured by the PCAS, has been linked to important patient outcomes such as adherence to physician's advice and improved health status (Safran et al. 1998).

The PCAS measures PCQ in the context of a specific physician–patient primary care relationship and references the entirety of that relationship (Safran et al. 1998). Nine summary scales (Table 1) cover two broad areas of PCQ: the patient–physician relationship (communication, interpersonal treatment, thoroughness of the physical exam, whole-person knowledge, preventive counseling, and trust) and structural/organizational features of care (organizational access, financial access, and visit-based continuity). Each scale has a range from 0 to 100 with higher scores indicating more of the underlying attribute.

Dependent Variable

Each of the three outcomes were derived from the Addiction Severity Index (McLellan et al. 1992): (1) alcohol addiction severity using the alcohol composite score (ASI-alc), (2) drug addiction severity using the drug composite score (ASI-drug), and (3) any drug or heavy alcohol use (more than three drinks in a day) in the past 30 days. The latter variable will be referred to as "any substance use." ASI-alc and ASI-drug composite scores range from 0 to 1, with higher scores indicating greater severity.

Statistical Analysis

We examined the relationship between higher PCQ and addiction outcomes by fitting separate multivariable longitudinal regression models for each outcome. Each regression model included a single PCAS scale as the main predictor variable. The unit of analysis was each study interview (i.e., observation). All study participants ($n = 183$) contributed to the analyses of any substance use ($n = 355$ observations). Only participants with alcohol as a first or second substance of choice ($n = 117$) contributed to the alcohol severity analyses ($n = 228$ observations). Likewise, only subjects with heroin or cocaine as a first or second drug of choice ($n = 145$) contributed to the drug severity analyses ($n = 284$ observations).

The longitudinal regression models accounted for the correlation from using repeated observations on the same subject. Continuous outcomes (ASI-alc and ASI-drug) were analyzed using a general linear model for correlated data with an unstructured correlation matrix (Liang and Zeger 1986). For the dichotomous outcome of any substance use, generalized estimating equations (GEE) logistic regression models with empirical standard errors were used to analyze the data (Zeger, Liang, and Albert 1988). Each regression model also included the following covariates: age, gender, race/ethnicity, education less than 12 years (yes versus no), homelessness (any night in a shelter or street in the past 6 months); (Kertesz et al. 2005), health insurance, addiction severity (baseline alcohol and drug ASI scores at HELP study entry), HELP randomization group, time of addiction outcome assessment after HELP study entry (12, 18, or 24 months), and interval between assessments of PCQ and addiction outcomes (6, 12, or 18 months). No adjustments were made for multiple comparisons due to the exploratory nature of the analyses.

In order to assess whether PCQ was a predictor of subsequent addiction severity and substance use, models included addiction outcomes measured at the study interview *after* the PCQ assessment interview. As research interviews

were scheduled approximately every 6 months over the 24-month study period, the interval between PCAS administration and the addiction assessment was 6 months if consecutive follow-up interviews were completed. However, if a participant missed a scheduled follow-up interview, then the substance abuse assessment at the next available interview was used (12 or 18 months after the assessment of PCQ). The interval between the PCQ assessment and addiction outcomes was 6 months for 269 (75.8 percent) of the 355 available observations; 12 months for 60 (16.9 percent); and 18 months for 26 (7.3 percent). A term for the interaction between each PCAS scale and the duration of time between PCQ assessment and addiction outcome assessment was included in initial models to assess whether the effect of PCQ depended upon the length of time between PCQ and addiction assessments. The interactions were not significant ($p > .1$) and therefore not included in the final models.

Longitudinal regression results for addiction severity are presented as the mean difference in alcohol or drug severity associated with a standard deviation increase in PCAS score. Similarly, regression results for substance use are presented as the risk of substance use associated with a standard deviation increase in PCAS score. Reported p-values were two-tailed and considered statistically significant if $< .05$. All analyses were completed using *SAS/STAT* software, version 8.2 (SAS/STAT 1999).

RESULTS

Study Subjects

This study's analytic sample was derived from the HELP cohort. Of the 470 subjects in the HELP cohort, two died before follow-up and 400/468 (85 percent) completed at least one interview during the two-year follow-up period. Of the 400 subjects with follow-up, 253 (63 percent) reported initiating primary medical care after being discharged from the detoxification unit. As previously reported, women were more likely to link with primary care as well as those with recent episodic medical visits, family support for abstinence, and health insurance (Saitz et al. 2004). Recent incarceration decreased the likelihood of linkage. Ethnicity, recent addiction or mental health treatment utilization, addiction severity, health status, substance problem recognition, and perceived need for medical care did not affect linkage.

Among the 253 subjects who initiated primary care, 183 (72 percent) returned for a third interview (the "addiction interview") comprising our study sample. No significant differences were found between the 70 participants

(70/253, 28 percent) who initiated primary care but were unavailable for the addiction interview and the study cohort in terms of age, gender, race/ethnicity, housing status, education level, alcohol abuse severity, or drug abuse severity.

The baseline sociodemographic and health characteristics of the study sample are displayed in Table 2. A majority of study participants were male

Table 2: Demographic and Clinical Characteristics of Participants ($N = 183$) at Study Entry

	N (%)
Male	124 (68)
Race/ethnicity	
Black	99 (54)
White	51 (28)
Hispanic	20 (11)
Less than 12 years education	55 (30)
Homeless	89 (49)
Uninsured	110 (60)
	Mean (SD)
Age	37 (8)
Physical health related quality of life*	46 (11)
Mental health related quality of life†	32 (12)
Alcohol addiction severity‡	0.45 (0.35)
Drug addiction severity§	0.26 (0.14)
Primary Care Assessment Survey¶	
Physician–patient interaction	
Communication	76 (20)
Interpersonal treatment	74 (21)
Whole person knowledge	54 (24)
Thoroughness of physical exam	74 (23)
Trust	73 (17)
Preventive counseling	56 (30)
Structural features of care	
Organization access	62 (21)
Financial access	81 (24)
Visit-based continuity	83 (24)

*Assessed with the Short-Form Health Survey (SF-36) Physical Component Summary (PCS), range 0–100.

†Assessed with the SF-36, Mental Component Summary (MCS), range 0–100.

‡Addiction Severity Index (ASI) alcohol composite score range 0–1 with higher scores indicating worse severity.

§ASI drug composite score range 0–1 with higher scores indicating worse severity.

¶Primary Care Assessment Survey (PCAS) scales range from 0 to 100 points with higher scores indicating more of the underlying attribute. PCAS results from the first interview that a participant reported having a primary care provider are presented.

(68 percent), nonwhite race/ethnicity (72 percent), and uninsured (60 percent). A substantial minority (30 percent) did not graduate from high school and about half were homeless. At baseline, the mean ASI-alc and mean ASI-drug scores were 0.45 and 0.26, respectively. These scores are similar to those of individuals entering the public treatment system in Massachusetts (Smith and Larson 2003) but more severe than those in a clinical addiction treatment sample in an HMO in California (Weisner, McLellan, and Hunkeler 2000). The mean Short-Form Health Survey (SF-36) Physical Component Summary (PCS) and Mental Component Summary (MCS; Ware 1994) scores were 46 and 32 respectively (the mean MCS and PCS score for the U.S. population is 50; 89 percent of adults with MCS scores of 30–34 screens positive for depression). The lowest ranked measure of PCQ was whole-person knowledge and the highest was visit-based continuity.

Multivariable Regression Results

The 183 subjects contributed 355 observations to the longitudinal regression models. Table 3 presents the results of regression analyses examining the nine PCAS scales as independent predictors of alcohol and drug addiction severity. We were able to detect significant associations between PCQ and addiction outcomes most consistently in the alcohol addiction severity analyses. Higher

Table 3: Adjusted Mean Change in Alcohol and Drug Addiction Severity Associated with Higher Primary Care Quality

Primary Care Quality Attribute	*Alcohol Addiction Severity Change**	*p-Value*	*Drug Addiction Severity Change§*	*p-Value*
Physician–patient interaction				
Communication	**−0.06**	**.001**	−0.005	.51
Interpersonal treatment	**−0.06**	**.004**	−0.006	.51
Thoroughness of physical exam	**−0.05**	**.005**	−0.0006	.93
Whole-person knowledge	**−0.06**	**.001**	**−0.02**	**.01**
Preventive counseling	−0.03	.06	−0.003	.73
Trust	**−0.04**	**.02**	−0.007	.40
Structural features of care				
Organizational access	**−0.06**	**.002**	**−0.01**	**.05**
Financial access	**−0.05**	**.01**	−0.0002	.97
Visit-based continuity	**−0.03**	**.05**	**−0.01**	**.03**

*Subjects with alcohol as a first or second substance of choice ($n = 115$) contributed to alcohol addiction severity analyses ($n = 224$ observations).

§Subjects with heroin or cocaine as a first or second drug of choice ($n = 145$) contributed to drug addiction severity analyses ($n = 280$ observations).

Statistically significant associations ($p < .05$) in bold.

scores on 8 of the 9 PCAS scales representing higher PCQ were associated with lower alcohol addiction severity at follow-up. Two measures of relationship quality (communication and whole person knowledge of the patient) were associated with the largest decrease in alcohol severity (ASI-alc -0.06); visit-based continuity was associated with the smallest decrease (-0.03).

The results of drug severity (Table 3) and any substance use analyses (Table 4) were mixed. Higher scores on three PCAS scale scores (whole-person knowledge, organizational access, and visit-based continuity) were associated with lower drug addiction severity: adjusted mean decrease in ASI-drug = -0.01 to -0.02 for each of the three PCAS scales. The other six PCAS scale scores were not significantly associated with drug addiction severity.

More whole-person knowledge and trust were significantly associated with lower odds of substance use (Table 4) (adjusted odds ratio, [AOR] = 0.65, 95 percent confidence interval [95% CI] = 0.51, 0.83 and AOR = 0.76, 95 percent CI = 0.59, 0.97, respectively). Other PCAS scales did not significantly predict substance use.

To address the possibility of higher order effects of the HELP intervention, we tested for an interaction between HELP randomization group and

Table 4: Adjusted Odds Ratio of Any Substance Use at Follow-up Associated with Higher Primary Care Quality

Primary Care Quality Attribute	*Adjusted Odds Ratio (95% Confidence Interval) of Any Substance Use**
Physician–patient interaction	
Communication	0.95 (0.73, 1.25)
Interpersonal treatment	0.80 (0.63, 1.02)
Thoroughness of physical exam	0.99 (0.76, 1.30)
Whole person knowledge	**0.65 (0.51, 0.83)**
Preventive counseling	0.91 (0.74, 1.13)
Trust	**0.76 (0.59, 0.97)**
Structural features of care	
Organizational access	0.86 (0.68, 1.09)
Financial access	0.83 (0.66, 1.03)
Visit-based continuity	0.94 (0.76, 1.17)

*Substance use is defined as any drug use (cocaine or heroin) or any alcohol intoxication (>3 drinks on any occasion) in past the 30 days. Substance use was assessed 6 to 18 months after the primary care quality evaluation.

Results are reported as the adjusted odds ratio associated with a standard deviation increase in PCAS scale score (indicating more of the underlying attribute). Statistically significant associations ($p<.05$) in bold. All study subjects ($n=183$) contributed to analyses of substance use ($n=355$ observations).

PCAS scale in each of the 27 main models. The interaction was only significant in the model examining organizational access and alcohol severity.

We explored whether primary care utilization affected the relationship between PCQ and addictions. In the main analyses, we did not adjust for primary care utilization because of the relatively low variability in the number of primary care visits. For 75 percent of the observations used in the analyses, patients reported between one and three primary care visits in the previous 6 months. Variability of the length of primary care relationships was also relatively narrow due to the fact that all study subjects initiated primary care during the study period. However, as primary care utilization is an important consideration, we included a covariate for the number of primary care visits (self-report) to the main models. Including this covariate did not alter our findings.

We also performed secondary analyses with covariates for exposure to substance abuse treatment (yes/no), AA participation (yes/no), and any mental health visit (yes/no), which again did not change the direction of the estimates or diminish the statistical significance of the results. However, in three of the models, the *p*-values increased: in the model examining visit-based continuity and alcohol addiction severity (*p*-value increased from .05 to .06), and in the models examining visit-based continuity and organizational access predicting drug addiction severity (*p*-values increased to .07). As the parameter estimates remained unchanged, higher *p*-values may have resulted from the addition of three more covariates to the models and diminished statistical power. Overall, the relationship between the quality of primary care and addiction outcomes did not appear to be mediated by these utilization variables.

DISCUSSION

In this cohort of adults recruited from a residential detoxification unit and prospectively assessed over a 24-month study period, higher quality primary care across multiple domains was associated with lower addiction severity and odds of substance use. These associations did not appear to be mediated by variables previously identified in the literature to be related to health care quality (i.e., health insurance, gender, race/ethnicity), primary care utilization, or baseline addiction severity. Two key PCQ attributes reflecting the quality of primary care relationships—physicians' whole-person knowledge of the patient and patient trust—were significant predictors of lower alcohol severity and lower risk of substance use.

Our findings align with those from previous studies that have demonstrated the importance of the physician–patient relationship in patient acceptance and receipt of preventive services (O'Malley et al. 2004), cancer screening measures (Safran et al. 1998; O'Malley and Forrest 2002), and depression treatment (O'Malley, Forrest, and Miranda 2003). Among HIV-infected individuals, trust and whole-person care has been linked with higher adherence to HIV medications (Schneider et al. 2004) as well as better physical and mental health functioning (Preau et al. 2004). Although the importance of interpersonal aspects of primary care has been reported in various populations, to our knowledge, this is the first study to examine their importance for patients with alcohol and drug use disorders.

In this study, whole-person knowledge emerged as the most consistent predictor of better addiction outcomes. Individualizing clinical decisions based upon the "contextual knowledge" of a patient's beliefs and values as well as responsibilities at work, home, or school (Weiner 2004) may have particular importance for individuals with addiction problems. It is notable that whole-person knowledge was one of the lowest ranked quality measure in this study, consistent with studies in other populations (Murphy et al. 2001). This suggests that careful attention to this aspect of primary care could be important when designing or evaluating programs that integrate addiction treatment with primary care.

It is important to note that the PCAS is not a measure of patient satisfaction but rather a well-validated measure of PCQ. The PCAS measures primary care in terms of a standard derived from the IOM's definition of primary care and may or may not relate to individual patient satisfaction. Other studies have measured PCQ in terms of delivery of specific services such as influenza vaccination or cervical screening. In light of our postulated mechanisms of higher quality primary care improving addictions, using the PCAS may have more relevance than using process of care measures.

Using this validated instrument to measure primary care, single standard deviation increases in PCQ scores were associated with a lower risk of subsequent substance use (i.e., 24–35 percent decrease in odds of any use) and moderate decreases in alcohol addiction severity (0.03–0.06 on scales ranging from 0 to 1). The reduction in ASI-alc scores demonstrated in this study are similar to the effect of two primary care visits versus none (ASI-alc = -0.04) found in a previous study (Saitz et al. 2005) but less than the effect of more intensive interventions such as case-managed residential care (Conrad et al. 1998) or work therapy for homeless veterans (ASI-alc = -0.16) (Kashner et al. 2002).

The clinical significance of the addiction differences in this study should be viewed in light of the fact that: (1) The magnitude of PCQ may be underestimated in this cohort with relatively new primary care relationships. Greater cumulative effects of higher quality primary care might be observed over a longer period of time. (2) The ability to robustly predict addiction outcomes with one exposure is rare. Since changes in addiction outcomes generally result from cumulative changes in the environment, it is not surprising that our findings were modest compared to more intensive interventions. We did not expect to find large differences in addiction severity among these patients, all of whom had initiated substance abuse treatment and were receiving primary care. As there is renewed interest in providing primary care to individuals with addictions, ensuring that core features of primary care exist in their primary care relationships might augment the postulated addiction benefits by the estimates found in this study.

While all of the effects of PCQ were in the hypothesized direction, we were unable to detect associations between attributes of PCQ and drug addiction severity. Despite adjusted analyses, the impact of the quality of primary care relationships initiated over a relatively short period of time may have been difficult to isolate in this cohort with significant homelessness and poverty. As drug dependence is more difficult to treat than alcohol dependence without adequate pharmacotherapy, the effect of primary care on addiction outcomes may have been overwhelmed without first reducing system-level barriers to opioid agonist therapy. Additionally, the prevalence of comorbid psychiatric conditions is generally higher in samples with drug addiction than those with alcohol addiction (Grant et al. 2004). Thus, detecting the effect of differences in PCQ may have been more difficult without addressing psychiatric comorbidities in patients with drug use disorders.

Several limitations should be considered when interpreting the study findings. The major limitation is the potential confounding resulting from this study's observational design. We cannot exclude the possibility that addiction severity was a determinant of the quality of primary care rather than PCQ predicting addiction outcomes. We attempted to minimize this possibility by: (1) using a lagged analysis, i.e., assessing the addiction severity of an individual at least 6 months after the receipt of primary care services; (2) including well-established determinants of PCQ, specifically, race/ethnicity, gender, and health insurance in the analyses; and (3) adjusting for baseline addiction severity. Still, unmeasured factors may have influenced which participants received better primary care and mediated addiction outcomes. A propensity

analysis might have better addressed this limitation, however, this study's sample size could not support this approach.

Another limitation relates to the fact these data were primarily collected for a randomized trial of an intervention to link adults to primary care. It is unlikely, however, that the HELP intervention confounded our results since previous analyses demonstrated that the HELP intervention was not associated with addiction benefit. In addition, we included HELP randomization group in the models and did not find significant interactions between PCAS scores and HELP randomization group (except for the alcohol model with organizational access). However, we acknowledge that the impact of primary care on addiction outcomes may have been evident in individuals who were "primed" by the HELP intervention to be responsive to the effects of primary care.

Finally, generalizability of the study's findings is another limitation. It is important to note that these data reflect a single urban adult population with substantial social and economic problems as well as alcohol and drug use disorders severe enough to require detoxification admission. The sample is not representative of patients with substance use disorders found in primary care, who may have less severe and persistent addiction trajectories. Hence, it is unclear whether our findings would be applicable to patients with less severe substance use disorders and these findings should be examined in other cohorts. Regardless of whether these findings are applicable to other patients in primary care, the population that we studied is an important one as this cohort is representative of many clinical samples in public-funded treatment settings (Smith and Larson 2003).

This study's findings are relevant to recent efforts to expand primary care's involvement in the management of addictive disorders. Financial and structural integration between substance abuse services and primary medical care have been proposed to increase coordination of traditionally separate systems of care (National Council for Community Behaviorial Healthcare 2003; New Freedom Commission on Mental Health 2003). As policy-makers move forward with increasing primary care's role in managing substance use disorders, our study suggests that certain core components of primary care, particularly the ability to provide whole-person care, are important to realizing the postulated benefits of increasing access to primary care.

In summary, multiple features of PCQ were associated with lower addiction severity and risk of future substance use. Our findings suggest that enhancing primary care's ability to deliver patient-centered, comprehensive, longitudinal care may have a beneficial impact on addiction outcomes.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute on Alcohol Abuse and Alcoholism (RO1-AA10870) and the National Institute on Drug Abuse (RO1-DA10019) (R25-DA13582). Support for Theresa Kim came from the National Institute Drug Abuse (R25-DA13582). This work was also supported, in part, by the Boston University General Clinical Research Center from the National Center for Research Resources (MO1-RR00533).

REFERENCES

Background Paper: Behavioral Health/Primary Care Integration Models, Competencies, and Infrastructure. 2003. National Council for Community Behavioral Healthcare.

Campbell, S., M. Roland, and D. Wilkin. 2001. "Primary Care Groups: Improving the Quality of Care through Clinical Governance." *British Medical Journal* 322: 1580–2.

Conrad, K. J., C. I. Hultman, A. R. Pope, J. S. Lyons, W. C. Baxter, and A. N. Daghestani. 1998. "Case Managed Residential Care for Homeless Addicted Veterans. Results of a True Experiment." *Medical Care* 36: 40–53.

De, A. I., J. H. Samet, and R. Saitz. 2004. "Burden of Medical Illness in Drug- and Alcohol-Dependent Persons without Primary Care." *American Journal of Addiction* 13: 33–45.

Fiellin, D. A., and P. G. O'Connor. 2002. "New Federal Initiatives to Enhance the Medical Treatment of Opioid Dependence." *Annals of Internal Medicine* 137: 688–92.

Friedmann, P. D., Z. Zhang, J. Hendrickson, M. D. Stein, and D. R. Gerstein. 2003. "Effect of Primary Medical Care on Addiction and Medical Severity in Substance Abuse Treatment Programs." *Journal of General Internal Medicine* 18: 1–8.

Grant, B. F., F. S. Stinson, D. A. Dawson, S. P. Chou, M. C. Dufour, W. Compton, R. P. Pickering, and K. Kaplan. 2004. "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." *Archives of General Psychiatry* 61: 807–16.

Institute of Medicine. 1996. *Primary Care: America's Health in a New Era.* Washington, DC: National Academy Press.

Johnson, J. G., R. L. Spitzer, J. B. Williams, K. Kroenke, M. Linzer, D. Brody, F. deGruy, and S. Hahn. 1995. "Psychiatric Comorbidity, Health Status, and Functional Impairment Associated with Alcohol Abuse and Dependence in Primary Care Patients: Findings of the PRIME MD-1000 Study." *Journal of Consulting and Clinical Psychology* 63: 133–40.

Kashner, T. M., R. Rosenheck, A. B. Campinell, A. Suris, R. Crandall, N. J. Garfield, P. Lapve, K. Prycz, T. Soyka, and A. Wicker. 2002. "Impact of Work Therapy on

Health Status among Homeless, Substance-Dependent Veterans: A Randomized Controlled Trial." *Archives of General Psychiatry* 59: 938–44.

Kertesz, S. G., M. J. Larson, N. J. Horton, M. Winter, R. Saitz, and J. H. Samet. 2005. "Homeless Chronicity and Health-Related Quality of Life Trajectories among Adults with Addictions." *Medical Care* 43: 574–85.

Kroenke, K. 2004. "The Many C's of Primary Care." *Journal of General Internal Medicine* 19: 708–9.

Liang, K., and S. L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73: 13–22.

Maisto, S. A., and R. Saitz. 2003. "Alcohol Use Disorders: Screening and Diagnosis." *American Journal of Addiction* 12 (suppl 1): S12–25.

McLellan, A. T., H. Kushner, D. Metzger, R. Peters, I. Smith, and G. Grissom. 1992. "The Fifth Edition of the Addiction Severity Index." *Journal of Substance Abuse Treatment* 9: 199–213.

McLellan, A. T., D. C. Lewis, C. P. O'Brien, and H. D. Kleber. 2000. "Drug Dependence, a Chronic Medical Illness: Implications for Treatment, Insurance, and Outcomes Evaluation." *Journal of the American Medical Association* 284: 1689–95.

Mertens, J. R., Y. W. Lu, S. Parthasarathy, C. Moore, and C. M. Weisner. 2003. "Medical and Psychiatric Conditions of Alcohol and Drug Treatment Patients in an HMO: Comparison with Matched Controls." *Archives of Internal Medicine* 163: 2511–17.

Murphy, J., H. Chang, J. E. Montgomery, W. H. Rogers, and D. G. Safran. 2001. "The Quality of Physician–Patient Relationships. Patients' Experiences 1996–1999." *Journal of Family Practice* 50: 123–9.

New Freedom Commission on Mental Health. 2003. *Achieving the Promise: Transforming Mental Healthcare in America. Executive Summary. SMA-03-3831.* Rockville, MD: DHHS.

O'Malley, A. S., and C. B. Forrest. 2002. "Beyond The Examination Room: Primary Care Performance and the Patient–Physician Relationship for Low-Income Women." *Journal of General Internal Medicine* 17: 66–74.

O'Malley, A. S., C. B. Forrest, and J. Miranda. 2003. "Primary Care Attributes and Care for Depression among Low-Income African American Women." *American Journal of Public Health* 93: 1328–34.

O'Malley, A. S., V. B. Sheppard, M. Schwartz, and J. Mandelblatt. 2004. "The Role of Trust in Use of Preventive Services among Low-Income African-American Women." *Preventive Medicine* 38: 777–85.

Preau, M., C. Leport, D. Salmon-Ceron, P. Carrieri, H. Portier, G. Chene, B. Spire, P. Chouter, F. Raffi, and M. Morin. 2004. "Health-Related Quality of Life and Patient-Provider Relationships in HIV-Infected Patients during the First Three Years after Starting PI-Containing Antiretroviral Treatment." *AIDS Care* 16: 649–61.

Safran, D. G. 2003. "Defining the Future of Primary Care: What Can We Learn from Patients?" *Annals of Internal Medicine* 138: 248–55.

Safran, D. G., M. Kosinski, A. R. Tarlov, W. H. Rogers, D. H. Taira, N. Lieberman, and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36: 728–39.

Safran, D. G., D. A. Taira, W. H. Rogers, M. Kosinski, J. E. Ware, and A. R. Tarlov. 1998. "Linking Primary Care Performance to Outcomes of Care." *Journal of Family Practice* 47: 213–20.

Safran, D. G., I. B. Wilson, W. H. Rogers, J. E. Montgomery, and H. Chang. 2002. "Primary Care Quality in The Medicare Program: Comparing the Performance of Medicare Health Maintenance Organizations and Traditional Fee-for-Service Medicare." *Archives of Internal Medicine* 162: 757–65.

Saitz, R., N. J. Horton, M. J. Larson, M. Winter, and J. H. Samet. 2005. "Primary Medical Care and Reductions in Addiction Severity: A Prospective Cohort Study." *Addiction* 100: 70–8.

Saitz, R., N. J. Horton, L. M. Sullivan, M. A. Moskowitz, and J. H. Samet. 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." *Annals of Internal Medicine* 138: 372–82.

Saitz, R., M. J. Larson, N. J. Horton, M. Winter, and J. H. Samet. 2004. "Linkage with Primary Medical Care in a Prospective Cohort of Adults with Addictions in Inpatient Detoxification: Room for Improvement." *Health Services Research* 39: 587–606.

Saitz, R., K. P. Mulvey, A. Plough, and J. H. Samet. 1997. "Physician Unawareness of Serious Substance Abuse." *American Journal of Drug and Alcohol Abuse* 23: 343–54.

Saitz, R., L. M. Sullivan, and J. H. Samet. 2000. "Training Community-Based Clinicians in Screening and Brief Intervention for Substance Abuse Problems: Translating Evidence into Practice." *Substance Abuse* 21: 21–31.

Samet, J. H., M. J. Larson, N. J. Horton, K. Doyle, M. Winter, and R. Saitz. 2003. "Linking Alcohol- and Drug-Dependent Adults to Primary Medical Care: A Randomized Controlled Trial of a Multi-Disciplinary Health Intervention in a Detoxification Unit." *Addiction* 98: 509–16.

SAS/STAT. 1999. *Version 8.2 of the SAS System for Windows. (8.2). 2001.* Cary, NC: SAS Institute Inc.

Schneider, J., S. H. Kaplan, S. Greenfield, W. Li, and I. B. Wilson. 2004. "Better Physician–Patient Relationships Are Associated with Higher Reported Adherence to Antiretroviral Therapy in Patients with HIV Infection." *Journal of General Internal Medicine* 19: 1096–103.

Smith, K. W., and M. J. Larson. 2003. "Quality of Life Assessments by Adult Substance Abusers Receiving Publicly Funded Treatment in Massachusetts." *American Journal of Drug and Alcohol Abuse* 29: 323–35.

Starfield, B., C. Cassady, J. Nanda, C. B. Forrest, and R. Berk. 1998. "Consumer Experiences and Provider Perceptions of the Quality of Primary Care: Implications for Managed Care." *Journal of Family Practice* 46: 216–26.

Ware, J. E. 1994. *SF-36 Physical and Mental Health Summary Scales: A User's Manual.* Boston: The Health Institute.

Weiner, S. J. 2004. "Contextualizing Medical Decisions to Individualize Care: Lessons from the Qualitative Sciences." *Journal of General Internal Medicine* 19: 281–85.

Weisner, C., and H. Matzger. 2003. "Missed Opportunities in Addressing Drinking Behavior in Medical and Mental Health Services." *Alcoholism Clinical and Experimental Research* 27: 1132–41.

Weisner, C., A. T. McLellan, and E. M. Hunkeler. 2000. "Addiction Severity Index Data from General Membership and Treatment Samples of HMO Members. One Case of Norming the ASI." *Journal of Substance Abuse Treatment* 19: 103–9.

Weisner, C., J. Mertens, S. Parthasarathy, C. Moore, and Y. Lu. 2001. "Integrating Primary Medical Care with Addiction Treatment: A Randomized Controlled Trial." *Journal of the American Medical Association* 286: 1715–23.

Welte, J. W., P. Perry, R. Longabaugh, and P. R. Clifford. 1998. "An Outcome Evaluation of a Hospital-Based Early Intervention Program." *Addiction* 93: 573–81.

Willenbring, M. L., and D. H. Olson. 1999. "A Randomized Trial of Integrated Outpatient Treatment for Medically Ill Alcoholic Men." *Archives of Internal Medicine* 159: 1946–52.

Zeger, S. L., K. Y. Liang, and P. S. Albert. 1988. "Models for Longitudinal Data: A Generalized Estimating Equation Approach." *Biometrics* 44: 1049–60.

Functional status outcomes among white and African-American cardiac patients in an equal access system

Nancy R. Kressin, PhD,[a,b] Mark E. Glickman, PhD,[a,b] Eric D. Peterson, MD, MPH,[c] Jeff Whittle, MD, MPH,[d,e] Michelle B. Orner, MPH,[a] and Laura A. Petersen, MD, MPH[f,g] ***Bedford and Boston, MA; Durham, NC; Milwaukee, WI; and Houston, TX***

Background Racial disparities exist in invasive cardiac procedure use and, sometimes, in subsequent functional status outcomes. We explored whether racial differences in functional outcomes occur in settings where differences in access and treatment are minimized.

Methods We conducted a prospective observational cohort study of 1022 white and African-American cardiac patients with positive nuclear imaging studies in 5 VA hospitals. Patients' functional status was assessed at baseline, 6, and 12 months later using the Seattle Angina Questionnaire and the SF-12, controlling for treatment received, clinical, sociodemographic, and psychological characteristics.

Results There were no significant baseline effects of race on functional status, after adjusting for sociodemographics, comorbid conditions, maximal medical therapy, severity of ischemia on nuclear imaging study, personal attitudes, and beliefs. Although there were no race differences in percutaneous transluminal coronary angioplasty use, there was a trend of African Americans being less likely to undergo coronary artery bypass graft, after 6 months (1.4% vs 6.5%) and 1 year (1.9 vs 6.9%). After adjustment, the decline in the SF12 Physical Component Summary from baseline to 6 months was, on average, 2.4 points less for African Americans than for whites, and at 12 months, Anginal Stability improved 8.4 points more for African Americans. The relative strength and direction of both findings persisted after removing covariates that might be confounded with race, and African Americans decreased less than whites on Physical Limitations, and improved more on Treatment Satisfaction, Anginal Frequency, and Disease Perceptions.

Conclusions In a setting where differences in access are minimized, so are racial differences in functional status outcomes. (Am Heart J 2007;153:418-25.)

From the [a]Center for Health Quality, Outcomes and Economic Research (a VA Health Services Research and Development National Center of Excellence), Bedford VA Medical Center, Bedford, MA, [b]Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, [c]Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, [d]Primary Care Division, Clement J Zablocki VA Medical Center, Milwaukee, WI, [e]Department of Medicine, Division of General Internal Medicine, Medical College of Wisconsin, Milwaukee, WI, [f]Division of Health Policy and Quality, Houston Center for Quality of Care and Utilization Studies (a VA Health Services Research and Development National Center of Excellence), Houston VA Medical Center, Houston, TX, and [g]Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX.

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

The research reported here was supported by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service (ECV 97-022.2, N Kressin, PI), and the American Heart Association and the Pharmaceutical Roundtable (9970113N, N Kressin, PI).

Submitted March 22, 2006; accepted November 30, 2006.

Reprint requests: Nancy R. Kressin, PhD, Center for Health Quality, Outcomes and Economic Research, VA Medical Center, 200 Springs Road, Building 70 (152), Bedford, MA 01730.

E-mail: nkressin@bu.edu

0002-8703/$ - see front matter

Published by Mosby, Inc.

doi:10.1016/j.ahj.2006.11.019

Despite the documentation of racial disparities in the use of invasive cardiac procedures in numerous settings,[1,2] African-American cardiac patients have *better* survival than white patients cared for in the Department of Veterans Affairs (VA),[3-5] where racial differences in access to care are minimized and where substantial efforts have been made to improve the quality and equity of treatment.[6,7]

Investigators have also examined disparities in functional status outcomes among cardiac patients, arguing that such information supplies necessary data beyond information on mortality, about the effects of various therapies on patients' functioning and well-being.[8,9] Functional status ratings vary by race in the general population, even after adjustment for demographic factors,[10] such that African Americans' functional status is generally worse than that of whites, but we know of only 1 study examining racial differences in functional status outcomes among potential candidates for revascularization.[11] In that study of angiography patients, before

taking into account patients' receipt of revascularization procedures, African Americans had worse physical and mental functioning, but these differences did not persist after controlling for treatments received. Based on the prior findings of better survival odds for African Americans, it seems likely that racial differences in functional status outcomes might also be equal or better for African-American patients in the VA setting, but the prior literature provides conflicting information.[4,5,11-14]

The purpose of our study was to examine whether there were differential functional status outcomes over 6 months and 1 year, among white and African American VA cardiac patients, after accounting for the care they received, patients' psychological characteristics and experiences, and clinical and sociodemographic factors, among a sample drawn from different geographic areas.

Method

Sample

Study setting and sample. The study methodology has been described in detail elsewhere.[15] Briefly, the research was conducted at 5 large, urban, academically affiliated tertiary care VA Medical Centers with on-site cardiac catheterization (Houston, St Louis, Durham, Atlanta, Pittsburgh). We established a prospective cohort of white and African American patients likely to have coronary artery disease by screening the results of all cardiac nuclear imaging studies performed between August 1999 and January 2001, selecting those with any evidence of reversible cardiac ischemia (evidenced by reversible defects or redistribution). A total of 1025 patients were included in the final baseline cohort (74% response rate), but for the present analyses we deleted 3 patients who received coronary artery bypass graft (CABG) after inclusion in the study but before they completed their baseline interview, as the recent receipt of CABG would likely affect functional status scores (N = 1022). At 6 months, 771 (75%) responded, and at 12 months, 760 (74%) responded.

This study was approved by the Human Studies Subcommittee of the VA medical centers involved in the study.

Data collection

Procedure. The functional status questions were included in baseline questionnaires administered shortly after the nuclear imaging study, and 6 and 12 months later.

Questionnaire measures

Independent Variables.

Demographics. We assessed the patient's age, self-reported race, education, income, and marital status.

Patient Psychological Characteristics and Experiences. At baseline, we assessed several dimensions related to functional status: patients' personality characteristics and experiences of racial discrimination. Negative Affectivity (NA; measured with the EPI-Q[16,17]) is a general disposition to experience subjective distress, which our prior work has demonstrated affects patient functional status ratings.[18] Optimism, a generalized predisposition to view things positively, is associated with health outcomes including rehospitalization after bypass surgery, so we reasoned that optimism might also be associated with patients' functional status ratings. We assessed this dimension with the Life Orientation Test.[19] Finally, as stressful experiences such as racial discrimination might negatively impact one's general well-being, including functional status, we used Krieger's well-known assessment of experiences of discrimination, to which we added an item specific to VA health care.[20]

Clinical and Treatment Variables. Trained nurses abstracted the medical records of each respondent, identifying demographics, cardiac symptoms, past medical history (including prior myocardial infarction, diabetes, hypertension, congestive heart failure, renal, or lung disease), laboratory values, test findings, and procedure utilization, and maximal medical therapy, following the definition used by the American College of Cardiology/American Heart Association guidelines for coronary angiography and the management of patients with chronic stable angina.[21,22]

We also tracked receipt of angiography and revascularization. We included all patients in the analysis, but accounted only for angiograms taking place within 120 days of the nuclear imaging study. In the postbaseline analyses, we excluded patients who had received CABG up to 2 months before each follow-up outcomes assessment (thus excluding patients still experiencing decrements in functional status due to the surgery).[23]

Nuclear Study Summary Scores. Two physicians classified the severity of each nuclear imaging study after review of the official report (blinded to patient race). We categorized the risk of severe coronary obstruction, using a modification of prior methods.[24] Patients with reversible lesions in the distribution of left anterior descending coronary artery or in both the right coronary artery and left circumflex artery were considered "high risk," as were patients with increased lung uptake or transient ischemic dilatation with exercise or pharmacologic stress. Patients with reversible lesions in just one of the right coronary artery or left circumflex artery were considered "moderate risk." Patients with very small or minimally reversible defects were considered "low risk." Mortality data were drawn from the VA Beneficiary Identification and Record Locator System.

Dependent Variables

Functional Status. We assessed functional status using 2 well-validated instruments, yielding 7 unique scales. First, we included the Seattle Angina Questionnaire (SAQ), which assesses patients' perceptions of several dimensions of coronary artery disease including anginal stability, anginal frequency, treatment satisfaction, disease perceptions, and physical limitations.[25] Second, we included a broad generic measure of health-related quality of life, the SF-12. This provides a mental component summary scale covering the dimensions of vitality, social functioning, role-emotional, and mental health, and a physical component summary scale covering the dimensions of physical functioning, role-physical, bodily pain, and general health perceptions. The SF-12 captures about 90% of the variance in the physical and mental component summary scales of the well-validated, but longer, SF-36.[26] For each functional status scale, higher scores indicate better functional status—for example, less frequent angina, better physical functioning.

Table I. Characteristics of the sample

	African Americans (n = 229)	Whites (n = 793)	P
Sociodemographic variables			
Age	61.6	63.4	.03
Education			.14
<12 y	32.6	27.9	
12 y/high school	30.0	36.7	
>12 y	37.4	35.4	
Married (% yes)	47.6	61.8	.0001
Employed (% yes)	16.6	19.8	.28
Income <$20,000	70.2	64.4	.12
Clinical variables			
Hypertension (% yes)	85.4	76.3	<.01
Angina (% yes)	65.6	64.3	.70
Congestive heart failure (% yes)	17.2	17.7	.85
Diabetes (% yes)	35.0	31.8	.37
Chronic obstructive pulmonary disease (% yes)	18.2	27.5	.01
Prior revascularization (% yes)	14.2	34.5	<.0001
Prior myocardial infarction (% yes)	26.2	33.9	.03
Renal dysfunction (% yes)	17.2	10.0	<.01
Maximal medical therapy (% yes)	32.3	35.6	.36
PTCA before baseline survey	2.6	2.5	.93
Catheterization within 120 d of nuclear imaging study	28.4	38.5	.01
PTCA between baseline and 6-m surveys (n = 989)	3.6	5.5	.27
PTCA between baseline and 6-m surveys among those receiving cardiac catheterization (n = 354)	13.3	14.3	.85
PTCA between baseline and 12-m surveys (n = 966)	3.8	5.8	.23
PTCA between baseline and 12-m surveys among those receiving cardiac catheterization (N = 350)	13.8	15.1	.80
CABG within 4 m of baseline survey (n = 989)	1.4	6.5	<.01
CABG within 4 m of baseline survey among those receiving cardiac catheterization (n = 354)	5.0	17.0	.02
CABG within 10 m of baseline survey (n = 966)	1.9	6.9	<.01
CABG within 10 m of baseline survey among those receiving cardiac catheterization (n = 350)	6.9	17.8	.03
Nuclear Study Summary scores			.90
High risk	42.2	43.2	
Moderate risk	40.4	38.7	
Low risk	17.4	18.1	
Died before 6 m (%)	3.5	3.0	.72
Died before 12 m (%)	6.6	5.2	.41
Personal attitudes/belief scales			
Optimism	58.7	59.2	.73
Racial discrimination*	35.3	5.6	<.0001
Negative affectivity*	4.4	3.9	.04

*Higher scores indicate more discrimination, greater negative affectivity.

Data analysis

Multiple analyses were performed on each of the 7 functional status scores, including all patients alive at each time point (baseline n = 1022, 6-month n = 989, 12-month n = 966). First, we examined racial differences in sociodemographics, clinical variables, and personal characteristics, and differences in functional status outcomes at each time point. Next, we examined racial differences in baseline scores adjusted for comorbidity, severity of the nuclear imaging study, medical therapy, demographics, and patients' NA, optimism, and experiences with racial discrimination (see full list of covariates in Table I). Then, we examined racial differences on 6-month (then 12-month) change in the scores adjusted for these covariates, and for baseline functional status. In a second set of analyses, we computed the results for models excluding variables that might be correlated with race: NA, experiences of discrimination, optimism, and severity of the nuclear imaging study. We also considered models where any interim revascularization procedures were included in the set of predictor variables. We performed a Bonferroni procedure to adjust for the multiple tests; instead of using a .05 significance level for individual tests on Tables I and II, we used .0024.

We adopted a Bayesian framework to fit all models; the advantages of this paradigm over a classical regression approach include the ability to model explicitly missing covariate information and to model dropout probabilities for longitudinal analyses without eliminating entire cases. We fit our models via Markov chain Monte Carlo simulation[27] from the posterior

Table II. Racial differences in functional status at 3 time points (unadjusted)

	African Americans	Whites	*P*
Baseline functional status			
PCS	32.4	32.2	.85
MCS	44.4	48.0	<.001
PL	74.1	76.5	.27
TS	83.7	86.8	.06
AF	73.0	75.8	.15
AS	65.0	70.0	.04
DP	64.7	70.2	.02
Six months			
PCS	33.9	32.0	.08
MCS	43.6	44.5	.44
PL	65.3	66.9	.60
TS	82.2	83.1	.70
AF	77.0	79.0	.39
AS	67.6	67.9	.93
DP	65.5	69.0	.19
Twelve months			
PCS	33.0	31.9	.34
MCS	42.6	45.2	.04
PL	68.5	67.4	.73
TS	82.3	83.4	.60
AF	79.8	79.5	.90
AS	70.4	67.5	.37
DP	68.5	70.5	.49

PCS, SF12 Physical Component Scale; *MCS*, SF12 Mental Component Scale; *PL*, SAQ Physical Limitations; *TS*, SAQ Treatment Satisfaction; *AF*, SAQ Anginal Frequency; *AS*, SAQ Anginal Stability; *DP*, SAQ Disease Perception.

distribution using the statistical software BUGS (MRC Biostatistics Unit, Cambridge, UK).[28] The analysis of baseline functional status was carried out using linear regressions with site-specific random intercepts. The effects of race were reported as the 95% central posterior intervals (roughly analogous to confidence intervals) for the coefficient of the race indicator variable. For the longitudinal analyses, we assumed that the probability of loss to follow-up is important to consider (ie, missing data may be "missing not at random"[29]), and that the probability of loss to follow-up may depend on the (unobserved) outcome value (eg, respondents with worse functional status may have been less likely to respond).[30] The model consists of 2 components, fit simultaneously; the first is a linear regression with the (possibly unobserved) outcome at follow-up as a function of the baseline value, race, site-specific random intercepts, and other covariates. The second is a logistic regression for the binary indicator of whether the outcome at follow-up was observed, modeled as a function of the (possibly unobserved) outcome at follow-up, adjusted for baseline score, race, and other covariates. We were interested in inferring the effect of race in the first model component. The assumption of a (normal) linear regression as the first model component essentially guarantees identifiability of the model parameters.[31] Note that 95% central posterior intervals do not need to be adjusted for multiplicity (eg, Bonferroni-type adjustments), as model summaries are considered simultaneous inferences for a single model.

To examine the sensitivity of the Bayesian selection models, we also fit least-squares regression (LSR) models of the first model component using the statistics package SAS (SAS Institute, Cary, NC), for patients only with complete cases.

Results

African Americans were less likely than whites to undergo cardiac catheterization within 120 days of the nuclear imaging study (28.4% vs 38.5%, $P < .005$). There were no racial differences in overall receipt of percutaneous transluminal coronary angioplasty (PTCA) between baseline and 6 months (3.6% vs 5.5%, not significant [NS]) or baseline and 12 months (3.8% vs 5.8%, NS), nor among the subset of patients who had received cardiac catheterization. However, there was a trend of whites receiving CABG more frequently during each period (1.4% vs 6.5%; 1.9% vs 6.9%; after the Bonferroni adjustment both were NS), with higher rates and more marked differences among the subset who had received cardiac catheterization (Table I). There were no significant racial differences in mortality over either period.

In unadjusted analyses, at baseline, African Americans had worse Mental Component Scores on the SF-12 (Table II). There were no race differences in any functional status scores at the other time points.

The baseline Bayesian regression analysis indicated that there were no significant effects of race, after adjusting for the covariates (results not shown), consistent with the complete-case LSR analysis. Although race was not significant, among the covariates, we observed that NA was negatively associated with all outcome measures (greater NA was associated with worse functional status). Among the clinical variables, angina was associated with worse outcomes in all dimensions except the Mental Component Summary scale, and receiving maximal medical therapy was associated with worse outcomes on Anginal Frequency, Anginal Stability, Disease Perceptions, and Physical Limitations, consistent with the notion that the sickest patients were receiving the most intensive therapy. Similarly, having a "high risk" nuclear scan was associated with worse Anginal Stability and Disease Perceptions. We fit 2 sets of models for each functional status outcome, one including revascularization received in the period between functional status measurements, and another model without. The effect of race was significant for the same functional status variables whether or not revascularization was included, so we present the results from the more comprehensive models here, including results with and without the 4 potentially race-related covariates.

As shown in Table III, which depicts changes in functional status over time, racial effects on change from baseline to 6 months were significant only for predicting 6-month change in the Physical Component Summary score of the SF-12 (see also Figure 1). After adjustment, the decline in Physical Component Summary from baseline to 6 months was, on average, 2.4 points less for African Americans than for whites (similar results were observed after excluding the race-related covariates: the decline was 2.31 points less for African Americans).

Table III. Bayesian analysis of racial differences in change in functional status outcomes at 6 and 12 m (whites relative to African Americans)

			Models excluding NA, experiences of discrimination, optimism, and magnitude of ischemia	
	Mean	**95% Posterior intervals**	**Mean**	**95% Posterior intervals**
Baseline—6 m				
PCS*	−2.40*	(−4.47, −0.48)	−2.31	(−4.03, −0.60)
MCS	1.00	(−1.32, 3.34)	0.44	(−1.68, 2.62)
PL	−1.17	(−6.56, 4.05)	0.18	(−4.49, 4.84)
TS	−1.93	(−5.89, 1.86)	−1.33	(−4.62, 2.05)
AF	−0.11	(−4.01, 3.98)	1.39	(−2.17, 5.01)
AS	−2.83	(−8.46, 2.64)	−1.32	(−6.25, 3.42)
DP	−0.67	(−5.66, 3.98)	0.43	(−3.79, 4.81)
Baseline—12 m				
PCS	−0.94	(−3.30, 1.52)	−0.22	(−2.30,1.85)
MCS	−0.99	(−3.87, 1.94)	−1.09	(−3.54, 1.36)
PL	1.47	(−7.51, 10.31)	8.21	(0.41, 15.91)
TS	−4.81	(−9.66, 0.07)	−5.97	(−10.12, −1.79)
AF	−3.10	(−8.21, 1.94)	−5.14	(−9.54, −0.68)
AS	−8.40	(−15.77, −1.13)	−10.26	(−16.87, −3.73)
DP	−4.71	(−11.27, 2.05)	−6.81	(−12.57, −1.10)

*Interpretation: The decline in PCS from baseline to 6 m was, on average, 2.4 points less for African Americans than for whites after adjusting for the same covariates as in the baseline analysis, baseline PCS score, revascularization (PTCA) received between baseline and 6 m, CABG received within the first 4 m after the baseline survey, and receipt of catheterization within 120 d after the nuclear imaging study.

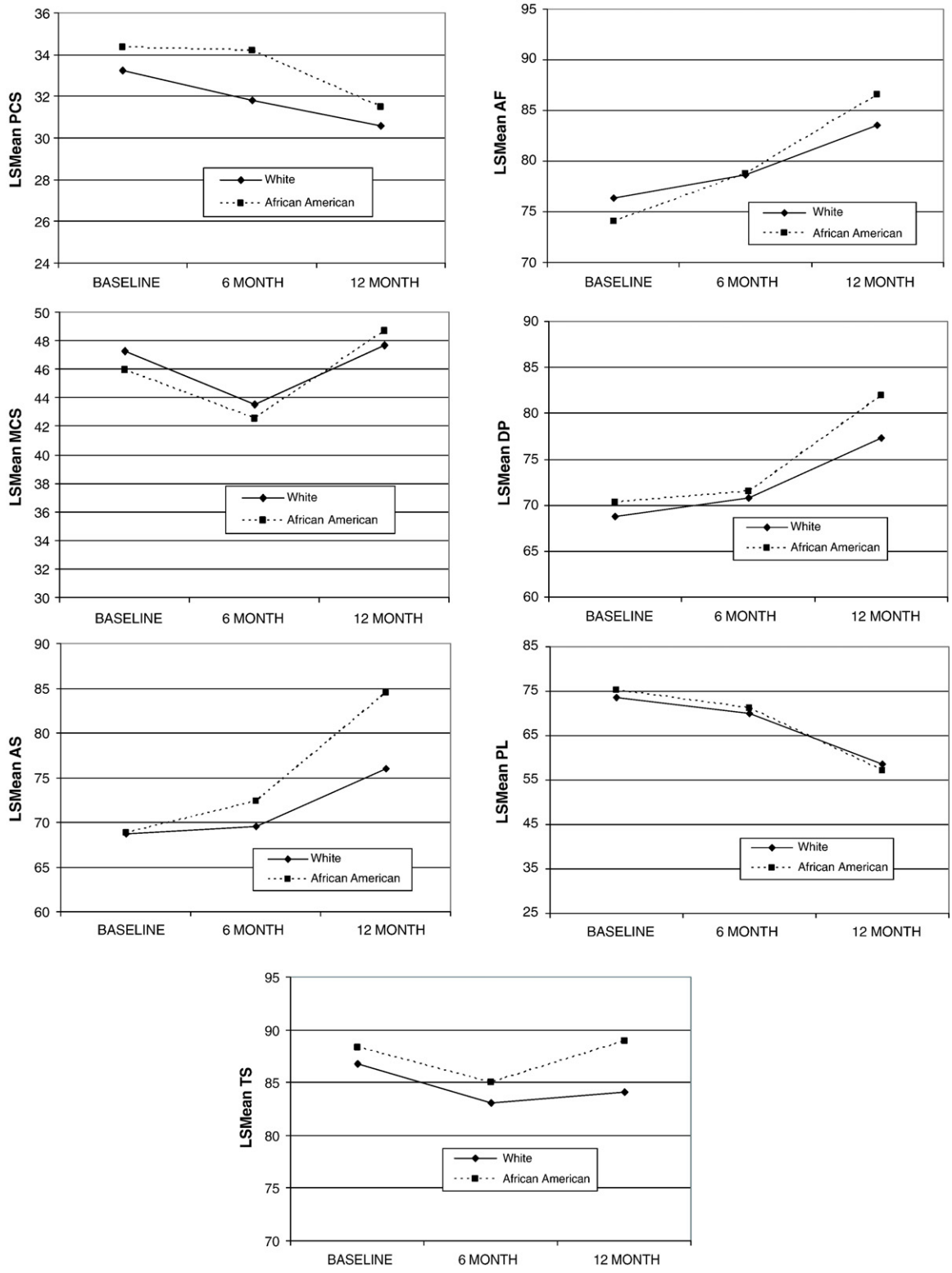
These are about the same magnitude of effect as the contrast between having angina versus not having angina.[32] Although the estimates of the coefficients in the complete-case LSR models were of comparable magnitude to those of the Bayesian regressions, none of these models evidenced race as significantly predictive of a functional status outcome score. Clinical and sociodemographic covariates consistently associated with greater decline in functional status across the models included angina (on all but Mental Component Summary), having received prior revascularization (on all but Treatment Satisfaction and Physical Component Summary), and having more education (on all but Disease Perception, Physical Component Summary, and Mental Component Summary; results not shown).

Racial effects for the full models predicting 12-month changes in functional status outcomes were only significant for the Anginal Stability model, where improvement from baseline to 12 months was, on average, 8.4 points higher for African Americans than for whites after adjusting for the same covariates as at baseline, as well as for revascularization (PTCA) received between baseline and 12 months, CABG received no later than 60 days before the 12-month survey, receipt of catheterization within 120 days after the nuclear imaging study, and baseline Anginal Stability score. Again, none of the complete-case LSR models found an association between race and functional status outcomes. After examining significant covariates, higher levels of NA were associated with worse functional status (except with the Anginal Frequency, Disease Perception, and Treatment Satisfaction scales). Prior revascularization was associated with worse functioning on the Anginal Frequency, Anginal Stability, and Disease Perception scales, but no other clinical variables were consistently associated with functional status outcomes (results not shown). Results from models excluding the race-related covariates indicated that African Americans improved more on Anginal Stability, Treatment Satisfaction, Anginal Frequency, and Disease Perceptions at 12 months, while declining more on Physical Limitations, although the posterior intervals nearly include 0, suggesting that these findings are not strongly significant.

Discussion

We examined racial differences in patients' functional status outcomes over 6 and 12 months, in a system of care where sociodemographic differences and difficulties in access to care are minimized, using data on a cohort of VA patients potentially eligible for revascularization, accounting for the cardiac care they received. We observed differences in the use of cardiac catheterization, and in CABG (but not PTCA), although all rates were relatively low. Contrary to prior research,[11] African Americans did not have worse functional status at baseline. Furthermore, over time, there were few racial differences in functional status, and the few observed differences indicated more favorable outcomes among African Americans than among whites. Thus, our results suggest that in a system where difficulties in access to care are minimized, and with an emphasis on minimizing differences in revascularization care, functional status outcomes are similar, whether or not we took treatment into account, contrary to earlier findings of differences in functional status before adjusting for treatment.[11]

Figure 1



Racial differences in functional status indices over time, adjusted for all covariates.

These results of similar or better outcomes for African-American patients are consistent with patterns observed in prior VA studies of mortality outcomes among patients with myocardial infarction, which showed that despite disparate receipt of invasive cardiac procedures including cardiac catheterization and revascularization, African Americans' mortality was better than whites'.[4,5] Our findings are also similar to those showing that African-American patients had worse unadjusted quality of life, as measured by the SF-36, but once results were adjusted for sociodemographic factors, comorbidity and functional status, no differences remained.[33]

However, our findings differ from other reports of racial differences in quality of life or functional status. Population-based studies have shown that African Americans consistently report worse health or functional status than whites even when results are adjusted by demographic factors and socioeconomic status.[10] In contrast to the general population, the relative sociodemographic homogeneity in VA across racial groups[34] might have influenced our results toward more similar outcomes.

How might our findings be explained? Our patients were farther upstream in the diagnostic process and were likely less sick than those of Kaul et al,[11] which would explain the better functional status, absence of racial differences in functional status, and the lower rates of revascularization we observed. Alternatively, the lower rate of revascularization could be explained by general underuse of cardiac procedures in the VA health care system,[35] which could have created a "floor effect," making it difficult to detect racial differences in procedure use or in functional status. Furthermore, we observed relatively slight differences in rates of revascularization between whites and African Americans, so similar functional status outcomes are not unexpected. If African Americans received worse care, we would expect them to have worse functional status, but an alternative explanation is that if African Americans who got worse over time died, then the surviving African Americans would have better functional status. However, we found no evidence of differential mortality rates (Table I). Another possibility is that the care given to each group was equivalent, but African Americans were not as chronically ill, rather they had more acute symptoms that led to their presentation and workup. However, this possibility is less plausible given research indicating that African-American patients are actually less likely to score positively for angina on the Rose scale.[36,37] The present findings are also consistent with our own results from other analyses of our data set where we found, among the subset of the cohort who underwent angiography, trends indicating that African Americans were less likely to have coronary obstruction and had less severe coronary disease than whites.[38] Thus, to the extent that African Americans' coronary artery disease was less severe in our cohort, their functional status should be better than whites', as demonstrated by the present results.

Our analyses were limited in several ways. First, we studied only patients actively undergoing evaluation of their cardiac status. Those patients selected for functional testing and those who chose to participate in the study may be healthier than those in the general population, potentially biasing the findings. However, this is not likely as the functional status scores in our population at baseline indicate notably poor health status overall. Our sample was representative of all nuclear imaging study patients at the included facilities by age and marital status, but there were fewer African Americans in the cohort than among those excluded from the study.[15] Finally, the VA system cares primarily for male patients, so our results may not be generalizable to women, or to patients in non-VA settings.

Strengths of our study included the fact that all facilities where patients were recruited had the capacity to conduct cardiac catheterization on-site. Unlike many prior studies, we included individual-level controls for sociodemographic and psychosocial characteristics. Because we studied VA patients, the effects of ability to pay for care, or physician financial incentives to recommend or deny procedures, were diminished. Furthermore, the sociodemographic gap between white and African-American patients is minimized in this setting.[34]

These results suggest that white and African American cardiac patients who had similar functional status at baseline and who were at a similar clinical starting point in the VA system received similar treatment and experienced similar or better functional status outcomes on most dimensions, whether or not we included perceptions of discrimination, negative affect, optimism, or magnitude of ischemia in our models. The absolute racial difference in revascularization rates was very small (among the full cohort), and we found largely similar functional status outcomes over a 1-year period, with the observed differences in outcomes indicating that African Americans fared slightly better, with only 1 exception. Thus, although African Americans and whites did not receive identical care, the few differences in outcomes suggest that, in this case, different care was not "disparate" care.[39]

Dr Kressin is a Research Career Scientist, Department of Veterans Affairs, Health Services Research and Development Service at the Edith Nourse Rogers Memorial Veterans Hospital; Dr Petersen was an Associate in the Career Development Award Program of the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service at the time that this work was conducted (Grant no. RCD 95-306), is a Robert Wood Johnson Foundation Generalist Physician Faculty Scholar, and an American Heart Association Established Investigator Awardee.

References

1. Kressin NR, Petersen LA. Racial variations in cardiac procedures: a review of the literature and prescription for future research. Ann Intern Med 2001;135:352-66.
2. Smedley B, Stith A, Nelson A. Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care, Board on Health Sciences Policy. Unequal treatment: confronting racial and ethnic disparities in health care. Washington (DC): National Academy Press; 2003.
3. Jha A, Shlipak M, Hosmer W, et al. Racial differences in mortality among men hospitalized in the Veterans Affairs health care system. JAMA 2001;285:297-303.
4. Petersen L, Wright S, Peterson E, et al. Impact of race on cardiac care and outcomes in veterans with acute myocardial infarction. Med Care 2002;40:I86-96.
5. Peterson ED, Wright SM, Daley J, et al. Racial variation in cardiac procedure use and survival following acute myocardial infarction in the Department of Veterans Affairs. JAMA 1994;271:1175-80.
6. Kizer K. The "new VA": a national laboratory for health care quality management. Am J Med Qual 1999;14:3-20.
7. Demakis J, McQueen L, Kizer K, et al. Quality Enhancement Research Initiative (QUERI): a collaboration between research and clinical practice. Med Care 2000;38:I-17-25.
8. Krumholz H, McHorney C, Clark L, et al. Changes in health after elective percutaneous coronary revascularization. Med Care 1996;34:754-9.
9. Hlatky M, Rogers W, Johnston I, et al. Medical care costs and quality of life after randomization to coronary angioplasty or coronary bypass surgery. N Engl J Med 1997;336:92-9.
10. Ren X, Amick B. Racial and ethnic disparities in self-assessed health status: evidence from the national survey of families and households. Ethn Health 1996;1:293-303.
11. Kaul P, Lytle B, Spertus M, et al. Influence of racial disparities in procedure use on functional status outcomes among patients with coronary artery disease. Circulation 2005;111:1284-90.
12. Maynard C, Ritchie J. Racial differences in outcomes of veterans undergoing coronary artery bypass grafting. Am J Cardiol 2001;88:893-5.
13. Rumsfeld J, Plomondon M, Peterson E, et al. The impact of ethnicity on outcomes following coronary artery bypass graft surgery in the Veterans Health Administration. J Am Coll Cardiol 2002;40:1786-93.
14. Peterson ED, Shaw LK, DeLong ER, et al. Racial variation in the use of coronary-revascularization procedures. Are the differences real? Do they matter? N Engl J Med 1997;336:480-6.
15. Kressin N, Chang B, Whittle J, et al. Racial disparities in cardiac catheterization as a function of patient perspectives. Am J Public Health 2004;94:2091-7.
16. Floderus B. Psychosocial risk factors in relation to coronary heart disease and associated risk factors. [Monograph] Nord Hyg Tidskr 1974;(Suppl 6):1.
17. Eysenck HJ, Eysenck SB. Manual of the Eysenck Personality Inventory. San Diego: Educational & Industrial Testing Service;1968.
18. Kressin NR, Spiro A III, Skinner K. Negative affectivity and health-related quality of life. Med Care 2000;38:858-67.
19. Scheier M, Carver C, Bridges M. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test. J Pers Soc Psychol 1994;67:1063-78.
20. Krieger N. Racial and gender discrimination: risk factors for high blood pressure? Soc Sci Med 1990;30:1273-81.
21. Scanlon PJ, Faxon DP, Audet AM, et al. ACC/AHA guidelines for coronary angiography: a report of the American College of Cardiology/American Heart Association task force on practice guidelines (Committee on Coronary Angiography). J Am Coll Cardiol 1999;33:1756-824.
22. Gibbons R, Chatterjee K, Daley J, et al. ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients With Chronic Stable Angina). J Am Coll Cardiol 1999;33:735-1097.
23. Borkon A, Muehlebach G, House J, et al. A comparison of the recovery of health status after percutaneous coronary intervention and coronary artery bypass. Ann Thorac Surg 2002;74:1526-30.
24. Bateman T, O'Keefe Jr J, Dong V, et al. Coronary angiographic rates after stress single-photon emission computed tomographic scintigraphy. J Nucl Cardiol 1995;2:217-23.
25. Spertus JA, Winder JA, Dewhurst TA, et al. Development and evaluation of the Seattle Angina Questionnaire: a New Functional Status measure for coronary artery disease. J Am Coll Cardiol 1995;25:333-41.
26. Ware JJ, Kosinski M, Keller S. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:22-233.
27. Gelfand A, Smith A. Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 1990;85:972-85.
28. Spiegelhalter DJ, Thomas A, Best NG, et al. BUGS: Bayesian Inference Using Gibbs Sampling (Version 0.60). Cambridge: Medical Research Council Biostatistics Unit; 1997.
29. Little R, Rubin D. Statistical analysis with missing data. 2nd ed. New Jersey: Wiley-Interscience; 2002.
30. Diggle P, Kenward M. Informative dropout in longitudinal data analysis (with discussion). Appl Stat 1994;43:49-94.
31. Hogan J, Roy J, Korkontzelou C. Biostatistics tutorial: handling dropout in longitudinal data. Stat Med 2004;23:1455-97.
32. Kazis L. The SF-36V short form health survey for veterans. Measuring and managing health care quality. Gaithersburg (Md): Aspen Publishers; 1998.
33. Johnson P, Goldman L, Orav E, et al. Comparison of the Medical Outcomes Study Short-Form 36-Item Health Survey in black patients and white patients with acute chest pain. Med Care 1995;33:145-60.
34. Wolinsky FD, et al. Veterans' and non-veterans' use of health services: a comparative analysis. Med Care 1985;23:1358-71.
35. Petersen L, Normand S, Leape L, et al. Regionalization and the underuse of angiography in the Veterans Affairs Health Care System as compared with a fee-for-service system. N Engl J Med 2003;348:2209-17.
36. Raczynski J, Taylor H, Cutter G, et al. Rose Questionnaire responses among black and white inpatients admitted for coronary heart disease: findings from the Birmingham-BHS project. Ethn Dis 1993;3:290-302.
37. Raczynski J, Taylor H, Cutter G, et al. Diagnoses, symptoms, and attribution of symptoms among black and white inpatients admitted for coronary heart disease. Am J Public Health 1994;84:951-6.
38. Whittle J, Kressin N, Peterson E, et al. Racial differences in prevalence of coronary obstructions among men with positive nuclear imaging studies. J Am Coll Cardiol 2006;47:2034-41.
39. Rathore S, Krumholz H. Differences, disparities, and biases: clarifying racial variations in health care use. Ann Intern Med 2004;141:635-8.

PERSPECTIVES

Career and Time Management Strategies for Clinical and Health Services Researchers

Nancy R. Kressin, PhD[1,2], Somnath Saha, MD, MPH[3,4], Frances Weaver, PhD[5,6], Lisa Rubenstein, MD, MSPH[7,8], and Morris Weinberger, PhD[9,10]

[1]Department of Veterans Affairs, Center for Health Quality, Outcomes & Economic Research, Bedford, MA, USA; [2]Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA; [3]Section of General Internal Medicine, Portland VAMC, Portland VAMC (P3MED), Portland, OR, USA; [4]Oregon Health and Science University, Portland, OR, USA; [5]Center for Management of Complex Chronic Care, and Spinal Cord Injury QUERI, Hines VA Hospital, Hines, IL, USA; [6]Institute for Healthcare Studies, Northwestern University, Chicago, IL, USA; [7]VA HSR&D Center of Excellence for the Study of Healthcare Provider Behavior, VA Greater Los Angeles Healthcare System, Sepulveda Ambulatory Care Center and Nursing Home, Sepulveda, CA, USA; [8]University of California, Los Angeles, CA, USA; [9]Center of Excellence for Health Services Research in Primary Care (152), Durham VAMC, Durham, NC, USA; [10]Department of Health Policy and Administration, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

J Gen Intern Med 22(10):1475–8
DOI 10.1007/s11606-007-0337-7

INTRODUCTION

Increased competition for research funding requires that clinical and health services researchers be focused and efficient if they are to sustain their research programs over their careers. Developing these skills is paramount for trainees and junior faculty. Although career development issues related to overall 'gestalt' concerns[1], general career building advice[2,3], and strategies for obtaining funding[4] have been previously addressed, there has been little published about strategies for maximizing one's focus and efficiency as a clinical or health services researcher[5]. In this article, we offer a set of such strategies for junior investigators.

First, Know Thyself (The Person)

We begin with a central guiding principle: understanding the goals, needs, and personal characteristics we bring to our careers is critical to making career decisions and selecting strategies to maximize efficiency and focus.

Understanding Your Values and Their Relative Priority. We all balance competing demands of work against personal commitments. The relative priority of these demands varies at different times in our lives, and explicitly recognizing their relative importance is essential. Further, it is important to clarify one's roles and what they require. How do you allocate your work time across research, teaching, and clinical activities? What are your familial responsibilities (both situational and ongoing)? Attaining a workable balance between one's career and family life is vital to success and satisfaction in each dimension, and reserving time to 'sharpen the saw'[6]—to rest, reflect, exercise, restore energy and perspective—is vital to having a full reservoir of energy and drive for continuing one's work.

We recommend you literally write down a "values" statement, indicating what is important to you in each aspect of your life (personal, family, community, career). This can help with explicating your relative priorities and with providing guidance for how to balance time among those different aspects and within each dimension.

Understanding Your Work Style, and Your Strengths and Weaknesses. Some people can adhere to internal deadlines by themselves; others need an external deadline or to be accountable to others (e.g., a mentor). Some investigators thrive on multi-tasking; others need to complete tasks serially. Some individuals are natural leaders; others are not. Some prefer large projects with complex multidisciplinary teams; others prefer smaller projects. Understanding what strengths we "bring to the table" helps maximize efficiency as we identify collaborators and resources that complement our strengths. The key is to recognize your values, work style, skills and limitations and then structure work activities based on this self-assessment. Mentors, supervisors, and trusted colleagues can help you identify your strengths and weaknesses and offer strategies to maximize focus and efficiency.

Be Clear on Your Goals. Academic medicine includes a range of activities including research, teaching, clinical activities, and mentoring; it is important to understand which dimensions are most and least satisfying. One strategy for clarifying this is to take your "affective temperature" during different activities; identifying those which you enjoy or look forward to the most

Received April 10, 2007
Revised August 1, 2007
Accepted August 6, 2007
Published online August 24, 2007

can help you make decisions about your priorities. Whichever you choose, be constantly mindful of your long-term goals and identify realistic short and intermediate milestones consistent with these goals. Develop a Gantt chart and/or "to do" list to organize goals and tasks and insure that your time commitments are consistent with your priorities. Using such tools to organize a timeline can help you scan the horizon for upcoming commitments, tasks, and deadlines, and help you determine whether to take on additional commitments. Your timeline must be updated regularly, especially if your long-term goals change.

You must discipline yourself to meet your deadlines, even informal ones. For example, it is easy to delay submitting a manuscript that is not perfect. However, "perfection is the enemy of the good", and at the point of diminishing returns, a paper should be submitted. Rewarding the completion of such tasks can give personal satisfaction and help reinforce completion of the next goal. We caution you, however, to not lose the forest for the trees—publications and funding are a means to an end, and a fulfilling career will be determined by work driven by values, and by progress towards vision, not by the number of publications and grants.

Organizing One's Schedule. We recommend creating and following a schedule, including scheduling time to write grants and manuscripts. The schedule should reflect your values, strengths, and weaknesses, and work style, as well as the urgency and importance of each task[5]. For example, if you write more effectively in the morning, protect this time for writing. From an efficiency standpoint, it may be more beneficial to arise 1 h earlier each day to write for an hour, than to spend three evening hours trying unsuccessfully to concentrate when one is tired from the day.

Second, Know Thy Environment (The Environment)

We all work in environments with formal and informal rules, expectations, and resources. Strategies for learning these rules have been covered in detail[2]; one must learn the rules and culture of one's specific environment so as to clearly understand what is valued, the metrics by which success is judged, and resources available to support one's career (e.g., pilot funds, professional development seminars). Mentors, supervisors, and other experienced investigators are critical to understanding the local environment.

Within your environment, identify mentors and potential collaborators. We advise you to query trainees and other junior faculty regarding strengths and weaknesses of potential mentors and collaborators before establishing working relationships[2]. It is also important to identify resources to help build your portfolio and career. We urge you to 'think outside the box' to actively identify such resources, whether or not they seem available or have been offered to you. Once identified, you should aim to develop skills at negotiating for those resources. Some are better than others at negotiating[7], so if necessary, strive to develop such skills[8,9].

Funded faculty with available data sets, analysts and statistical resources can often help pave the way for publications and ideas for new funding proposals, which may be more efficient than starting a study from the beginning. It is important to identify potential mentors and co-investigators who have a track record of following through with projects—evidence of a successful publication record, a continued funding stream, and successful former mentees are each signs of a desirable mentors and colleagues.

It is sometimes difficult to find a single mentor to provide both content and methodological expertise. One can often learn from mentors with clinical or health services research skills, even if their focus differs from yours. With creativity, you can organize a mentoring team that capitalizes on local expertise and resources that will support development of your core interests. Your mentoring team can include a senior investigator whose time is more limited, coupled with a more junior investigator. You may seek one or more external mentors with specific technical or content expertise. You also might identify opportunities to participate as a site collaborator in a relevant multi-site study. Here, you could build relationships with mentors, add on small grants, or write papers that amplify the study beyond its main products. Regardless of the specific strategy, choose mentors and colleagues whose personal characteristics complement your own.

Developing successful mentoring relationships is hard work. Mentees must learn when to ask for help and develop focused agendas for meetings to respect mentor time. Effective mentees are also able to disagree with their mentors, while remaining open-minded. If you disagree with your mentor, pose an alternative. We advise against either blindly sticking to your own perspective or applying answers from your mentor that you do not believe in; rather, you should develop skills in discussing your concerns and finding common ground. If you find yourself feeling criticized or overwhelmed by your mentor's responses, talk to others. This will help you gain perspective and ensure that you are not taking critiques of your work personally. Generally, a much greater cause for alarm is not being adequately critiqued.

You should identify local resources for junior investigators by querying your mentors, senior faculty, and office of research administration. Many medical schools have intramural junior faculty research funding to support pilot work preparatory to a grant proposal, or to provide initial funding for a program of research which can later be leveraged into more substantial funding.

Third, Determine What You Need to Succeed (Person–Environment Fit)

Our recommendations are designed to help you maximize the fit between yourself and your environment. When such "Person–Environment Fit" is not optimal, "strain…develops when there is a discrepancy between the demands of the job and the ability of persons to meet those demands"[10], which saps needed energy away from doing your work. Armed with knowledge about yourself and your environment, you and your mentor can identify strategies to enhance your success. Perhaps you spend too much time on administrative tasks, which others can perform-learning to identify persons to whom to delegate, and then to actually delegate such tasks, is an essential survival skill. Perhaps you devote too much time to nonresearch activities, including committee assignments, teaching or clinical activities. Learning to protect one's time is essential, and this is often a challenge if "opportunities" are

presented by someone with direct authority over you. Mentors are essential to helping junior investigators decide what is reasonable.

Manage Volume. We recommend careful thought and constant vigilance regarding your work volume. Taking on too many tasks and responsibilities inevitably leads to missed deadlines, personal frustration, and disappointment from mentors, colleagues and superiors. Thus, it is best to manage this volume at the outset, by making careful and thoughtful decisions about what to take on. To make these decisions, one must constantly refer back to one's mission and goals, and how well each possible activity fits within this framework. If necessary, you may need to offload activities through delegation, or remove activities from your portfolio. In doing this, it is useful to consult your mentor or supervisor, to ensure that you are considering not only your immediate needs but also your career interests and those of your institution.

Learn How to Say "No". Knowing when to say "no" and developing skills to do so are crucial to managing work volume. Remember that your time is valuable, and time spent in low priority activities (even when they are "opportunities" from your superiors) detracts from your ability to engage in tasks that are critical for your career. Chin encouraged postponing commitment-making whenever possible, to allow one to think carefully before responding (Figure 1). Thus, following his advice, we urge you to rehearse the response: "Thank you for asking but let me think about it and get back to you" so as to give yourself enough time to discern if a new commitment fits your priorities. In doing so, think about the political realities of the environment—whether doing something will help your center, department, or group—and whether it is important to being a good citizen, or whether it is someone else's turn to step up for such activities. Engaging one's mentor/chief to "play the heavy" can be a useful way of saying no, or providing a reason for saying no (e.g. "My mentor advised me against participating in this project") when it is politically difficult.

Minimize Switching Costs. 'Switching costs' are those associated with excessive multitasking in which switching back and forth between projects incurs significant losses of time and energy associated with refamiliarizing oneself with a prior project, remembering where one left off, and organizing a new set of papers and files[11]. We encourage minimizing switching costs by focusing primarily on high priority tasks each day or week. The oft-given advice to focus one's efforts within a certain research content area partly stems from the recognition of the intellectual switching costs involved in mastering multiple content domains (as well as the fact that lack of focus slows one's progress towards building a national reputation in a specific area). However, sometimes, we must switch tasks: for example, from manuscript to grant proposal preparation; in such cases it becomes important to minimize the switching costs.

How one does this depends upon one's personality. Some work more effectively by designating entire days for certain activities (e.g., Wednesdays are 'meeting days', Thursdays are 'writing days'). Others are comfortable switching tasks after shorter periods. There is no one right answer; you must uncover the strategy that works best for you, and apply it in a way that minimizes wasted time.

1. Does it fit your mission and agenda?
   Is the opportunity something you are excited or passionate about?
2. What impact will you have?
   Do you have skills and perspectives that will be a valuable contribution, and do you have the resources to accomplish the goal?
3. What is the time commitment?
   Can you offload undesireable parts of the task, and will you have administrative support to complete it?
4. Can you make it more academic?
   Is the opportunity purely service or are there academic possibilities that might lead to a publication opportunity?
5. Can you say no or negotiate the responsibility?
   Are you pulling your fair share of the weight? Is there an alternative service obligation that more closely fits your interests?
6. How stable is your research program and/or funding streams?
   The weaker your funding and power base, the more you need to concentrate upon shoring this up before taking on new responsibilities.
7. How stable is your family/personal situation?
   If you just had a baby or have ill parents, it might be better to wait for future opportunities (more will always come).

Figure 1. When to Say Yes and when to Say No (reprinted and excerpted with permission from Marshall Chin, MD, MPH, and the Society of General Internal Medicine "Forum" publication)

Various strategies can help minimize switching costs. When stopping work on a project, leave a 'to do' list for your return, thereby minimizing the need to rethink where you left off when you reinitiate that project. Always document and date your work, so that if an unexpected interruption to the project occurs, you can more easily reconstruct where you left off and where you need to begin again. Having a written overall project work plan, or "to do" list, helps the sequence of tasks remain clear even if interrupted. Regardless of strategies, maintain an updated "to do" list with deadlines, and review this list with mentors. In addition, consider scanning documents to create electronic, rather than paper, folders, which can be more easily searched using search engines when materials need to be located.

Minimize Interruptions. A common interruption is email, which can be ubiquitous and overwhelming. Consider dealing with emails during 1–2 blocks per day. Similar strategies can be adopted from other potential distracters. You might ask to be paged only in emergencies, requesting that nonurgent clinical or other calls be directed to your voice mail. Then, return pages and calls once or twice a day. Try to avoid being called into impromptu meetings. If acceptable in your workplace, consider working somewhere where you will be harder to find, for example, the library. In addition, endeavor to minimize psychic interruptions, that is, distractions to one's train of thought from oneself. For example, if your priority is writing the science of a grant but you are repeatedly distracted by the many related administrative details, consider keeping a side list of administrative tasks, to which you can attend after your writing time.

In summary, career management involves understanding and managing yourself and your environment, as well as the fit between the two. This involves prioritizing and balancing family and work, emphasizing your strengths and supplementing to counter your weaknesses, and knowing your own personal clock. Knowing your work environment will allow you to maximize your productivity by maximizing your use of resources, minimizing switching costs, managing your work volume and learning when and how to say no. No one does this perfectly, but we believe that striving for such balance helps one to be more successful in life.

Acknowledgement: *Drs. Kressin, Weaver and Weinberger are supported by Research Career Scientist awards, and Dr. Saha is supported by an Advanced Research Career Development Award from the Health Services Research & Development Service, Department of Veterans Affairs. Dr. Saha is also supported by a Generalist Physician Faculty Scholar Award from the Robert Wood Johnson Foundation. We thank Victoria Parker, DBA, for introducing us to the concept of switching costs. This work was previously presented at the 2006 Annual Meeting of the Department of Veterans Affairs Health Services Research and Development Service. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the Robert Wood Johnson Foundation.*

Conflicts of Interest: *None disclosed.*

Corresponding Author: *Nancy R. Kressin, PhD; Department of Veterans Affairs, Center for Health Quality, Outcomes & Economic Research, Bedford VA 200 Springs. Rd. Bldg 70 (152), Bedford, MA 01730, USA (e-mail: nkressin@bu.edu).*

REFERENCES

1. **Goldman L.** Blueprint for a research career in general internal medicine. J Gen Intern Med. 1991;6:341–4.
2. **Chin M, et al.** Building a research career in general internal medicine. J Gen Intern Med. 1998;13(2):117–22.
3. **Applegate S, Williams M.** Career development in academic medicine. Am J Med. 1990;88:263–7.
4. **Gill T, et al.** Getting funded: Career development awards for aspiring clinical investigators. J Gen Inter Med. 2004;19:472–8.
5. Burroughs Wellcome Fund and Howard Hughes Medical Institute, Making the Right Moves: a Practical Guide to Scientific Management for Postdocs and New Faculty. 2006, Howard Hughes Medical Institute: Research Triangle Park, North Carolina, and Chevy Chase, Maryland.
6. **Covey SR.** The 7 Habits of Highly Effective People. NY, NY: Simon & Schuster; 2004.
7. **Babcock L, Laschever S.** Women Don't Ask: Negotiation and the Gender Divide. Princeton, NJ: Princeton University Press; 2003.
8. **Fisher R, Ury W.** Getting to Yes. New York, NY: Houghton Mifflin; 1991.
9. **McCormack M.** What They Don't Teach You at Harvard Business School. New York, NY: Bantam Books; 1984.
10. **Caplan R, et al.** Job Demands and Worker Health. Cincinnati, OH: National Institute for Occupational Safety and Health; 1975.
11. **Rothman J.** Multitasking Overhead. 2003.

Hypertensive Patients' Race, Health Beliefs, Process of Care, and Medication Adherence

Nancy R. Kressin, PhD[1,2], Fei Wang, PhD[1,3], Judith Long, MD[4,5], Barbara G. Bokhour, PhD[1,3], Michelle B. Orner, MPH[1], James Rothendler, MD[1,3], Christine Clark, PharmD[6,7], Surekha Reddy, MD[6,7], Waldemar Kozak, MPH[1], Laura P. Kroupa, MD[8,9], and Dan R. Berlowitz, MD, MPH[1,3]

[1]Center for Health Quality, Outcomes and Economic Research (a VA Health Services Research and Development National Center of Excellence), Bedford VA Medical Center, Bedford, MA, USA; [2]Division of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA; [3]Health Policy and Management Department, Boston University School of Public Health, Boston, MA, USA; [4]CHERP, Philadelphia VAMC, Philadelphia, PA, USA; [5]Department of Internal Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA; [6]Jesse Brown VAMC, Chicago, IL, USA; [7]University of Illinois at Chicago, Chicago, IL, USA; [8]St. Louis VAMC, St. Louis, MO, USA; [9]St. Louis University, St. Louis, MO, USA.

BACKGROUND: African Americans have higher rates of hypertension and worse blood pressure (BP) control than Whites, and poorer medication adherence may contribute to this phenomenon. We explored associations among patients' race, self-reported experiences with clinicians, attitudes and beliefs about hypertension, and ultimately, medication adherence, among a sample with no racial disparities in BP control, to determine what lessons we could learn from patients and providers in this setting.

METHODS: We recruited 793 White and African-American (58%) patients previously diagnosed with hypertension from 3 VA medical centers to participate in survey assessments of each of the above dimensions, subsequent to a primary care clinic visit.

RESULTS: African-American patients' providers were significantly more active in advising and counseling about hypertension care and medication adherence. African-American patients indicated greater knowledge or heightened awareness of the importance of controlling their BP, but there were no race differences on a summary adherence measure. In multivariate models modeling medication adherence, race was not significant, but having been told to split one's pills, believing one's BP continues to be high, and having one's provider discuss things to do to make it easier to take BP medications were each significantly associated with worse adherence, whereas having more confidence in one's ability to take BP medications as prescribed was associated with better adherence (all p's $\leq .02$).

CONCLUSION: When both physicians and patients take BP management seriously, disparities in BP adherence and control may be reduced.

Received June 21, 2006
Revised January 4, 2007
Accepted February 27, 2007
Published online March 16, 2007

KEY WORDS: physician–patient relations; patient compliance; attitude to health.
DOI: 10.1007/s11606-007-0165-9
 2007;22:768–774

BACKGROUND

Hypertension affects nearly 50 million Americans,[1] and is more frequent among African Americans (www.americanheart.org., cited 2000 November 2), accounting for a significant portion of racial differences in mortality[2]. Many patients with established hypertension have poorly controlled blood pressure (BP), and African Americans are disproportionately represented among this group[3].

Poor medication adherence is one of the most important barriers to blood pressure control[1], but there is mixed evidence as to whether African Americans are less likely to adhere to prescribed therapies than Whites[4,5]. Provider–patient communication is an important determinant of adherence[6], and poor communication is thought to contribute to worse care for African Americans and other ethnic minority individuals[7,8]. In addition, patients' beliefs and attitudes about their blood pressure, and their experiences with providers regarding BP management, may ultimately affect adherence to treatment recommendations[9,10]. Yet little is known about how each of these dimensions varies by patient race. Such information is vital for the development of future clinical interventions to ultimately address disparities in blood pressure control.

Whereas most prior studies of race, medication adherence, and hypertension care have utilized samples with racial differences in BP control, we posit that the opposite approach might help identify factors associated with *success* in eliminating differences in BP care and outcomes. Thus, drawing from a multisite Department of Veterans Affairs (VA) sample with no racial disparities in BP control rates, we examined the links between patient race, experiences with clinicians, attitudes and beliefs about hypertension, and ultimately, adherence to antihypertensive medications. We first studied patterns of interactions with clinicians and patients' beliefs about

hypertension by race. Second, we examined patterns of antihypertensive medication adherence by race. Third, we evaluated whether any racial differences in interactions with clinicians or beliefs about blood pressure were related to observed differences in adherence.

METHODS

Sample

Initially, we identified all White and African-American patients with outpatient diagnoses of hypertension on at least 2 separate occasions in 2001 at 3 urban tertiary care Department of Veterans Affairs (VA) Medical Centers (ICD9 diagnosis codes: 401—401.0, 401.1, 401.9; 405—405.11, 405.19, 405.9, 405.91, 405.99). The study was approved by the Institutional Review Boards of all participating facilities, and patients provided informed consent.

Using this "universe" of 11,731 hypertensive patients from the 3 medical centers, study staff tracked patients' primary care visits over a 14-month period, and as they presented for care, approached 1,210 of them to request participation in the study. A total of 203 were excluded owing to their race not being African American or White (n=18), poor mental status (n=41), denying hypertension (n=59), participation in another hypertension study (n=6), or miscellaneous other reasons (n=79), leaving 1,007 eligible patients. Two hundred and fourteen patients (18% of the 1,210 approached) refused to participate. Thus, 793 patients were included in the final cohort (78.7% response rate).

Measures

Overview. We assessed 3 primary domains of interest: 1. Patient experiences with providers, 2. Patient characteristics, including sociodemographic factors and health beliefs, and 3. Antihypertensive medication adherence. In addition, we assessed numerous covariates that we thought were relevant to either the independent or dependent variables. Our selection of covariates was informed by the Health Decision Model (HDM;[11]). Figure 1 presents our expanded conceptualization of the factors that affect individuals' health decisions (e.g., medication adherence), to include patient personality, preferences and knowledge, prior experiences, and social interactions (such as with physicians and family), in addition to health beliefs.

1. Experiences with providers
   Using questions adapted from Ockene[12], we assessed the content of the doctor–patient interaction focusing on hypertension and antihypertensive medication adherence, through an exit interview with each patient after his/her visit. Such "Patient Exit Interviews" (PEIs; see Table 2 for items [all yes/no responses]) have been demonstrated to accurately measure the actual content of clinic visits, through comparisons of audiotapes of such interactions to patient reports[13]. The question answers were summed to create 1 PEI scale score, with a range from 1 to 12 (higher scores indicate greater frequency of counseling).
2. Patient sociodemographic characteristics, and health beliefs
   Patients completed an interview including questions about sociodemographic characteristics: date of birth, highest grade in school completed, employment, marital status,

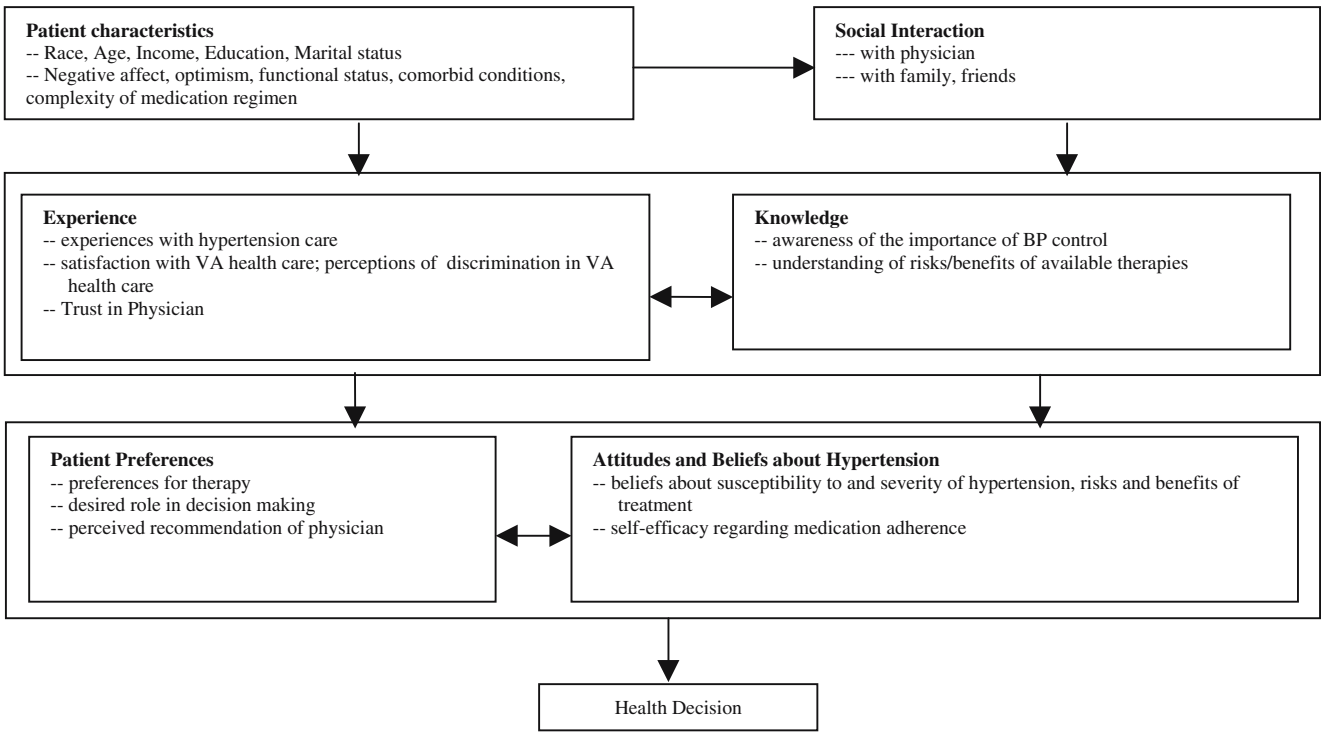


Figure 1. The health decision model, adapted from Eraker, Kirscht, and Becker, 1984 [12].

income, and race. We also obtained information about comorbid diagnoses and BP measurements from the medical record.

Following the Health Belief Model (HBM), which posits that patients' appraisals of disease risk (susceptibility) and severity influence the practice of health-related behaviors[14], we assessed patients' perceptions of the seriousness of blood pressure, and consequences resulting from not taking BP medications (answer categories ranged from "extremely serious—a threat to one's life', to "not at all serious"). The HBM posits that an individual's considerations of the perceived barriers and incentives for performing specific behaviors are important to consider. Thus, we asked patients if they thought BP medications would make them feel worse, or better, help them live longer, improve their quality of life, or prevent BP-related illnesses (all yes/no responses). Finally, the HBM considers cues to action, which prompt individuals to act by reminding them of the need to change their behaviors. We assessed 1 particular set of cues by asking patients if their family or friends help them to remember to take their BP medication.

3. Covariates

There is evidence that trust in one's physician, experiences of racial discrimination in the health care setting, self-efficacy (e.g., knowledge about one's BP, and confidence in one's ability to take BP medications), complexity of one's medical regimen, patient personality (both negative affect and optimism), and functional status might affect adherence. Thus, we assessed each dimension, as follows.

To assess trust in one's physician, we included the 11-item Trust in Physician scale[15], which assesses the extent to which patients feel their doctor cares about him/her, is considerate of his/her needs, tells the truth, and makes good judgments. The patients were asked to respond to each statement on a 5-point scale (1=totally agree; 5=totally disagree). Appropriate items were reverse coded and the answers to these questions were summed and converted to a 100-point scale, so higher scores represented greater trust. To assess patients' experiences of racial discrimination in the health care setting, we included the "Experiences of Discrimination" scale. Item responses were summed and converted to a 100-point scale so that higher scores indicate less discrimination[16]. To assess self-efficacy, we asked patients how confident they were about their knowledge about BP and how to deal with it, and in their ability to take their BP medication as prescribed (0=not confident at all; 10=totally confident)[6]. To assess the complexity of a patient's medication regimen[17], we asked patients how many different types of BP pills they take, how many times each day they take it, and whether they were told to split their pills. Negative Affectivity (NA) was measured with the EPI-Q[18]. Optimism was assessed with the Life Orientation Test[19]. We included a broad generic measure of functional status, the VR-12 (the SF-12 adapted for veterans;[20]), which provides a mental component summary scale (MCS) and a physical component summary scale (PCS).

Outcome Assessment – Antihypertensive Medication Adherence. We asked patients to self-report their medication adherence, using items from 2 well-validated measures from the literature to assess multiple dimensions of medication adherence. Following Choo et al.[21], we assessed how many days in the past week patients *forgot* to take their blood pressure medication (0, 1, 2 or more days), how many days they did *not* take their medication on purpose (0, 1, 2, or more days), how many days they *added* an extra pill (0, 1, 2, or more days), whether they ever took less medicine because they felt they needed less (yes or no), and following Morisky et al.[22] we asked, "sometimes if you feel worse when you take the medicine, do you stop taking it?" (yes or no). Finally, we asked, "Some people have difficulty taking blood pressure medication as prescribed. Do you have difficulty with this?" (yes or no).

We evaluated several different methods for combining these items, assessing the validity of each possible summary measure by examining its association with blood pressure control, choosing the measure with the strongest association. Thus, our final measure of adherence included the items assessing forgetting to take medications, not taking medications on purpose, taking less medications because of perceptions that one needed less, and having difficulties taking medications as prescribed. We specified that the patient must answer positively to at least 1 of the items to be considered as having adherence problems. We calculated this variable only for those who answered all of these questions. For the variables with 3 responses, a person was considered adherent if he/she indicated adherence problems for 0 day or 1 day each week, and nonadherent if he/she indicated adherence problems for 2 or more days; the other questions had yes/no responses (adherence problems or not).

Statistical Analyses. We first examined whether there were racial differences in interactions with clinicians, in patients' beliefs and attitudes regarding hypertension, or in medication adherence, using bivariate analyses of each variable by race (chi square or *t* tests, as appropriate). Next, we performed bivariate analyses to determine which sociodemographic variables, health beliefs, and PEI items and other covariate scales were associated with the dichotomous adherence summary outcome. We then conducted stepwise analyses

Table 1. Sociodemographic and Clinical Characteristics of the Cohort, by Race

	African Americans *N*=460	Whites *N*=333	*p* value
Sociodemographic variables			
Age (mean)	64.4	67.3	<.001
Education (%≥12 years)	72.4	75.0	.42
Income (% ≥$20,000)	45.5	47.9	.53
Employed	17.7	13.6	.12
Married	37.4	54.5	<.001
Blood pressure (% ≥140/90)	53.6	56.6	.41
Blood pressure (% ≥160/100)	19.4	14.5	.07
Benign prostatic hypertrophy	33.5	30.6	.40
Coronary artery disease	43.0	52.9	.006
Congestive heart failure	20.4	24.3	.19
Cerebrovascular disease	20.0	22.5	.39
Diabetes	51.5	48.7	.42
Hyperlipidemia	64.4	69.1	.17
Peripheral vascular disease	26.3	23.4	.36
Renal disease	31.3	16.5	<.001
Tobacco use (% yes)	21.3	14.7	.02
Body mass index (mean)	30.1	30.9	.08

Table 2. Bivariate Results—Racial Differences in Experiences with Providers, and Beliefs and Attitudes Regarding High Blood Pressure (BP)

Item text	African American (AA)	White	*p* value
Experiences with BP care/Patient Exit Interview items			
PEI overall score	6.6	5.8	.01
Did your primary care provider...			
– talk to you about your BP and medications for BP during your appointment? (% yes)	86.6	83.4	.23
– ask if you take your BP medication as prescribed?	88.2	82.4	.04
– discuss how important your BP medication is for controlling your BP?	84.6	78.0	.03
– discuss other health problems that might develop if someone does not take their BP medication?	67.8	52.6	<.001
– advise you to take your BP medication as prescribed?	90.3	88.8	.54
– discuss your efforts to manage your BP medication?	77.4	73.6	.28
– discuss things that get in the way of taking your BP medication?	34.6	31.0	.35
Did you and your provider...			
– discuss things you can do to make it easier to take your BP medication?	39.2	27.8	.004
– discuss any specific goals to help you take your BP medication as prescribed?	37.4	34.4	.46
– agree on any specific goals for taking your BP medication?	57.3	54.8	.53
Did your provider ask you to make another appointment to discuss your BP?	59.5	67.2	.05
Did your provider give you any written materials about BP during your appointment?	12.4	6.7	.02
Beliefs about BP and BP medications			
Do you believe that taking BP medication...(% yes)			
will make you feel worse?	5.1	3.9	.46
will make you feel better?	96.3	91.7	.008
will help you live longer?	98.5	97.4	.27
will improve the quality of your life?	95.4	93.4	.25
will prevent future high BP related illnesses?	96.0	94.6	.38
How serious do you think high BP is, in general?	Extr. serious=73.0% Quite serious=17.2 Mod. serious=6.4 Little serious=2.2 Not serious=1.2	Extr. serious=59.9% Quite serious=25.1 Mod. serious=11.4 Little serious=2.6 Not serious=1.0	.003
How serious do you think your high BP is, given your current use of medication?	Extr. serious=31.8% Quite serious=19.9 Mod. serious=28.7 Little serious=10.3 Not serious=9.3	Extr. serious=18.3% Quite serious=17.3 Mod. serious=37.3 Little serious=11.1 Not serious=16.0	<.001
If you did not take your BP medication, how likely do you think it would be that you would develop other health problems over the next year?	Very likely=71.3% Likely=11.6 50–50=12.6 Unlikely=2.4 Very unlikely=2.2	Very likely=64.8% Likely=19.1 50–50=12.4 Unlikely=2.7 Very unlikely=1.0	.63
If you did not take your BP medication, how likely do you think it would be that your BP would get worse over the next year?	Very likely=75.9% Likely=8.7 50–50=12.1 Unlikely=0.7 Very unlikely=2.7	Very likely=79.2% Likely=9.9 50–50=8.9 Unlikely=1.0 Very unlikely=1.0	.09
How would you describe your BP currently?	Good control=78.8% High=20.2 No longer a prob=0.9	Good control=81.1% High=14.4 No longer a prob=4.6	.69
BP medication adherence			
Did you ever take less medicine because you felt you needed less? (% yes)	12.6	8.8	.11
Sometimes if you feel worse when you take the medicine, do you stop taking it? (% yes)	10.2	5.5	.02
How many days in the past week did you forget to take your BP medication?	0 days=78.9% 1 day=10.7 2+ days=10.4	0 days=86.4% 1 day=7.5 2+ days=6.2	.01
How many days in the past week did you not take your medication on purpose?	0 days=90.7% 1 day=4.3 2+ days=5.0	0 days=95.5% 1 day=1.6 2+ days=2.9	.04
How many days in the past week did you add an extra pill?	0 days=95.7% 1 day=3.1 2+ days=1.2	0 days=97.1% 1 day=1.6 2+ days=1.3	.54
Some people have difficulty taking BP medication as prescribed. Do you have difficulty with this? (% yes)	6.9	5.5	.45
Summary adherence measure (% adherent)	74.2	79.8	.10
Covariates			
Optimism scale (higher=more optimistic; mean scores)	14.9	14.9	.89
Negative affectivity (higher=more negative affect; mean scores)	3.3	3.4	.53
VR-12 Physical component (higher score is better) (mean score)	40.1	38.0	.02
VR-12 Mental component (higher score is better) (mean score)	48.8	50.7	.03
Trust in Physician (higher=more trust)	78.2	82.5	<.001

Table 2. (Continued)

Item text	African American (AA)	White	p value
Racial discrimination (lower=more discriminated against)	57.5	91.6	<.001
Medication Regimen complexity			
Split pills	33.1	29.7	.33
# types of medications	2.2	1.9	.002
# times/day medications taken	1.4	1.5	.30
Family help taking medications (% yes)	36.6	30.5	.09
Self-efficacy			
Ability to take medications as prescribed?	9.5	9.6	.34
Knowledge about BP and how to deal with it?	8.8	8.4	.01

including all significant independent variables from the bivariate analyses, as well as race and site of care, retaining items significant at the $p<.05$ level. In the final step, using PROC LOGISTIC for the binary adherence outcome, we included all significant variables from the prior stepwise analysis, into the multivariate regression model (race and site of care were included in the final model regardless of their significance in prior analyses). All analyses were conducting using SAS 9.1.3 (SAS Institute, Cary, NC).

RESULTS

White patients were older (67.3 vs 64.4,) and more likely to be married (54.5 vs 37.4, both p's<.0001; (Table 1)). African Americans had a lower frequency of coronary artery disease diagnoses (43% vs 53%), but nearly twice the rate of renal disease (31% vs 17%), and higher rates of tobacco use (21% vs 15%; all p's<.05). African-American and White patients were equally likely to have a blood pressure greater than 140/90 mmHg (54% vs 57%, $p=.41$). Race was differentially distributed across sites, with the proportion of African Americans ranging from 42% (St. Louis), to 56% (Philadelphia), to 77% (Chicago; 58% overall). To account for these differences across sites, site was included as a covariate in all models. Adherent patients were 1.5 times more likely to have controlled BP than nonadherent patients, after controlling for comorbid conditions, BMI, age, and site of care ($p=.0433$; results not shown).

There were important bivariate racial differences on each of the 3 primary dimensions assessed (Table 2).

Patient–provider interactions. African Americans (AA) reported that their providers were more active in counseling and advising them about BP, with a higher total patient exit interview score than Whites (6.6 vs 5.8, $p<.01$), indicating that AAs' providers asked about 1 more question on average regarding BP. However, Whites' (W) providers were more likely to request a follow-up appointment for BP care (67% vs 60%, $p=.05$). Whereas relatively high rates of patients of both races (~80%) reported that doctors discussed BP and its management with them, only about a third of patients reported that providers explicitly discussed barriers to BP medication adherence, and only about 10% of patients reported that their providers provided them with written materials about BP (although AA patients more often reported that their doctors had done so: 12% vs 7%). African Americans reported that their doctors more often discussed other health problems that might result from high BP (68% vs 53%), whether they take their medications as prescribed (88% vs 82%), how important medications are for controlling BP (85% vs 78%), and discussed things to do to make it easier to take BP medications (39% vs 28%; all p's<.05).

Patients' beliefs about BP and BP medications. Whereas there were some similarities in perceptions about BP medications, AAs were more likely to report that BP medications would make them feel better, and to rate high BP as a more serious health concern (p's$\leq .01$; Table 2). There were no significant racial differences in patients' perceptions of whether their BP is under good control.

Medication adherence. African Americans were less adherent on 3 of the 6 single items assessing adherence (more likely to stop taking medications if one feels worse; not taking medication on purpose, and forgetting to take medications [all p's<.05]; Table 2). However, such differences remained only at the trend level on the summary adherence variable (74% vs 80%, $p=.10$).

Covariates. With regard to medication regimen complexity, there were no racial differences in patients being told to split pills (on average, approximately 30% of both groups were told to do so) or in the number of times per day BP medications were taken. However, African Americans were taking 2.2 types of medications versus Whites' 1.9 ($p=.002$). African Americans felt more knowledgeable about their BP and how to deal with it, and had better physical and mental functioning. Finally, African Americans were more likely to have experienced racial

Table 3. Multivariate Logistic Regression Modeling Self-reported Medication Adherence

Independent variables	Odds ratio	95% CI	p value
White race vs. Black race	1.24	0.80–1.92	.33
How confident about ability to take BP meds as prescribed	1.41	1.20–1.67	<.001
Did you and provider discuss things you can do to make it easier to take your BP meds?	0.56	0.37–0.85	.006
Told to split any BP pills by MD or pharmacist?	0.58	0.37–0.90	.02
How would you describe your BP currently?			
Continues to be high vs. under good control	0.51	0.32–0.82	.006
No longer a problem vs. under good control	0.41	0.07–2.32	.32
Site of care			
Chicago vs. St. Louis	1.00	0.61–1.65	.99
Philadelphia vs. St. Louis	1.28	0.75–2.19	.37

discrimination and White patients indicated greater trust in their physicians (p's<.05).

Last, we calculated a multivariate logistic regression model for self-reported adherence (Table 3). Confidence in one's ability to take medications as prescribed was associated with better adherence (p<.0001). However, patient perception that one's BP continues to be high was associated with worse adherence (p=.006), as were having the provider discuss things the patient can do to make it easier to take BP medications (p=.006) and being told to split one's pills (p=.016).

DISCUSSION

This study is among the first to simultaneously examine racial differences in patients' experiences with physicians, attitudes and beliefs about hypertension and antihypertensive medication, and the association of each of these dimensions with patients' antihypertensive medication adherence. In our sample with similar BP control among Whites and African Americans, we found that African Americans believed BP to be a more serious health threat, suggesting that public health, VA system-wide, and clinician-provided messages about the significance of high BP, especially among African Americans, seem to be reaching their target. These results also suggest that African Americans are receiving enhanced BP care—their physicians counseled them more about blood pressure and prescribed more medications for their BP. Thus, these VA providers seem to have received and acted upon the message regarding poorer outcomes for African-American patients with hypertension. We conclude that such methods and strategies, when used by clinicians, may positively influence patients' beliefs about antihypertensive medications.

Our findings differ from previous reports about racial/ethnic differences in specific knowledge and beliefs about hypertension[23,24], and suggest that in the VA setting, African Americans perceive BP to be a more serious threat to health than do White patients. In contrast, whereas earlier reports had noted that White patients were more likely to be counseled about hypertension by their providers[25], our findings suggest the opposite dynamic is occurring in the VA.

Our study was limited in several ways. First, we only studied regular users of the VA care system, which may have biased our sample toward more adherent patients (e.g., appointment keepers). However, we still observed a range of adherence behaviors. Also, the VA system cares primarily for male patients, so our results may not be generalizable to women, or to patients in non-VA settings. Further, racial disparities in health care may be minimized in the VA setting[26], which could limit our ability to generalize these findings.

These results suggest that several patient health beliefs and practices are associated with adherence, although we were unable to determine the causal direction of the associations. The fact that a patient's understanding that his blood pressure was high was negatively associated with adherence, as were reports of providers counseling about medication taking suggests that providers may have been working harder with patients with harder-to-control BP, or with patients who did not seem to be adherent.

Notably, patients with greater perceived self-efficacy in medication taking had better adherence. Such beliefs can be fostered by primary care providers during clinic visits, using strategies developed for patient-centered counseling[12]. Clinicians can ask open-ended questions of their patients about medication adherence such as, "What kinds of problems are you having taking your blood pressure medications?" Then, using barriers identified by the patients, clinicians can help to strategize ways to address such barriers, thus enhancing patients' self-efficacy, and adherence. Indeed, our prior work has demonstrated that there is room for improvement in primary care clinicians' antihypertensive medication adherence counseling skills[27].

African Americans are disproportionately affected by hypertension, with lower rates of blood pressure control in the general population, although not in this VA sample. Notably, less than half of this sample overall had controlled BP, indicating much room for improvement in BP care. Our findings suggest that patient beliefs are significantly associated with blood pressure medication adherence. Thus, we encourage providers to actively learn more about their patients' beliefs about both hypertension and its therapies, to provide targeted counseling to help patients improve medication adherence, and ultimately, blood pressure control.

Acknowledgments: *This research was presented as a poster at the American Heart Association Compliance Conference, May 2005 (NR Kressin, MB Orner, F Wang, J Long, W Kozak, C Clark, S Reddy, L Kroupa, B Bokhour, J Rothendler, D Berlowitz. Racial Differences in Antihypertensive Medication Adherence, Attitudes, Beliefs and Experiences with Blood Pressure Care). This research was supported by grants from the Department of Veterans Affairs (VA) Health Services Research and Development Service (TRH01-038, N. Kressin, P.I). Dr. Kressin is a Research Career Scientist, Department of Veterans Affairs, Health Services Research & Development (RCS 02-066-1); Dr. Long was an Associate in the Career Development Award Program of the VA HSR&D Service when this work was performed (CDA # 00-023).*

Conflict of Interest: *No authors have any affiliation, financial agreement, or other involvement with any company whose product figured prominently in the submitted manuscript.*

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

Corresponding Author: *Nancy R. Kressin, PhD; Center for Health Quality, Outcomes and Economic Research (a VA Health Services Research and Development National Center of Excellence), Bedford VA Medical Center, Bedford, MA, USA (e-mail: nkressin@bu.edu).*

REFERENCES

1. Joint National Committee. The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. Arch Intern Med. 1997;157:2413–46.
2. **Wong M, Shapiro M, Boscardin W, et al.** Contributions of major diseases to disparities in mortality. N Engl J Med. 2002;347:1585–92.
3. **Cushman WC, Ford CE, Cutler JA, et al.** Success and predictors of blood pressure control in diverse North American settings: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). J Clin Hypertens (Greenwich). 2002;4(6):393–404.
4. **Monane M, Bohn RL, Gurwitz JH, Glynn RJ, Levin R, Avorn J.** Compliance with antihypertensive therapy among elderly Medicaid enrollees: the roles of age, gender, and race. Am J Public Health. 1996;86 (12):1805–8.
5. **Sharkness CM, Snow DA.** The patient's view of hypertension and compliance. Am J Prev Med. 1992;8:141–6.
6. **Dunbar-Jacob JM, Schlenk EA, Burke LE, Matthews JT.** Predictors of patient adherence: patient characteristics. In: Shumaker SA, Schron EB,

Ockene JK, McBee WL, eds. The Handbook of Health Behavior Change. 2nd ed. New York: Springer Publishing Company; 1998:491–512.

7. **Kressin NR, Petersen LA.** Racial variations in cardiac procedures: a review of the literature and prescription for future research. Ann Intern Med. 2001;135:352–66.
8. **Kleinman A, Eisenberg L, Good B.** Culture, illness, and care: clinical lessons from anthropologic and cross-cultural research. Ann Intern Med. 1978;88:251–8.
9. **Taylor DW, Sackett DL, Johnson AL, Gibson ES, Roberts RS.** Compliance with antihypertensive drug therapy. Ann NY Acad Sci. 1978:390–403.
10. **Norman SA, Marconi KM, Schezel GW, Schechter CF, Stolley PD.** Beliefs, social normative influences, and compliance with antihypertensive medication. Am J Prev Med. 1985;1(3):10–7.
11. **Eraker SA, Kirscht JP, Becker MH.** Understanding and improving patient compliance. Ann Intern Med. 1984;100:258–68.
12. **Ockene IS, Hebert JR, Ockene JK, Merriam PA, Hurley TG, Saperia GM.** Effect of training and a structured office practice on physician-delivered nutrition counseling: the Worcester-Area Trial for Counseling in Hyperlipidemia. Am J Prev Med. 1996;12(4):252–8.
13. **Pbert L, Adams A, Quirk M, Hebert J, Ockene J, Luippold R.** The Patient exit interview as an assessment of physician-delivered smoking intervention: a validation study. Health Psychol. 1999;18(2):183–8.
14. **Rosenstock I.** Why people use health services. Milbank Mem Fund Q. 1966;44(94).
15. **Thom D, Ribisl K, Stewart A, Luke D, and the Stanford Trust Study Physicians.** Further validation and reliability testing of the Trust in Physician scale. Med Care. 1999;37(5):510–7.
16. **Krieger N.** Racial and gender discrimination: risk factors for high blood pressure? Soc Sci Med. 1990;30(12):1273–81.
17. **Miller NH, Hill M, Kottke T, Ockene IS.** The multilevel compliance challenge: recommendations for a call to action–a statement for health-care professionals. Circulation. 1997;95:1085–90.
18. **Eysenck HJ, Eysenck SB.** Manual of the Eysenck Personality Inventory. San Diego: Educational & Industrial Testing Service; 1968.
19. **Scheier M, Carver C, Bridges M.** Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test. J Pers Soc Psychol. 1994;67(6):1063–78.
20. **Kazis LE, Miller DR, Clark JA, et al.** Improving the response choices on the veterans SF-36 health survey role functioning scales: results from the Veterans Health Study. J Ambul Care Manage. 2004;27 (3):263–80.
21. **Choo PW, Rand CS, Inui TS, et al.** Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. Med Care. 1999;37 (9):846–57.
22. **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.
23. **Okonofua E, Cutler N, Lackland D, Egan B.** Ethnic differences in older americans: awareness, knowledge, and beliefs about Hypertension. Am J Hypertens. 2005;18:972–9.
24. **Hertz R, Unger A, Cornell J, Saunders E.** Racial disparities in hypertension prevalence, awareness, and management. Arch Intern Med. 2005;165:2098–104.
25. **Tschann JM, Adamson TE, Coates TJ, Gullion DS.** Behaviors of treated hypertensive patients and patient demographic characteristics. J Commun Health. 1988;13(1):19–32.
26. **Jha AK, Shlipak MG, Hosmer W, Frances CD, Browner WS.** Racial differences in mortality among men hospitalized in the Veterans Affairs health care system. JAMA. 2001;285(3):297–303.
27. **Bokhour BG, Long JA, Berlowitz DR, Kressin NR.** Assessing patient adherence in medical encounters: how do providers talk with patients about antihypertensive medication taking? J Gen Intern Med. 2006;21 (6):577–83.

doi:10.1111/j.1360-0443.2007.01759.x

Persistent pain is associated with substance use after detoxification: a prospective cohort analysis

Mary Jo Larson[1], Michael Paasche-Orlow[2], Debbie M. Cheng[2,5], Christine Lloyd-Travaglini[3], Richard Saitz[2,4] & Jeffrey H. Samet[2,6]

New England Research Institutes, Watertown MA, USA,[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] Data Coordinating Center, Boston University School of Public Health, Boston MA, USA,[3] Youth Alcohol Prevention Center, and Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA,[4] Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA[5] and Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA[6]

ABSTRACT

Aims To test the hypothesis that persistent pain is associated with an increased odds of substance use after detoxification. **Design** Analysis of data from a prospective cohort enrolled in a randomized controlled trial (RCT) to improve linkage with primary medical care. **Setting** An urban residential detoxification program. **Participants** Adults ($n = 397$) enrolled in the RCT with heroin, alcohol or cocaine as a substance of choice and at least one follow-up interview. **Measurements** The key independent variable was pain status: persistent pain (moderate to very severe pain at all available interviews), no pain (mild pain or less at all available interviews) and intermittent pain (all others). There were four outcomes of interest: self-reported use of any substance; heroin/opioid use; heavy alcohol use; and cocaine use 24 months after detoxification. Multivariable logistic regression controlled for several covariates including demographics, physical/sexual abuse, depressive symptoms, duration of follow-up and addiction severity at study entry. **Findings** Pain in detoxification patients was common; 16% had persistent pain and 54% had intermittent pain. Persistent pain was associated with an increased odds for use of any substance [adjusted odds ratio (AOR) 4.2, 95% confidence interval (CI) 1.9–9.3], heroin/opioid use (AOR 5.4, 95% CI 2.1–13.8) and heavy alcohol use (AOR 2.2, 95% CI 1.0–4.5) at the 24-month follow-up. A statistically non-significant increase in the odds of cocaine use (AOR 2.0, 95% CI 0.9–4.6) was also observed. **Conclusions** Among individuals leaving residential detoxification, chronic pain is a common problem and is associated independently with long-term substance use after detoxification. Addressing pain as a treatable chronic condition among adults receiving detoxification presents a potential opportunity to improve long-term clinical outcomes and warrants further intervention research.

Keywords Detoxification, medical comorbidity, outcomes, pain, prospective data, substance use.

Correspondence to: Mary Jo Larson, New England Research Institutes, 9 Galen Street, Watertown, MA 02472, USA. E-mail: mjlarson@neriscience.com
Submitted 28 November 2005; initial review completed 25 April 2006; final version accepted 26 November 2006

INTRODUCTION

A high prevalence of pain has been reported in cross-sectional studies of patients with drug use disorders [1–5], including those from methadone programs (37%) and short-term residential treatment programs (24%) [1]. Inadequate treatment of pain is a particularly challenging problem among patients with substance use disorders due to the concern that opiates for the treatment of pain may contribute to exacerbation of the substance use disorder [6,7]. Indeed, apprehension of promoting substance use has led to widespread under-treatment of pain and prejudicial care patterns [8].

There is some evidence that chronic severe pain influences the drug use patterns of patients in addiction treatment programs. Chronic pain is generally defined as unrelenting, lasting more than 6 months and often occurring after healing from disease or injury is complete [9,10]. Methadone and residential treatment patients with chronic severe pain are more likely than patients without pain to report use of illicit drugs to relieve pain [1]. Trafton and colleagues [4] suggest that substance abuse patients with chronic pain may have different

addiction treatment needs from other substance abuse patients. In their study of 251 veterans seeking methadone/levo-alpha-acetyl-methadol (LAAM) treatment, those with moderate or severe pain had greater misuse of non-prescribed analgesic medications than other patients, suggesting selective self-administration of substances that relieve pain [4].

Despite the high prevalence of pain, and the common use of illicit drugs to remedy it, there is scant literature on the impact of chronic pain on recovery from alcohol and drug dependence. In a retrospective study of veterans in residential substance abuse treatment, back pain was an independent factor contributing to a lower probability of completing treatment [4]. The presence of pain among adults in a national sample of drug abuse treatment programs was associated with poorer self-reported health status at follow-up 12 months after treatment terminated [11].

As unrelieved pain is related to continued illicit drug use and leads to negative medical and psychosocial consequences, it is possible that pain will also affect return to substance use after seeking care for one's addiction. However, we found no empirical prospective studies that have addressed whether chronic pain affects return to alcohol or drug use. In this study we examine whether patients with persistent pain seeking detoxification in an urban, residential program have different addictions-related outcomes in a prospective study. This study addresses two main questions: (1) what is the prevalence of persistent and intermittent pain over a 2-year period among adults receiving residential substance abuse detoxification; and (2) is persistent or intermittent pain associated with substance use 24 months after detoxification?

METHODS

Subjects

The study involved a prospective cohort of adults admitted to a freestanding urban residential alcohol and drug detoxification unit between 1 June 1997 and 1 April 1999. Patients were screened for eligibility and enrolled on their second day or later in the detoxification unit [12]. Inclusion criteria were the following: (1) heroin, alcohol or cocaine as the patient's first or second drug of choice; (2) age greater than 17 years; and (3) residence in proximity to a referral primary care clinic or homelessness. The exclusion criteria were: (1) an established primary care relationship that the patient intended to continue; (2) cognitive impairment limiting the subject's ability to provide pertinent history (score of less than 21 of 30 on the Mini-Mental State Examination) [13]; (3) specific plans to leave the area in the next 12 months; (4) inability to provide three contact names for follow-up tracking; (5) pregnancy; and (6) not fluent in English or Spanish. Of 642 eligible subjects, 470 provided consent and were enrolled in the cohort. All 470 participated in the randomized clinical trial, the Health Evaluation and Linkage to Primary care (HELP) study [12]. All subjects were assigned randomly to receive either standard medical care referral by clinical addictions staff on an as-needed basis (usual care) or an enhanced effort for referral to primary medical care. This enhanced effort and the results of the clinical trial are described elsewhere [12]. The results reported here are from prospectively collected data on subjects from the HELP sample who completed at least one follow-up interview and had no missing covariate information ($n = 388$). Of 470 subjects in the cohort, two died before any scheduled follow-up and 400/468 (85%) completed at least one interview during the 2-year follow-up period. Of the 400 subjects who completed at least one follow-up interview, 12 were missing a value on one or more covariates and were excluded from further analyses. White subjects were significantly less likely to be lost to follow-up than non-white subjects; otherwise, subjects in this analysis did not differ significantly from those lost to follow-up on any other study variable [14].

Assessments

After initial resolution of the symptoms of acute withdrawal, trained research associates interviewed subjects at the detoxification unit. Assessments included: demographics, substance of choice, substances used and addiction severity [Addiction Severity Index (ASI) alcohol and drug scales] [15], depressive symptoms [Center for Epidemiologic Studies Depression scale (CES-D)] [16], health-related quality of life [Short Form Health Survey (SF-36)] [17] and questions regarding primary medical care [12]. Except for demographics, all the assessments were repeated every 6 months over 24 months. At follow-up interviews, alcohol breath tests were performed to encourage truth telling [18]. Pain was measured by items of the SF-36, which focuses specifically on the severity of bodily pain and resulting limitations in activities [17].

Outcome variables

There were four dichotomous outcomes of interest based on type of substance use at the 24-month follow-up interview. The outcomes were: (1) heavy alcohol use (i.e. intoxication or three or more drinks on at least 1 day); (2) cocaine use; (3) heroin/opioid use (i.e. non-therapeutic—does not include prescribed pain medications or methadone); and (4) any substance use (i.e. heavy alcohol use or illicit use of cocaine or

opioids). Each outcome was self-reported and assessed using items from the alcohol and drug ASI regarding the most recent 30 days from the interview date [19]. Heavy alcohol use was dichotomized as 1 if intoxication or 3 or more drinks on at least 1 day was reported (else zero). Similarly, for cocaine, heroin and any drug use, the dichotomous measure took on a value of 1 if there was use on any 1 day in the past 30 days and 0 if there was no use in the past 30 days. The 24-month measurement was selected as the primary outcome because the interest was in substance use after a prolonged period following detoxification.

Main independent variable

The main independent variable of interest was a composite of the severity and duration of pain. The severity of pain was assessed using the SF-36 [17] pain item: 'how much bodily pain have you had during the past 4 weeks?'. One of six responses was possible: none, very mild, mild, moderate, severe and very severe. Three categories were created combining severity and duration of pain: (1) persistent pain (moderate or a higher level of pain reported at each available interview); (2) intermittent pain (moderate or a higher level of pain reported during at least one but not all available interviews); and (3) no pain (mild or a lower level of pain reported at each available interview). The SF-36 includes an additional question relating to pain. The second item is 'how much did pain interfere with your normal work (including both work outside the home and housework)?', with five possible responses ranging from not at all to extremely. The two pain-related questions are combined to form the SF-36 Bodily Pain scale (BP). Although we do not focus upon disruption of normal work due to pain in this paper, these data are also reported for descriptive purposes. For the BP scale, a lower score indicates worse pain and more limitation due to pain. The range is 0–100; the reported reliability coefficient is 0.90 in normative data for the general US population [17] and is 0.87 in this study sample.

Other predictor variables

Additional predictor variables of interest measured at time of study entry were the following: gender; age; health insurance in the 6 months prior; unemployed in 6 months prior; self-report of life-time physical/sexual abuse; highest level of education (high school or greater); health literacy score from the Rapid Assessment of Adult Literacy in Medicine (REALM), a 66-item literacy screening instrument based on word pronunciation [20]; race/ethnicity (white or non-white); homeless (1 or more nights in a shelter or on the street in the preceding 6 months); severity of alcohol and drug dependence as measured by the ASI alcohol and drug scores (ASI-alcohol and ASI-drug, respectively); indicator variables for substance of choice (heroin, alcohol or cocaine); the mental component summary (MCS) score of the SF-36; depressive symptoms (CES-D) [16]; and indicator variable for told by a doctor they had one of 13 chronic conditions [21]. Duration of follow-up was measured as 6, 12, 18 or 24 to reflect the month of the last completed survey.

Analysis

Two analytical samples were used to assess the relationship of variables with substance use outcomes. The primary analysis included 397 subjects (3 of 400 had missing pain variable): 275 who completed the 24-month interview and 122 subjects for whom the substance use outcome obtained at the last follow-up visit was carried forward for the 24-month outcome measurements; due to item missing responses, the analytical sample was 388. A secondary analysis was conducted that included only participants who completed the 24-month follow-up interview.

Bivariate associations between pain status and baseline characteristics and pain status and substance use outcomes 24 months after detoxification were assessed using analysis of variance (ANOVA) for continuous measures and the χ^2 test for categorical measures.

Multiple logistic regression models were used to assess the association between pain status and substance use 24 months after detoxification. Separate models were fitted for each of the four outcome measures. The primary and secondary analyses were adjusted for age, gender, race, unemployment, health insurance, experienced physical or sexual abuse, depressive symptoms, alcohol and drug severity at study entry and duration of follow-up. Regression models did not control for the presence of a chronic medical condition as these conditions were hypothesized to be a probable source of chronic pain—we were interested in whether pain was related to substance use, not whether that relationship would be moderated by the cause of the pain. Reported *P*-values are two-tailed, and a *P*-value less than 0.05 was considered statistically significant. Analyses were carried out using SAS/STAT software, version 8.2. [22].

Two confirmatory analyses were conducted to assess the sensitivity of results to using alternative measures for the following covariates: alcohol and drug use at study entry (drug of choice instead of ASI alcohol and drug scores); mental health status (SF-36 MCS instead of depressive symptoms). Additional models were also fitted to assess whether homelessness, health literacy or educational level were confounding factors. A final exploratory analysis was conducted to assess whether adjusting for substance use outcome at 6-month follow-up changed

Table 1 Demographic characteristics of residential detoxification patients by pain group ($n = 397$).

Characteristic	*All subjects*	*Pain category*[1] *Persistent pain*	*Intermittent pain*	*No pain*	*Stat. sign.**
n	397	64	214	119	
Gender (%)					$P < 0.001$
Male	75.6	14.3	50.7	35.0	
Female	24.4	21.7	63.9	14.3	
Age, mean (SD)	36.0 (7.7)	39.0 (8.4)	36.1 (7.6)	34.2 (7.1)	$P < 0.001$
Health insurance (%)					$P < 0.0001$
Yes	39.3	25.6	55.8	18.6	
No	60.7	10.0	52.7	37.3	
Unemployed past 6 months (%)					$P < 0.001$
Yes	38.0	21.2	59.6	19.2	
No	62.0	13.0	50.4	36.6	
Physical/sexual abuse (%)					$P = 0.02$
Yes	71.1	16.4	57.5	26.1	
No	28.9	14.9	44.7	40.4	
Education (%)					$P = 0.868$
Less than high school	29.9	15.3	55.9	28.8	
High school graduate	70.1	16.6	53.1	30.3	
Health literacy (%)					$P = 0.20$
Low (0–44)	13.8	11.5	53.9	34.6	
Moderate (45–60)	32.1	13.2	50.4	36.4	
High (61–66)	54.1	19.6	54.4	26.0	
Race/ethnicity (%)					$P < 0.0001$
White	33.8	28.3	47.8	23.9	
Non-white	66.2	9.9	57.0	33.1	
Homeless (%)					$P = 0.09$
Yes	46.8	20.4	51.1	28.5	
No	53.2	12.3	56.4	31.3	
Duration of follow-up, median months	23.3	22.0	23.5	23.4	$P = 0.02$

*F-statistic for means; Kruskall–Wallis test for medians. [1]Based on all interviews in a 24-month period.

the association between pain status and substance use outcomes.

RESULTS

Subject characteristics

Of the 397 subjects eligible for this prospective cohort study, the majority was male (76%), lacked health insurance (61%), was employed in the 6 months prior to study entry (62%), had experienced physical or sexual abuse (71%), completed a high school education or more (70%), identified as black, Hispanic or another minority group (66%) and demonstrated health literacy equivalent to 9th grade or above (i.e. REALM score 61 or greater) (54%). Mean age was 36.0 (SD 7.7). More than 46% had at least 1 day of homelessness (on the street or in a shelter) in the 6 months prior to study entry (Table 1).

Almost half the subjects reported at least one chronic medical condition (48%) (Table 2). The substance of choice was heroin for about one-quarter of subjects, alcohol for more than one-third of subjects and cocaine/crack for one-third of subjects. Average ASI-alcohol and drug composite scores at baseline were 0.47 and 0.26, respectively. Measures of mental health status were indicative of high depressive symptoms (CES-D, mean 33.1, SD 12.5), where > 16 is suggestive of a possible depressive disorder, and low mental health status (SF-36 MCS mean 31.5, SD 12.7) where the US population norm is 50.

Several bivariate associations were statistically significant between baseline demographic characteristics and pain status (i.e. persistent, intermittent, no pain) (Table 1). Pain status differed by all demographic characteristics except education, homelessness (marginal significance) and health literacy. Subjects classified with persistent pain were more likely than other subjects to be

Table 2 Health indicators and main outcome variables: residential detoxification patients by pain group ($n = 397$).

Characteristic	*All subjects*	*Pain category*[4] *Persistent pain*	*Intermittent pain*	*No pain*	*Stat. sign.*
n	397	64	214	119	
Pain characteristics					
Pain at baseline (%)					$P < 0.0001$
None	18.9	–	30.7	69.3	
Very mild	14.9	–	27.1	72.9	
Mild	10.8	–	44.2	55.8	
Moderate	25.9	22.3	77.7	–	
Severe	21.9	34.5	65.5	–	
Very severe	7.6	36.7	63.3	–	
SF-36 pain index, mean (SD)	53.1 (29.9)	27.1 (13.6)	44.6 (26.4)	82.3 (16.5)	$P < 0.0001$
Other health condition characteristics					
Any chronic condition (%)					$P < 0.0001$
Yes	48.1	25.7	55.5	18.9	
No	51.9	7.3	52.4	40.3	
CES-D (depressive symptoms), mean (SD)	33.1 (12.5)	38.4 (11.9)	34.4 (12.4)	28 (11.1)	$P < 0.0001$
SF-36 mental component summary, mean (SD)	31.5 (12.7)	28.0 (10.6)	29.7 (12.5)	36.6 (12.5)	$P < 0.0001$
Substance use characteristics					
First substance of choice (%)					$P < 0.0001$
Heroin	27.3	14.8	66.7	18.5	
Cocaine/crack	34.9	8.7	49.3	42.0	
Alcohol	37.9	24.0	48.7	27.3	
ASI–alcohol (mean)	0.473 (0.34)	0.538 (0.33)	0.459 (0.36)	0.462 (0.31)	$P = 0.25$
ASI–drug (mean)	0.259 (0.15)	0.219 (0.16)	0.272 (0.14)	0.256 (0.14)	$P = 0.04$
Outcome measures[1]					
Any substance use[2] (%)					$P < 0.0001$
Yes	59.6	21.6	55.1	23.3	
No	40.4	8.1	51.9	40.0	
Heroin/opioid use (%)					$P < 0.001$
Yes	22.5	24.7	60.7	14.6	
No	77.5	13.7	51.8	34.5	
Heavy alcohol use[3] (%)					$P = 0.054$
Yes	43.4	20.9	52.9	26.2	
No	56.6	12.5	54.5	33.0	
Cocaine use (%)					$P = 0.27$
Yes	26.0	20.4	54.4	25.2	
No	74.0	14.7	53.6	31.7	

[1]Use is measured in the past 30 days at last follow-up interview in 2-year period; [2]any substance use is heavy alcohol use or other drug use; [3]heavy alcohol use is alcohol to intoxication or three or more drinks on at least 1 day; [4]based on all interviews in a 24-month period.

female, older, white, unemployed, report physical/sexual abuse and have health insurance.

A total of 275 (69%) subjects had the last observation measured at the 24-month follow-up interview; 24-month completers were 67% of the no pain group, 73% of the intermittent pain group, and 59% of the persistent pain group. The median months of follow-up among participants of each pain group were 23.4, 23.5 and 22.0 months for the no pain, intermittent pain, and persistent pain groups, respectively. There was a statistically significant difference in the median follow-up across the three groups (Kruskal–Wallis test, χ^2 7.7, $P = 0.02$).

Pain and health indicators

The majority of study subjects reported moderate or greater pain at study entry (55%) (Table 2). Consistent with reports of greater pain severity, the SF-36 BP scale mean score was very low (53.1, SD 29.9) relative to the US norm for adults ages 45–54 years (73.1, SD 24.0), the

Table 3 Multivariable logistic regression adjusted odds ratios (confidence intervals): substance use[1] at 24 months ($n = 388$).

	Substance use outcomes			
Predictors	*Any substance*[2]	*Heavy alcohol*[3]	*Heroin*	*Cocaine*
Pain category[4]				
Persistent pain	4.21 (1.90–9.33)**	2.15 (1.03–4.51)*	5.36 (2.09–13.75)**	2.05 (0.91–4.62)
Intermittent pain	1.74 (1.05–2.87)*	1.35 (0.81–2.27)	2.57 (1.24–5.32)*	1.12 (0.63–2.01)
No pain (reference)	–	–	–	–
Female	0.55 (0.32–0.95)*	0.45 (0.25–0.79)**	0.64 (0.32–1.28)	1.11 (0.61–2.02)
Unemployed past 6 months	1.47 (0.92–2.36)	1.38 (0.86–2.21)	1.39 (0.79–2.46)	0.88 (0.52–1.47)
Health insurance past 6 months	1.67 (1.05–2.66)*	1.25 (0.79–1.98)	1.24 (0.70–2.22)	1.03 (0.61–1.72)
Physical/sexual abuse history	1.03 (0.63–1.68)	1.29 (0.79–2.11)	0.96 (0.52–1.75)	1.32 (0.76–2.29)
Race/ethnicity white	1.37 (0.86–2.20)	1.32 (0.82–2.12)	1.59 (0.89–2.84)	0.70 (0.41–1.21)
Age (years)	0.98 (0.95–1.01)	0.96 (0.93–0.99)**	0.98 (0.95–1.02)	0.99 (0.96–1.03)
Depressive symptoms (CES-D)	1.00 (0.98–1.02)	1.00 (0.98–1.01)	1.01 (0.98–1.03)	1.00 (0.98–1.02)
ASI–alcohol (standardized)[5]	1.09 (0.87–1.37)	1.72 (1.36–2.18)**	0.55 (0.41–0.73)**	1.06 (0.83–1.35)
ASI–drug (standardized)[5]	1.24 (0.98–1.56)	0.95 (0.75–1.19)	2.24 (1.59–3.15)**	1.66 (1.26–2.18)**
Duration of follow-up	1.00 (0.97–1.04)	1.00 (0.96–1.03)	0.99 (0.94–1.03)	1.02 (0.98–1.06)

**$P < 0.01$; *$P < 0.05$. [1]Use is measured in the past 30 days at last follow-up interview in 2-year period; [2]any substance use is heavy alcohol use or any other drug use; [3]heavy alcohol use is alcohol to intoxication or three or more drinks on at least 1 day; [4]based on all interviews in a 24-month period; [5]ORs are per standard deviation increase in ASI.

general US population (75.2, SD 23.7), and comparable to the published US norm for a sample of patients with clinical depression (58.4, SD 26.7) [17]. Persistent pain was observed in 16% of subjects.

The observed proportions of subjects reporting substance use 24 months after detoxification were 43% for return to heavy alcohol use, 26% for cocaine use, 23% for heroin/opioid use and 60% for any illicit drug or heavy alcohol use.

Bivariate associations also existed between pain status and several health status indicators measured at study entry. Subjects with at least one chronic medical condition, those reporting alcohol or heroin as the first substance of choice, and those with more depressive symptoms (CES-D) and worse mental health (SF-36 mental component summary score), were more likely to report pain (Table 2).

In unadjusted analyses (Table 2), pain status was associated with any substance use and heroin/opioid use outcomes. Of subjects with persistent pain, 80% reported any substance use, 34% heroin/opioid use, 56% heavy alcohol use and 33% cocaine use (not mutually exclusive) (data not shown). Relative to the group with no pain, the persistent pain group had a significantly increased odds of any substance [odds ratio (OR) 4.6, confidence interval (CI) 2.3–9.3], heroin/opioid (OR 4.3, CI 1.0–9.3) and heavy alcohol (OR 2.1, CI 1.1–3.9) use; a non-significant increase in the odds of cocaine use (OR 1.7, CI 0.9–3.4) was also observed. Relative to the group with no pain, the intermittent pain group had an increased odds of any substance use (OR 1.8, CI 1.2–2.9) and heroin/opioid use (OR 2.8, CI 1.4–5.3).

Adjusted analyses

In adjusted models, the association between pain status and outcomes remained (Table 3). Relative to the group with no pain, the persistent pain group had a significantly increased odds of any substance use (adjusted odds ratio (AOR) 4.2, CI 1.9–9.3), heroin or other opioid use (AOR 5.4, CI 2.1–13.8) and heavy alcohol use (AOR 2.1, CI 1.0–4.5); a non-significant increase in the odds of cocaine use (AOR 2.05, CI 0.9–4.6) was also observed. Relative to the group with no pain, the intermittent pain group had an increased odds of any substance use (AOR 1.7, CI 1.1–2.9) and heroin/opioid use (AOR 2.6, CI 1.2–5.3) but not heavy alcohol use (AOR 1.3, CI 0.8–2.3) or cocaine use (AOR 1.1, CI 0.6–2.0).

Females had reduced odds of heavy alcohol use and any substance use. Having health insurance was associated with increased odds of any substance use. Older age was associated with reduced odds of heavy alcohol use.

The findings were not sensitive to other model specifications that were tested. Models fitted with alternative measures of alcohol and drug use and mental health status at study entry and models including omitted variables (homelessness, health literacy and educational level) produced comparable results for the association of pain status. In addition, when analyses were repeated on the subset of subjects who completed the 24-month follow-up interview, the main findings did not change from the primary analysis. The data are not presented for these additional covariate and confirmatory analyses. In a final exploratory model that added the 6-month substance use outcome as a covariate (resulting in a reduced

sample size of 295), the association between pain status and long-term outcome was attenuated for all outcomes. In the adjusted models with the 6-month covariate, persistent pain remained significantly associated with any substance use (AOR 3.4, CI 1.2–10.1), but was no longer associated significantly with heroin or other opioid use (AOR 1.8, CI 0.6–5.9) or heavy alcohol use (AOR 1.7, CI 0.6–4.5). The intermittent pain group had non-significantly increased odds of any substance use (AOR 1.45, CI 0.8–2.7) and heroin or other opioid use (AOR 1.1, CI 0.5.–2.5).

DISCUSSION

More than half the urban adult population seeking detoxification from drugs and alcohol reported moderate to very severe pain upon entry to a residential program, and for about one-third of those with pain reports the pain was persistent for up to 2 years. The extent to which the presence of persistent pain or even intermittent pain was associated with an increase in the odds of substance use after detoxification was notable. While it has been reported previously that pain is prevalent in substance using addiction treatment populations, little prior empirical evidence demonstrated that chronic pain has an independent role in hindering recovery from substance dependence in patients seeking help.

These findings are consistent with the high level of pain reports in studies conducted in veterans' methadone treatment clinics (52% reported moderate/severe pain) [4] and non-veterans' methadone clinics (61%) [3] and we extend prior knowledge with prospective data collection documenting that for one-third of the population moderate/severe pain continues over a 2-year period.

Some pain reports while in detoxification care may be associated with withdrawal and not require any further intervention. The withdrawal hypothesis was explored explicitly in samples entering methadone and residential programs by Rosenblum *et al.* [1], and withdrawal was not found to be a major contributor to pain prevalence. In this study, the mean bodily pain score did improve during the 6 months immediately after detoxification. However, mean bodily pain scores from 6 months to 24 months were stable (between 66 and 69) and thus remained much more severe than the normative pain level of the US population of similar age. In the long term, pain levels were less severe than two US samples with published normative data (back pain and depression).

It has been suggested previously that co-occurring pain and substance use disorders could complicate the treatment of substance use disorders [4,23]; this study provides empirical evidence in support of this idea. There are other reasons that treatment centers and detoxification programs should be aware of pain problems and seek adequate care for patients with moderate and severe pain. Severe pain is indicative of other chronic disorders requiring treatment such as HIV, and can contribute independently to reduced functioning among those with chronic medical conditions [24,25]. In a longitudinal study of community-dwelling older adults, problem drinkers reported more frequent use of alcohol to manage pain than did non-problem drinkers and had more chronic health problems and recent serious injury at 3-year follow-up [26].

The current clinical practice literature in this area focuses on the management of acute and chronic pain among patients with a history of opioid abuse, and the significant risk of recurrent opioid abuse in such patients when opioids are part of medical treatment [6,7,27,28]. The findings of the current study highlight an additional countervailing risk: namely, that under-treatment of pain may itself be a significant factor in recurrent opioid abuse. Further, these findings suggest that chronic pain is an issue that should also be addressed among alcohol-dependent people, as addiction treatment outcomes were observed to be worse due to persistent pain for both opioid and alcohol use.

Several limitations of the study should be noted. First, we do not address, nor did we ask subjects about, the source of pain. Understanding the source of pain is important for tailoring an appropriate clinical response and may influence the odds of substance use after detoxification. We also could not incorporate additional measures relevant to the management of chronic pain. Insights gained from instruments such as the Treatment Outcomes of Pain Survey or brief pain index [29,30] can be useful to guide management as can measures for ascertaining patients' perceptions of whether their drug use was motivated in part by seeking pain relief [1]. We do not believe that opioid 'pseudoaddiction' (i.e. mistakes in presuming patients have addiction when indeed they are using drugs to seek adequate pain management) explains these findings, in part because we found increased likelihood of any substance use and heavy alcohol use after detoxification, not simply heroin use.

Secondly, there is a complex relationship that exists between pain, early return to substance use and long-term substance use that we were not able to study thoroughly. Exploratory analyses we conducted support the contention that return to substance use at 6-month follow-up may, in part, mediate the association between persistent and intermittent pain with long-term substance use. In addition, hyperalgesia might explain the high prevalence of both chronic pain and long-term substance use. There also are other plausible covariates not captured in this analysis. For example, there is a high prevalence of co-occurring anxiety, pain and substance use; while this study did capture childhood sexual and

physical abuse as a stressor and included a measure of depressive symptoms, we did not measure anxiety symptoms [2,31,32].

Thirdly, because the primary analysis carried forward the substance use reports from the last observation for 122 of the 397 sample observations, it is plausible that differential loss to follow-up influenced study findings. However, we do not expect that the association of duration of follow-up and pain categories explains these findings. In addition to adjusting for duration of follow-up in the primary analyses, the results from a secondary analysis that included only subjects who completed the 24-month follow-up were not substantially different from the main findings. Fourthly, it is possible that we misclassified pain status for subjects who did not complete all four possible follow-up interviews. For example, subjects who completed three interviews and were categorized as having persistent pain could have been classified as having intermittent pain if they had completed four interviews. Misclassification of pain status, as described, would bias the findings towards the null hypothesis. Further, we did not ask subjects if they had been treated for pain symptoms, so we cannot assess directly whether the high prevalence of pain reflects the absence or inadequacy of pain treatment. However, it is possible that most subjects were not involved with any regular source of pain treatment. At study entry all of the cohort lacked a source of primary medical care; while the majority linked with a primary medical care source during the course of the study [14], 36.8% received no primary care visits and 13.5% received a total of only one visit in the 2-year period [33]. Fifthly, the subjects in this study may not generalize to other groups of substance abuse treatment samples. In particular, the rate of depressive symptoms in this sample may be higher than in other substance abuse samples. The mean CES-D score in this sample is similar to one study where all patients had comorbid psychiatric and substance use disorders [34], but much higher than a second study of 12 residential treatment programs [35]. In this second study symptom reports may have been lower as measurement occurred 15 days after admission [36], rather than in the first few days as with this study. Finally, although the study was prospective, using these data we cannot determine whether the presence of chronic pain contributed to the development of an addictive disorder or whether the history of drug use behaviors contributed in some way to risk of developing painful conditions, as can occur with HIV/AIDS, chronic venous insufficiency or hyperalgesia. None the less, regardless of this etiological question, the presence of pain was a risk factor for substance use after detoxification.

The findings that persistent pain among people with substance use disorders is associated with recurrent alcohol and opioid use suggest clinical approaches that could potentially enhance the effectiveness of substance use services. The implication of this study is that clinicians and researchers should pay additional attention to the management of chronic pain for individuals with substance dependence. Treatment centers, detoxification programs and primary care physicians should be aware of pain problems and seek adequate care for patients with moderate to very severe pain. Furthermore, the current study provides new empirical evidence that co-occurring pain and substance use disorders could complicate the treatment of substance use disorders. Chronic pain would be a fruitful area for future research to confirm these findings in other treatment settings and consider interventions to address the problem. Addressing the comorbidities of pain and substance dependence will require coordination of care between general medical and substance use providers. In many cases this will involve linking patients with substance use disorders with primary medical care or other care providers in which adequate resources and expertise for chronic pain management exist. One successful intervention for people with concurrent chronic pain and a substance use disorder integrated cognitive–behavioral treatment of pain management and relapse prevention [37]. Attending to a patient's pain makes good medical sense and is one avenue by which to potentially improve the quality of addiction treatment.

Acknowledgements

Preliminary results were presented at the June 2004 annual national meeting of the College on Problems of Drug Dependence, San Juan Puerto Rico. Data management was performed by the Data Coordinating Center, Boston University School of Public Health. The National Institute on Alcohol Abuse and Alcoholism (R01-AA10870) and the National Institute on Drug Abuse (R01-DA10019) provided support for the study. This work was supported in part by General Clinical Research Center grant M01-RR00533 from the National Center for Research Resources.

References

1. Rosenblum A., Joseph H., Fong C., Kipnis S., Cleland C., Portenoy R. K. Prevalence and characteristics of chronic pain among chemically dependent patients in methadone maintenance and residential treatment facilities. *JAMA* 2003; **289**: 2370–8.
2. Trafton J. A., Oliva E. M., Horst D. A., Minkel J. D., Humphreys K. N. Chronic pain and medical, psychiatric and drug abuse problems in opioid dependent veterans (Abstract 2023). *Health Services Research and Development 2003 National Meeting: Diverse Veteran Populations: Challenges and Opportunities.* Washington, DC: Center for Mental Healthcare and Outcomes Research; 2003. Available from: http://www.hsrd.research.va.gov/about/national_meeting/2003/abstracts/2023.html [accessed 28 May 2005].

3. Jamison R. N., Kauffman J., Katz N. P. Characteristics of methadone maintenance patients with chronic pain. *J Pain Symptom Manage* 2000; **19**: 53–62.
4. Trafton J. A., Oliva E. M., Horst D. A., Minkel J. D., Humphreys K. Treatment needs associated with pain in substance use disorder patients: implications for concurrent treatment. *Drug Alcohol Depend* 2004; **73**: 23–31.
5. Chandragiri S., Vael B. J., King S., Hardy T. Chronic pain in veterans with substance abuse and posttraumatic stress disorder: a hidden diagnosis. *Pain Med* 2000; **1**: 202.
6. Compton P., Athanasos P. Chronic pain, substance abuse and addiction. *Nurs Clin North Am* 2003; **38**: 525–37.
7. Alford D. P., Compton P., Samet J. H. Acute pain management for patients receiving maintenance methadone or buprenorphine therapy. *Ann Intern Med* 2006; **144**: 127–34.
8. Green C. R., Anderson K. O., Baker T. A., Campbell L. C., Decker S., Fillingim R. B. *et al.* The unequal burden of pain: confronting racial and ethnic disparities in pain. *Pain Med* 2003; **4**: 277–94.
9. Jamison R. N. *Mastering Chronic Pain: a Professional's Guide to Behavioral Treatment.* Sarasota, FL: Professional Resource Press; 1996.
10. Merskey H., Bogduk N. *Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms*, 2nd edn. Seattle, WA: IASP Press; 1994.
11. Friedmann P. D., Lemon S. C., Anderson B. J., Stein M. D. Predictors of follow-up health status in the Drug Abuse Treatment Outcome Study (DATOS). *Drug Alcohol Depend* 2003; **69**: 243–51.
12. Samet J. H., Larson M. J., Horton N. J., Doyle K., Winter M., Saitz R. Linking alcohol and drug dependent adults to primary medical care: a randomized controlled trial of a multidisciplinary health evaluation in a detoxification unit. *Addiction* 2003; **98**: 509–16.
13. Folstein M. F., Folstein S. E., McHugh P. R. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; **12**: 189–98.
14. Saitz R., Larson M. J., Horton N. J., Winter M., Samet J. H. Linkage with primary medical care in a prospective cohort of adults with addictions in inpatient detoxification: room for improvement. *Health Serv Res* 2004; **39**: 587–606.
15. McLellan A. T., Luborsky L., Woody G. E., O'Brien C. P. An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index. *J Nerv Ment Dis* 1980; **168**: 26–33.
16. Radloff L. S. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Measure* 1977; **1**: 385–401.
17. Ware J. E. Jr. *SF-36 Health Survey: Manual and Interpretation Guide.* Boston, MA: The Health Institute—New England Medical Center; 1993.
18. Welte J. W., Perry P., Longabaugh R., Clifford P. R. An outcome evaluation of a hospital-based early intervention program. *Addiction* 1998; **93**: 573–81.
19. McLellan A. T., Kushner H., Metzger D., Peters R., Smith I., Grissom G. *et al.* The fifth edition of the Addiction Severity Index. *J Subst Abuse Treat* 1992;9: 199–213.
20. Davis T. C., Long S. W., Jackson R. H., Mayeaux E. J., George R. B., Murphy P. W., *et al.* Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Fam Med* 1993; **25**: 391–5.
21. De Alba I., Samet J. H., Saitz R. Burden of medical illness in drug and alcohol dependent persons without primary care. *Am J Addict* 2004; **13**: 33–45.
22. *SAS/STAT Software*, version 8.2. *SAS System for Windows.* Cary, NC: SAS Institute, Inc.; 2001.
23. Savage S. R. Preface: pain medicine and addiction medicine—controversies and collaboration. *J Pain Symptom Manage* 1993; **8**: 254–6.
24. Pieper B., Templin T. Lower extremity changes, pain, and function in injection drug users. *J Subst Abuse Treat* 2003; **25**: 91–7.
25. Marcus K. S., Kerns R. D., Rosenfeld B., Breitbart W. HIV/AIDS-related pain as a chronic pain condition: implications of a biopsychosocial model for comprehensive assessment and effective management. *Pain Med* 2000; **1**: 260–73.
26. Brennan P. L., Schutte K. K., Moos R. H. Pain and use of alcohol to manage pain: prevalence and 3-year outcomes among older problem and non-problem drinkers. *Addiction* 2005; **100**: 777–86.
27. Portenoy R. K., Payne R. Acute and chronic pain. In: Lowinson J. H., Ruiz P., Millman R. B., Langrod J. G., editors. *Substance Abuse: a Comprehensive Textbook*, 3rd edn. Philadelphia, PA: Williams & Wilkins; 1997, p. 563–89.
28. Savage S. R. Addiction in the treatment of pain: significance, recognition, and management. *J Pain Symptom Manage* 1993; **8**: 265–78.
29. Ho M. J., LaFleur J. The treatment outcomes of pain survey (TOPS): a clinical monitoring and outcomes instrument for chronic pain practice and research. *J Pain Palliative Care Pharmacother* 2004; **18**: 49–59.
30. Passik S. D., Kirsh K. L., Whitcomb L., Portenoy R. K., Katz N. P., Kleinman L. *et al.* A new tool to assess and document pain outcomes in chronic pain patients receiving opioid therapy. *Clin Ther* 2004; **26**: 552–61.
31. Green C. R., Flowe-Valencia H., Rosenblum L., Tait A. R. Do physical and sexual abuse differentially affect chronic pain states in women? *J Pain Symptom Manage* 1999; **18**: 420–6.
32. McWilliams L. A., Cox B. J., Enns M. W. Mood and anxiety disorders associated with chronic pain: an examination in a nationally representative sample. *Pain* 2003; **106**: 127–33.
33. Saitz R., Horton N. J., Larson M. J., Winter M., Samet J. H. Primary medical care and reductions in addiction severity: a prospective cohort study. *Addiction* 2004; **100**: 70–8.
34. Havassy B. E., Alvidrez J., Owen K. K. Comparisons of patients with comorbid psychiatric and substance use disorders: implications for treatment and service delivery. *Am J Psychiatry* 2004; **161**: 139–45.
35. Bobo J. K., McIlvain H. E., Leed-Kelly A. Depression screening scores during residential drug treatment and risk of drug use after discharge. *Psychiatr Serv* 1998; **49**: 693–5.
36. Brown R. A., Monti P. M., Myers M. G., Martin R. A., Rivinus T., Dubreuil M. E. *et al.* Depression among cocaine abusers in treatment: relation to cocaine and alcohol use and treatment outcome. *Am J Psychiatry* 1998; **155**: 220–5.
37. Currie S. R., Hodgins D. C., Crabtree A., Jacobi J., Armstrong S. Outcome from integrated pain management treatment for recovering substance abusers. *J Pain* 2003; **4**: 91–100.

PTSD in Urban Primary Care: High Prevalence and Low Physician Recognition

Jane Liebschutz, MD, MPH[1,2], Richard Saitz, MD, MPH[1,3,4], Victoria Brower, MPH[5], Terence M. Keane, PhD[6,7,8], Christine Lloyd-Travaglini, MPH[9], Tali Averbuch, MPP[1], and Jeffrey H. Samet, MD, MA, MPH[1,2]

[1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, 91 East Concord St., Suite 200, Boston, MA 02118, USA; [2]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA; [3]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; [4]Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA; [5]Massachusetts Behavioral Health Partnership, Boston, MA, USA; [6]Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA; [7]National Center for Post-Traumatic Stress Disorder, Boston, MA, USA; [8]Veterans Affairs Boston Healthcare System, Boston, MA, USA; [9]Data Coordinating Center, Boston University School of Public Health, Boston, MA, USA.

BACKGROUND: Posttraumatic stress disorder (PTSD) is associated with medical and psychological morbidity. The prevalence of PTSD in urban primary care has not been well described.

OBJECTIVE: To measure the prevalence of PTSD in primary care patients overall and among those with selected conditions (chronic pain, depression, anxiety, heavy drinking, substance dependence (SD), irritable bowel syndrome (IBS), and immigrant status).

DESIGN: Cross-sectional study.

PARTICIPANTS: English-speaking patients aged 18–65 years old, awaiting primary care appointments in an urban academic medical center, were eligible for enrollment to determine PTSD prevalence (N=509). Additional eligible participants (n=98) with IBS or SD were subsequently enrolled.

MEASUREMENTS: PTSD (past year) and trauma exposure were measured with Composite International Diagnostic Interview. We calculated the prevalence of PTSD associated with depression, anxiety, heavy drinking, SD, IBS, and chronic pain. Only the analyses on heavy drinking, SD, and IBS used all 607 participants.

RESULTS: Among the 509 adults in primary care, 23% (95% CI, 19–26%) had PTSD, of whom 11% had it noted in the medical record. The prevalence of PTSD, adjusted for age, gender, race, and marital and socioeconomic statuses, was higher in participants with, compared to those without, the following conditions: chronic pain (23 vs 12%, p=.003), major depression (35 vs 11%, p<.0001), anxiety disorders (42 vs 14%, p<.0001), and IBS (34 vs 18%, p=.01) and lower in immigrants (13 vs 21%, p=.05).

CONCLUSIONS: The prevalence of PTSD in the urban primary care setting, and particularly among certain high-risk conditions, compels a critical examination of optimal approaches for screening, intervention, and referral to PTSD treatment.

KEY WORDS: underserved populations; PTSD; prevalence.
DOI: 10.1007/s11606-007-0161-0

Portions of this work were presented at the annual meeting of the Society of General Internal Medicine, May 2005, New Orleans, LA, at the annual meeting of the College on Problems of Drug Dependence, June 2005, Orlando, FL, and at the annual meeting of the American Public Health Association, November 2004, Washington, DC.
Received November 21, 2006
Revised February 1, 2007
Accepted February 7, 2007
Published online March 10, 2007

INTRODUCTION

Media coverage of the 2001 World Trade Center attacks increased American awareness of posttraumatic stress disorder (PTSD) as an important psychiatric diagnosis.[1,2] Recent documentation of the burden of mental disorders, including PTSD, among US soldiers returning from Iraq[3,4] and people affected by Hurricane Katrina[5–7] further increased awareness of PTSD risk after exposure to trauma. Posttraumatic stress disorder is characterized by disabling symptoms of reexperiencing, avoidance, and numbing after a significant traumatic event.[8] The World Health Organization and others have substantiated the major impact of posttraumatic mental disturbance on social and health functioning.[9,10]

In primary care settings, the prevalence of PTSD depends upon the features of the population studied and the methods for participant recruitment and assessment. In a nationally representative US sample, Kessler et al.[11] found a lifetime PTSD prevalence of 6.8%; 61% of males and 51% of females reported trauma exposure.[12] Breslau et al.[13] reported 90% lifetime prevalence of trauma and 9% probability of developing PTSD after trauma in an urban sample in Detroit. Stein and colleagues[14] found a 2% prevalence of current PTSD in a middle-class, university-affiliated primary care practice in San Diego. In a representative sample of national health clinics in Israel, 9% of patients had current PTSD.[15] One-third of African-American primary care patients in an urban setting met PTSD diagnostic criteria.[16] Prevalence at veterans affairs (VA) health facilities ranges from 11.5% (current) to 27% (lifetime) PTSD.[17,18] Female gender, being divorced or widowed, other psychiatric or substance-use disorders, and exposure to certain traumas increase PTSD risk.[12,13,19]

However, these studies may not accurately estimate PTSD prevalence because of selective recruitment, nondiagnostic assessments, and low study participation.[14,16,19–21] Only one of these studies was conducted in a primary care practice in an urban, impoverished, and under represented minority population, a group at risk for PTSD because of high trauma exposure and low levels of social support.[22] PTSD is clinically under-recognized; of patients with PTSD, primary care providers recognized 2% in Israel and 47% in a VA setting.[15,17] These findings are consistent with extensive literature on clinician under-recognition of other mental disorders.[23]

Accurate determination of PTSD prevalence in primary care settings serving patients at high risk is important for several reasons. First, PTSD is associated with increased health-care utilization.[4,24–26] Second, PTSD and trauma exposure have been associated with conditions commonly treated by primary care clinicians: somatization functional impairment, depression, anxiety, chronic pain, irritable bowel syndrome (IBS) and substance-use disorders.[15,18,20,27–42] Apart from depression,[16] these associations have not been evaluated in primary care. Additionally, immigrant and refugee populations have high exposure to violence and merit further evaluation.[43] If PTSD is sufficiently prevalent in urban primary care patients or in those with certain common conditions, then it fulfills a key criterion for a clinical screening recommendation.[44] Determining if primary care patients with specific comorbid conditions have particularly high prevalence of current PTSD strengthens the case and may even suggest an approach for targeted screening in primary care.

We conducted a cross-sectional study to evaluate the prevalence of and risk factors for PTSD in primary care patients at an urban hospital-based outpatient department. We assessed the prevalence of PTSD among those with immigrant status, depression, generalized anxiety or panic disorder, chronic pain, IBS, heavy drinking, and substance dependence (SD), and we determined also whether PTSD was identified by physicians caring for these patients. We hypothesized that PTSD prevalence would be high, especially in patients with selected characteristics, and under-recognized by clinicians. We also hypothesized that clinicians might identify PTSD as depression, a more familiar mental illness.

METHODS

Study Design

We interviewed a consecutive sample of primary care patients at the outpatient department of an urban university-affiliated hospital to examine overall prevalence of PTSD and its prevalence among patients with seven selected conditions: being an immigrant to the United States, major or other depression, generalized anxiety or panic disorder, heavy drinking, SD, IBS, and chronic pain. We also reviewed medical records for documentation of PTSD and depression.

Participants

Patients visiting the Internal Medicine and Family Medicine clinics were eligible if they spoke English, were 18–65 years old, and had a scheduled appointment with a primary care clinician. Patients were excluded if they could not be interviewed alone or did not appear to understand study procedures. After a planned minimum study enrollment of 100 participants per selected condition was met for depression, anxiety, pain, and immigrant status with the first 509 enrolled participants ("consecutive sample"), enrollment was limited to eligible patients with heavy drinking, drug use or IBS, or none of the selected conditions (control group) to permit planned subgroup analyses ("oversampled group", n=98). Enrollment occurred from February 2003 to September 2004.

Recruitment and Enrollment

Interviewers approached consecutive patients upon arrival for appointments and screened them for eligibility. Eligible patients were asked to participate in an interview about stress and health. Each participant gave written informed consent. Study participants were compensated with $10, and they received safety referrals upon interview completion. Interviews not completed before clinician visit were completed immediately afterward. Boston University Medical Center's Institutional Review and HIPAA Privacy Review Boards approved the study. A Certificate of Confidentiality was obtained from the National Institutes of Health.

Data Collection

Screening. The self-administered screening tool contained 19 questions about the following: demographics, symptoms of depression and anxiety (2 questions from the PRIME-MD screener),[45] IBS (4 definition questions from the ROME II Integrative questionnaire),[46] quantity and frequency of alcohol use in the past 30 days,[47] use of heroin or cocaine in the past month, and chronic pain in the past 3 months.[48]

Interview. Interview included demographic questions, the Composite International Diagnostic Interview (CIDI) version 2.1 PTSD module,[8,49] the Chronic Pain Definitional Questionnaire,[48] the Patient Health Questionnaire (PHQ)[45] modules measuring depression diagnoses (major and other depression in the past 2 weeks) and anxiety disorder (generalized anxiety and/or panic disorder in the past 4 weeks), and the CIDI-Short Form (CIDI-SF) substance-use disorder modules.[49] The interview lasted for approximately 30 min.

Medical Record Review. Using standardized data-collection forms, researchers trained in chart abstraction reviewed patient electronic medical records (EMRs), including all outpatient and emergency department records and inpatient discharge summaries for the 12 months before and including the date of interview, to identify documentation of ICD-9 coded PTSD or depression in the problem list or in the visit assessment. All clinical encounters used EMR documentation.

Dependent Variables: Trauma Exposure, PTSD, Medical Record Documentation. Current (past 12 months) and lifetime PTSD were diagnosed using CIDI version 2.1.[49] The CIDI includes 9 questions about specific traumatic events and 2 open-ended questions. The research team classified open-ended responses as traumatic or nontraumatic according to DSM-IV guidelines for qualifying trauma.[8] We defined trauma exposure as self-reported experience of any category of qualifying trauma. Multiple experiences of a single category of trauma (e.g., rape

or physical assault) were counted as one trauma exposure. Clinician recognition of PTSD and depression was defined as documentation of this diagnosis (PTSD or depression, respectively) in the active problem list or in the free text of a note in the medical record.

Independent Variables and Covariates. Independent variables included immigrant status, IBS, chronic pain, anxiety, depression, heavy drinking, and SD. Immigrants were participants born outside the US or Puerto Rico. Irritable bowel syndrome was present if participants endorsed diagnostic criteria from the Rome II Integrative Questionnaire.[46] Chronic pain was present if participants endorsed pain for 3 months or more on the Chronic Pain Definitional Questionnaire.[48] Current anxiety disorder was present if participants met diagnostic criteria for generalized anxiety and/or panic disorder in the past month according to the PHQ.[45] Current depressive disorder was classified into major or other depression in the past 2 weeks according to the PHQ.[45] Heavy drinking in the past month was defined using the National Institute of Alcohol Abuse and Alcoholism guidelines.[47] Substance dependence included alcohol dependence and/or drug dependence in the past 6 months as determined by the CIDI-SF.[50] Covariates were age, gender, income, employment, marital status, and race.

Analyses

We used descriptive analyses to determine the prevalence of current and lifetime PTSD and trauma exposure. Bivariate analyses compared the prevalence of PTSD and trauma exposure in people with and without 7 selected conditions. We employed multivariable logistic regression to adjust for covariates. Adjusted proportions were calculated using the mean values of background covariates; *P* values comparing adjusted proportions are based on the regression coefficients for the group variable. Bivariate analyses compared PTSD or depression EMR documentation with research interview diagnoses. We used the consecutive sample (N=509) for most analyses. Analyses examining heavy drinking, SD, and IBS utilized the enhanced sample (the consecutive sample plus the oversampled group, n=98) of 607 participants.

RESULTS

Participant Demographic and Clinical Characteristics

Of 627 eligible patients, 509 (81%) were enrolled in the consecutive sample (Fig. 1). See Table 1 for the demographic and clinical characteristics of the sample. Gender differences were notable for race, immigrant status, income, other depression, anxiety, and SD. Compared to US-born people, immigrants were more likely to be employed (63 vs 47%) and to be married or living with a partner (37 vs 23%).

PTSD and Trauma Prevalence

Thirty-four percent (95% CI, 30–38%) of the consecutive sample met the diagnostic criteria for lifetime PTSD and 23% (CI 19–26%) for current PTSD. Median duration of PTSD symptoms was 6 years (range 1 month–44 years). Twenty-six percent (CI 21–32%) of women and 19% (CI 14–24%) of men met the criteria for current PTSD (p=.04). Trauma exposure was very common (79%), with 65% reporting exposure to more than one trauma category. See Appendix for details of trauma exposure.

Current PTSD was more common in participants with incomes <$20,000 (30 vs 14%, p<.0001), unemployed or disabled (30 vs 16%, p=.0001), or separated/divorced (31 vs 15% for married, p=.009; or 31 vs 22% for never married, p=.009).

Comorbid Conditions

PTSD was significantly more common among participants with at least one comorbid condition (depression, pain, IBS, anxiety, and/or SD) than among those with none of these conditions (27 vs 8%, p<.0001). Ninety-one percent of participants with current PTSD had at least one comorbid condition compared to 71% of those without PTSD. Of note, 75% of the entire consecutive sample (including those with PTSD) met the diagnostic criteria for at least one comorbid condition. In those with a comorbid condition, 28% had PTSD, whereas in those without a comorbid condition, 91% did not have PTSD.

Relationship of Comorbid Conditions with PTSD Prevalence: Adjusted Analyses

In the consecutive sample, current PTSD was less common among immigrants compared to nonimmigrants. Current PTSD was more common among participants with chronic pain, anxiety disorders, and both major and other depression. Adjusted analyses of lifetime PTSD prevalence in people with comorbid conditions echoed the current PTSD findings (Table 2).

Adjusted analyses in the enhanced sample revealed that current, but not lifetime, PTSD was significantly more common in participants with IBS. Current PTSD prevalence was not

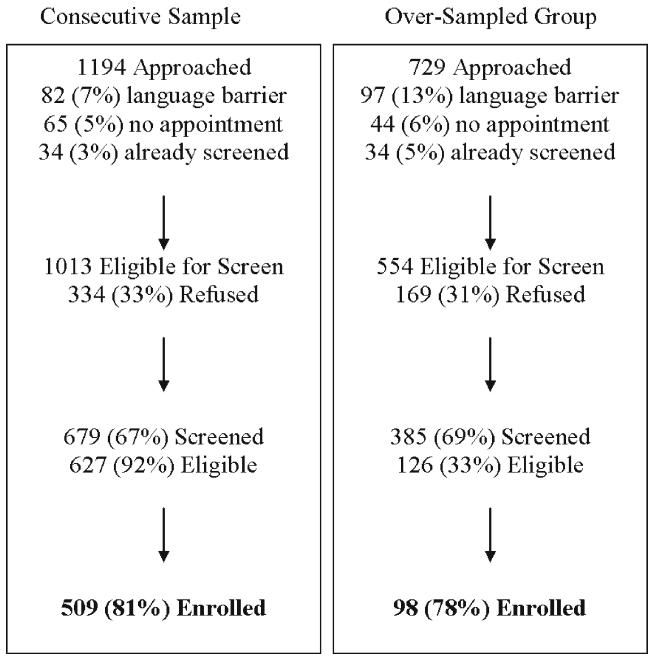


Figure 1. Study enrollment of patients in primary care. Consecutive sample: patients approached consecutively after registering for appointment. Oversampled group: consecutive sample criteria plus positive screen for irritable bowel symptoms, heavy drinking in past month, use of drugs in past month, or none of 6 select conditions (see text for details).

Table 1. Demographic and Clinical Characteristics of Participants Recruited in Primary Care (Consecutive Sample)

Variable	Total (*N*=509)	Male (*n*=251)	Female (*n*=258)
		N (%)	
Mean age in years (SD)	42 (11.3)	42 (11.1)	41 (11.5)
Race/Ethnicity			
Black/African American	300 (59)	137 (55)	163 (63)
White	98 (19)	60 (24)	38 (15)
Hispanic/Latino	40 (8)	25 (10)	15 (6)
Other	71 (14)	29 (12)	42 (16)
Immigrant status			
USA- or Puerto Rico-born	365 (72)	194 (77)	171 (66)
Marital Status			
Married/Living with partner	137 (27)	67 (27)	70 (27)
Separated/Divorced	110 (22)	55 (22)	55 (21)
Widowed	19 (4)	5 (2)	14 (5)
Never married/not living with a partner	243 (48)	124 (49)	119 (46)
Employment status			
Full-/Part-time	234 (46)	118 (47)	116 (45)
Student	28 (6)	11 (4)	17 (7)
Unemployed or disabled	247 (49)	122 (49)	125 (48)
Annual income			
<*$20,000	254 (52)	109 (44)	145 (59)
Irritable bowel syndrome	27 (5)	12 (5)	15 (6)
Chronic pain*	324 (64)	156 (62)	168 (65)
Major depression*	109 (21)	46 (18)	63 (24)
Other depressive disorder*	110 (22)	45 (18)	65 (25)
Anxiety disorder†	109 (22)	44 (18)	65 (25)
Heavy drinking	48 (9)	24 (10)	24 (9)
Substance dependence‡	73 (15)	52 (21)	21 (8)

*N=508
†Generalized anxiety and/or panic disorders (N=500)
‡N=502

significantly higher among heavy drinkers or those with SD, whereas lifetime PTSD was (Table 3).

Clinician Recognition: Medical Record Documentation

Based on EMRs of the participants in the consecutive sample, 11% of patients with CIDI-determined current PTSD had a PTSD diagnosis documented in the medical record. Of participants with current PTSD, 51% had depression noted in the medical record. Depression was noted in the EMR of 55% of those with both PTSD and depression (major and other), 43% of those with PTSD alone, 32% of those with depression but not PTSD, and 18% of those with neither PTSD nor depression ($p<.0001$ for differences between groups).

DISCUSSION

Among the patients attending an urban hospital-based primary care practice, PTSD is an exceptionally common diagnosis.[51] Almost one-quarter of patients met the criteria for current PTSD and one-third met the criteria for lifetime PTSD. Importantly, documentation of this diagnosis was exceedingly uncommon. Only 11% of those with current PTSD were correctly identified in the medical record. In patients with certain conditions (chronic pain, IBS, depression, anxiety disorder, and SD), PTSD prevalence was two to three times as high; patients with these conditions accounted for over 90% of all cases of PTSD. While over a quarter of those with a co-occurring condition had PTSD, not having a co-occurring condition made the diagnosis of PTSD very unlikely.

Our findings are consistent with the published studies of people with PTSD reporting more physical symptoms.[52–58] Multiple factors may contribute to the association between PTSD and physical symptoms including organic illnesses,[31,57,59–61] altered physiology,[62–64] and psychological mechanisms.[30]

In this study, lifetime PTSD was present in half of patients with current heavy drinking and SD. These findings are similar to those from substance-abuse treatment settings and primary care.[38,39,65] Surprisingly, current SD and heavy drinking were strongly associated with lifetime PTSD but only weakly with current PTSD. One explanation may be that substance use dampens PTSD symptoms, an assertion supported by reports of PTSD patient self-medication in other studies.[66–68] Another explanation may be that patterns of substance use are initiated in a period of current PTSD but persist beyond its resolution.

The higher PTSD prevalence in those with a diagnosis of depression and anxiety was not unexpected based on the literature detailing the psychiatric conditions comorbid with PTSD.[12,69] While primary care clinicians have become relatively more comfortable with identifying and managing depression in the past 10 years, more work remains to optimize screening and treatment for other psychiatric disorders.[70–73] More than half of our participants with EMR documentation of depression met the diagnostic criteria for PTSD, suggesting that clinicians may have labeled psychological distress as the more familiar diagnosis of depression.

Unlike depression, for which medication alone can be effective, evidence-based care for PTSD includes psychotherapy with or without medication.[74] Trauma-focused psychotherapies can result in significant improvement of symptoms in more than half of patients.[75] Thus, treatment by primary care clinicians may require new strategies beyond psychoactive medication prescription. In particular, the close relationship between physical problems and PTSD suggests an avenue to develop new interventions, better coordinating the care of general medicine and mental-health clinicians. Such coordination of care to address mental-health conditions in the primary care setting has been encouraged by a recent Institute of Medicine report.[76] Based on this study's findings, clinicians should consider evaluating patients who present with more than 3 months of physical pain or IBS for PTSD and referring those with PTSD for care. In addition, effective PTSD treatment may improve pain. Extensive literature suggests that treating depressive symptoms can improve pain and disability.[77–81]

An unexpected finding was the lower PTSD prevalence among immigrants. The study eligibility requirement to speak English likely selected acculturated immigrants, producing a "healthy immigrant effect".[82] Immigrants in this study had more social support as well. Other studies suggest lower psychiatric morbidity among some immigrant groups, depending on home country, socioeconomic status and subsequent adjustment to their adopted land.[83] Additionally, it is likely that immigrants with PTSD would have had substantial barriers to learning English

Table 2. PTSD Prevalence Among Participants with Selected Conditions (Consecutive Sample, *N*=509)

			Current PTSD				Lifetime PTSD			
Condition		*N*	Unadjusted prevalence (%)	*p* value	Adjusted prevalence‡ (95% CI)	*p* value§	Unadjusted prevalence (%)	*p* value	Adjusted prevalence‡ (95% CI)	*p* value§
Immigrant*	Yes	140	17		13% (8–20)	0.046	24		22% (15–30)	
	No	359	25	0.06	21% (17–26)		38	0.004	36% (31–42)	0.006
Chronic pain*	Yes	319	27		23% (18–29)	0.003	40		38% (33–44)	
	No	180	14	0.0009	12% (8–18)		23	<0.0001	21% (15–28)	0.0003
Anxiety†	Yes	104	50		42% (31–53)	<0.0001	68		64% (53–74)	
	No	389	16	<0.0001	14% (11–18)		25	<0.0001	25% (21–30)	<0.0001
Major depression*	Yes	104	43		35% (25–45)	<0.0001	60		54% (44–65)	
				<0.0001				<0.0001		<0.0001
Other depression	Yes	109	31		26% (18–36)		41		38% (29–48)	
No depression		286	12		11% (8–16)		22		22% (17–28)	

*N=499

†*Generalized anxiety disorder and/or panic disorder (N=493)*

‡*Models are adjusted for age, gender, race, income, employment, and marital status. Adjusted proportions were calculated using the mean values on background covariates.*

§*p values comparing adjusted proportions are based on the regression coefficients for the group variable.*

fluently enough to participate in this study.[84] Future studies should examine both English and non-English speaking immigrants for PTSD to better understand its prevalence.

The prevalence of PTSD in these urban primary care patients was markedly higher than that found in most other studies of primary care settings, including those involving veterans, where PTSD prevalence is expected to be high.[14,15,18,19,21,31,37,85] Possible explanations for this population's high PTSD prevalence are rooted in its socioeconomic profile with high exposure to trauma and low levels of social support. In a recent study of high school students in Boston, 71% witnessed violence and 44% were directly victimized in the prior year.[86] Social support is an important source of protection from the development of PTSD after trauma exposure,[87–90] and several demographic characteristics can serve as proxies for social support in this study. For example, half of our participants earned less than $20,000 per year. Likewise, half of the participants, all of whom were of working age, were unemployed or disabled. In addition, only 27% were married or living with a partner.[91,92]

There were limitations to this study. As a diagnostic instrument, the CIDI PTSD module only assessed for symptoms after one trauma selected as most stressful by the participant, whereas other diagnostic interviews assessed for symptoms after multiple traumas, offering more opportunities for PTSD diagnosis. However, the instrument used in this study is the same or very similar to those instruments used in national PTSD studies.[12,13] Our sample was drawn from a single hospital care system whose mission is to serve vulnerable populations and may not represent settings with a broader range of patients. However, our sample is similar to other urban medical settings serving low-income populations in demographic characteristics as well as prevalence of: depression, trauma exposure, substance-use disorders, and co-occurring chronic conditions.[16,93–97] Further research in primary care settings serving other populations is warranted. Lastly, the prevalence would only apply to patients aged 18–65 years old presenting for care, not to a population of all primary care patients.

Despite these limitations, the study had numerous strengths. We enrolled a large primary care sample of patients, oversampled for less prevalent comorbid conditions of interest, utilized a well-regarded structured diagnostic interview for PTSD, and analyzed data with multivariable approaches. These methodological features address some of the concerns with prior studies of PTSD prevalence in primary care.

Table 3. PTSD Prevalence Among Participants with Comorbid Conditions (Enhanced Sample, *N*=607)

			Current PTSD				Lifetime PTSD			
Condition		*N*	Unadjusted prevalence (%)	*p* value	Adjusted prevalence‡ (95% CI)	*p* value§	Unadjusted prevalence (%)	*p* value	Adjusted prevalence‡ (95% CI)	*p* value§
IBS*	Yes	47	45		34% (22–49)	0.012	53	0.007	45% (31–60)	
	No	550	21	0.0002	18% (15–22)		34		32% (28–37)	0.104
Heavy†	Yes	73	32		26% (16–38)	0.174	53	0.0005	49% (37–62)	
Drinking	No	524	21	0.05	19% (15–23)		33		31% (27–36)	0.007
Substance	Yes	105	30		26% (18–36)	0.069	52	<0.0001	52% (41–62)	
Dependence†	No	488	21	0.056	18% (14–22)		31		29% (25–34)	<0.0001

*N=597

†N=593

‡*Models are adjusted for age, gender, race, income, employment, and marital status. Adjusted proportions were calculated using the mean values on background covariates.*

§*p values comparing adjusted proportions are based on the regression coefficients for the group variable.*

Our results underscore the need to focus on the identification and treatment of PTSD in urban primary care settings as well as to explore the relationship between PTSD and certain conditions (chronic pain, IBS, SD, heavy drinking, depression, anxiety, and immigrant status) in varied settings with lower baseline PTSD prevalence. Before recommending broad screening for PTSD, screening tools and interventions in the primary care setting need to be developed and tested. In the meantime, it is appropriate for primary care clinicians treating patients in high-risk areas, particularly those with the conditions identified in this study, to assess for PTSD and refer those identified to effective care.

Acknowledgments: *This work was supported by a Generalist Physician Faculty Scholar Award from the Robert Wood Johnson Foundation, Princeton, New Jersey (RWJF #045452) and by a career development award from the National Institute on Drug Abuse, National Institutes of Health (K23 DA016665).*

We thank Jessica Geier, Minga Claggett-Borne, Lauren Kelly, Michael Rosas, Mary Reyes, Pavan Sekhar, Jen Tran, Eric Holder, and Mary Benitta Schickel for their aid in data collection, and Joann Elmore (MD MPH), Roger Weiss (MD), and Larry Culpepper (MD) for their comments on research design and analysis.

Conflict of Interest Summary: *None disclosed.*

Corresponding Author: *Jane Liebschutz, Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, 91 East Concord St., Suite 200, Boston, MA 02118, USA (e-mail: jliebs@bu.edu).*

APPENDIX

Table 4. Trauma exposure of participants recruited in primary care (consecutive sample)

Traumatic event	Total (*N*=509)	Male (*N*=251)	Female (*N*=258)
		N (%)	
Witness someone being badly injured or killed*	233 (46)	150 (60)	83 (32)
Seriously physically attacked or assaulted†	209 (41)	120 (48)	89 (35)
Threatened with a weapon, held captive, kidnapped*	191 (38)	127 (51)	64 (25)
Life-threatening accident*	163 (32)	98 (39)	65 (25)
Fire, flood, other natural disaster*	112 (22)	61 (24)	51 (20)
Sexually molested*	104 (20)	31 (12)	73 (28)
Raped†	80 (16)	11 (4)	69 (27)
Violent death of family/friend**	66 (13)	30 (12)	36 (14)
Violent acts to family/friend**	60 (12)	29 (12)	31 (12)
Non-violent death of family/friend**	57 (11)	31 (12)	26 (10)
Sexual violence to family/friend**	41 (8)	15 (6)	26 (10)
Tortured or victim of terrorism*	23 (5)	12 (5)	11 (4)
Direct combat experience in a war*	17 (3)	9 (4)	8 (3)
Health complications/illness	16 (3)	9 (4)	7 (3)
Accident (non-life-threatening)	17 (3)	11 (4)	6 (2)
Natural disaster to family/friend**	10 (2)	4 (2)	6 (2)

All traumas met the DSM-IV criterion A for PTSD: sudden, violent, or unexpected with corresponding horror, helplessness, or fear for one's life. Some included events occurring to a person close to the participant as per DSM-IV. Only trauma exposures reported by >2% participants are listed.
*N=508
†N=507
***Event occurred to person close to participant.*

REFERENCES

1. **Yehuda R.** Post-traumatic stress disorder. N Engl J Med. 2002;346 (2):108–14 (Jan 10).
2. **Goode E.** A nation challenged: psychological trauma. The New York Times. September 18, 2001;B:1.
3. **Mishra R.** Mental toll on troops detailed. The Boston Globe. July 1, 2004:A1.
4. **Hoge CW, Auchterlonie JL, Milliken CS.** Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. JAMA. 2006;295 (9):1023–32 (Mar 1).
5. **Weisler RH, Barbee JG, Townsend MH.** Mental health and recovery in the Gulf Coast after Hurricanes Katrina and Rita. JAMA. 2006;296 (5):585–8.
6. **Centers for Disease Control and Prevention.** Health hazard evaluation of police officers and firefighters after Hurricane Katrina, New Orleans, Louisiana, October 17–28 and November 30–December 5, 2005: Centers for Disease Control and Prevention; 2006.
7. **Lamberg L.** Katrina survivors strive to reclaim their lives. JAMA. 2006;296(5):499–502.
8. **American Psychiatric Association, ed.** Diagnostic and Statistical Manual of Mental Disorders Fourth Edition. Washington, DC: American Psychiatric Association; 1994.
9. **Mollica RF, Cardozo BL, Osofsky HJ, Raphael B, Ager A, Salama P.** Mental health in complex emergencies. Lancet. 2004;364(9450):2058–67 (Dec 4–10).
10. **Murray CJ, Lopez AD, eds.** The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Cambridge, MA: Harvard University Press; 1996; No. 1.
11. **Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE.** Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005;62(6):593–602.
12. **Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB.** Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry. 1995;52:1048–60.
13. **Breslau N, Kessler R, Chilcoat H, Schultz L, Davis G, Andreski P.** Trauma and posttraumatic stress disorder in the community: The 1996 Detroit Area Survey of Trauma. Arch Gen Psychiatry. 1998; 55:626–32.
14. **Stein M, McQuaid J, Pedrelli P, Lenox R, McCahill M.** Posttraumatic stress disorder in the primary care medical setting. Gen Hosp Psychiatry. 2000;22:261–9.
15. **Taubman-Ben-Ari O, Rabinowitz J, Feldman D, Vaturi R.** Posttraumatic stress disorder in primary care settings: prevalence and physicians' detection. Psychol Med. 2001;31:555–60.
16. **Alim TN, Graves E, Mellman TA, Aigbogun N.** Trauma exposure, posttraumatic stress disorder and depression in an African-American primary care population. J Natl Med Assoc. 2006;98(10):1630–6.

17. **Magruder KM, Frueh BC, Knapp RG, et al.** Prevalence of posttraumatic stress disorder in Veterans Affairs primary care clinics [see comment]. Gen Hosp Psychiatry. 2005;27(3):169–79.
18. **Escalona R, Achilles G, Waitzkin H, Yager J.** PTSD and somatization in women treated at a VA primary care clinic. Psychosomatics. 2004;45 (4):291–6 (Jul–Aug).
19. **Zlotnick C, Rodriguez BF, Weisberg RB, et al.** Chronicity in posttraumatic stress disorder and predictors of the course of posttraumatic stress disorder among primary care patients. J Nerv Ment Dis. 2004;192 (2):153–9 (Feb).
20. **Weisberg RB, Bruce SE, Machan JT, Kessler RC, Culpepper L, Keller MB.** Nonpsychiatric illness among primary care patients with trauma histories and posttraumatic stress disorder. Psychiatr Serv. 2002;53 (7):848–54 (Jul).
21. **Samson AY, Bensen S, Beck A, Price D, Nimmer C.** Posttraumatic stress disorder in primary care. J Fam Pract. 1999;48(3):222–7.
22. **Breslau N, Davis GC, Andreski P.** Risk factors for PTSD-related traumatic events: a prospective analysis. Am J Psychiatry. 1995;152(4):529–35.
23. **Edlund MJ, Unutzer J, Wells KB.** Clinician screening and treatment of alcohol, drug, and mental problems in primary care: results from healthcare for communities. Med Care. 2004;42(12):1158–66 (Dec).
24. **Walker EA, Katon W, Russo J, Ciechanowski P, Newman E, Wagner AW.** Health care costs associated with posttraumatic stress disorder symptoms in women. Arch Gen Psychiatry. 2003;60(4):369–74.
25. **Calhoun PS, Bosworth HB, Grambow SC, Dudley TK, Beckham JC.** Medical service utilization by veterans seeking help for posttraumatic stress disorder. Am J Psychiatry. 2002;159(12):2081–6 (Dec).
26. **Spertus IL, Yehuda R, Wong CM, Halligan S, Seremetis SV.** Childhood emotional abuse and neglect as predictors of psychological and physical symptoms in women presenting to a primary care practice. Child Abuse Negl. 2003;27(11):1247–58 (Nov).
27. **Andreski P, Chilcoat H, Breslau N.** Post-traumatic stress disorder and somatization symptoms: a prospective study. Psychiatry Res. 1998;79 (2):131–8 (Jun 15).
28. **Asmundson GJ, Coons MJ, Taylor S, Katz J.** PTSD and the experience of pain: research and clinical implications of shared vulnerability and mutual maintenance models. Can J Psychiatry. 2002;47(10):930–7.
29. **Bryant RA, Marosszeky JE, Crooks J, Baguley IJ, Gurka JA.** Interaction of posttraumatic stress disorder and chronic pain following traumatic brain injury. J Head Trauma Rehabil. 1999;14(6):588–94.
30. **Otis JD, Keane TM, Kerns RD.** An examination of the relationship between chronic pain and post-traumatic stress disorder. J Rehabil Res Dev. 2003;40(5):397–405.
31. **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(12):1306–12.
32. **Bonin MF, Norton GR, Asmundson GJ, Dicurzio S, Pidlubney S.** Drinking away the hurt: the nature and prevalence of PTSD in substance abuse patients attending a community-based treatment program. J Behav Ther Exp Psychiatry. 2000;31(1):55–66 (Mar).
33. **Rodriguez BF, Weisberg RB, Pagano ME, Machan JT, Culpepper L, Keller MB.** Mental health treatment received by primary care patients with posttraumatic stress disorder. J Clin Psychiatry. 2003;64(10):1230–6.
34. **Smith MV, Rosenheck RA, Cavaleri MA, Howell HB, Poschman K, Yonkers KA.** Screening for and detection of depression, panic disorder, and PTSD in public-sector obstetric clinics. Psychiatr Serv. 2004;55 (4):407–14 (Apr).
35. **Hankin CS, Spiro A, 3rd, Miller DR, Kazis L.** Mental disorders and mental health treatment among U.S. Department of Veterans Affairs outpatients: the Veterans Health Study. Am J Psychiatry. 1999;156 (12):1924–30 (Dec).
36. **Kubany ES, McKenzie WF, Owens JA, Leisen MB, Kaplan AS, Pavich E.** PTSD among women survivors of domestic violence in Hawaii. Hawaii Med J. 1996;55(9):164–5 (Sep).
37. **Davis TM, Wood PS.** Substance abuse and sexual trauma in a female veteran population. J Subst Abuse Treat. 1999;16(2):123–7.
38. **Najavits LM, Weiss RD, Shaw SR.** The link between substance abuse and posttraumatic stress disorder in women. A research review. Am J Addict. 1997;6(4):273–83.
39. **Reynolds M, Mezey G, Chapman M, Wheeler M, Drummond C, Baldacchino A.** Co-morbid post-traumatic stress disorder in a substance misusing clinical population. Drug Alcohol Depend. 2005;77 (3):251–8 (Mar 7).
40. **Brown PJ, Recupero PR, Stout R.** PTSD substance abuse comorbidity and treatment utilization. Addict Behav. 1995;20(2):251–4 (Mar–Apr).
41. **Drossman DA, Leserman J, Nachman G, et al.** Sexual and physical abuse in women with functional or organic gastrointestinal disorders. Ann Intern Med. 1990;113:828–33.
42. **Drossman DA, Talley NJ, Leserman J, Olden KW, Barreiro MA.** Sexual and physical abuse and gastrointestinal illness: review and recommendations. Ann Intern Med. 1995;123:782–94.
43. **Holman EA, Silver RC, Waitzkin H.** Traumatic life events in primary care patients: a study in an ethnically diverse sample. Arch Fam Med. 2000;9(9):802–10 (Sept/Oct).
44. **Force UPST.** Guide to Clinical Preventive Services, 2nd Edition. Bethesda, MD; 1996.
45. **Spitzer RL, Kroenke K, Williams JB.** Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA. 1999;282(18):1737–44.
46. **Drossman DA, Corazziari E, Talley NJ, Thompson WG, Whitehead WE.** Rome II. The Functional Gastrointestinal Disorders. Diagnosis, Pathophysiology and Treatment: A Multinational Consensus. McLean, VA: Degnon Associates; 2000.
47. **NIAAA.** Helping Patients Who Drink Too Much: A Clinician's Guide. http://pubs.niaaa.nih.gov/publications/Practitioner/CliniciansGuide2005/clinicians_guide.htm; 2006.
48. **Purves AM, Penny KI, Munro C, et al.** Defining chronic pain for epidemiological research assessing a subjective definition. Pain Clinic. 1998;10(3):139–47.
49. **World Health Organization.** Composite International Diagnostic Interview (CIDI): Version 2.1. Geneva: World Health Organization; 1997.
50. **Kessler R, Andrews G, Mroczek D, Ustun B, Wittchen HU.** The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF). Int J Methods Psychiatr Res. 1998;7(4):171–85.
51. Health, United States, 2004 With Chartbook on Trends in the Health of Americans. Hyattsville, MD: National Center for Health Statistics; 2004.
52. **Dobie DJ, Kivlahan DR, Maynard C, Bush KR, Davis TM, Bradley KA.** Posttraumatic stress disorder in female veterans: association with self-reported health problems and functional impairment. Arch Intern Med. 2004;164(4):394–400 (Feb 23).
53. **Sherman JJ, Turk DC, Okifuji A.** Prevalence and impact of posttraumatic stress disorder-like symptoms on patients with fibromyalgia syndrome. Clin J Pain. 2000;16(2):127–34 (Jun).
54. **Engel CC, Jr., Liu X, McCarthy BD, Miller RF, Ursano R.** Relationship of physical symptoms to posttraumatic stress disorder among veterans seeking care for gulf war-related health concerns. Psychosom Med. 2000;62(6):739–45 (Nov–Dec).
55. **Asmundson GJ, Wright KD, Stein MB.** Pain and PTSD symptoms in female veterans. Eur J Pain. 2004;8(4):345–50 (Aug).
56. **Amir M, Kaplan Z, Neumann L, Sharabani R, Shani N, Buskila D.** Posttraumatic stress disorder, tenderness and fibromyalgia. J Psychosom Res. 1997;42(6):607–13 (Jun).
57. **Lauterbach D, Vora R, Rakow M.** The relationship between posttraumatic stress disorder and self-reported health problems. Psychosom Med. 2005;67(6):939–47 (Nov–Dec).
58. **Beckham J, Crawford A, Feldman ME, et al.** Chronic posttraumatic stress disorder and chronic pain in Vietnam combat veterans. J Psychiatr Res. 1997;43(4):379–89.
59. **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(6):835–45 (Jul–Aug).
60. **Liebschutz JM, Feinman G, Sullivan L, Stein MD, Samet J.** Physical and sexual abuse in women infected with the human immunodeficiency virus: increased illness and health care utilization. Arch Intern Med. 2000;160:1659–64.
61. **Liebschutz JM, Mulvey KP, Samet JH.** Victimization among substance-abusing women: worse health outcomes. Arch Intern Med. 1997;157:1093–97.
62. **Sharp TJ, Harvey AG.** Chronic pain and posttraumatic stress disorder: mutual maintenance? Clin Psychol Rev. 2001;21(6):857–77 (Aug).
63. **Griffin MG, Resick PA, Yehuda R.** Enhanced cortisol suppression following dexamethasone administration in domestic violence survivors. Am J Psychiatry. 2005;162(6):1192–9 (Jun).
64. **Nixon RD, Bryant RA, Moulds ML, Felmingham KL, Mastrodomenico JA.** Physiological arousal and dissociation in acute trauma victims during trauma narratives. J Trauma Stress. 2005;18(2):107–13 (Apr).
65. **Zlotnick C, Bruce SE, Weisberg RB, Shea MT, Machan JT, Keller MB.** Social and health functioning in female primary care patients with post-

traumatic stress disorder with and without comorbid substance abuse. Compr Psychiatry. 2003;44(3):177–83.
66. **Jacobsen LK, Southwick SM, Kosten TR.** Substance use disorders in patients with posttraumatic stress disorder: a review of the literature. Am J Psychiatry. 2001;158(8):1184–90.
67. **Brady KT, Sinha R.** Co-occurring mental and substance use disorders: the neurobiological effects of chronic stress. Am J Psychiatry. 2005;162 (8):1483–93.
68. **Bremner JD, Southwick SM, Darnell A, Charney DS.** Chronic PTSD in Vietnam combat veterans: course of illness and substance abuse. Am J Psychiatry. 1996;153(3):369–75.
69. **Bruce SE, Weisberg RB, Dolan RT, et al.** Trauma and posttraumatic stress disorder in primary care patients. Prim Care Companion J Clin Psychiat. 2001;3(5):211–7 (Oct).
70. **Kessler RC, Berglund P, Demler O, et al.** The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA. 2003;289(23):3095–105 (Jun 18).
71. **Wang PS, Berglund P, Kessler RC.** Recent care of common mental disorders in the United States: prevalence and conformance with evidence-based recommendations. J Gen Intern Med. 2000;15(5): 284–92.
72. **Oquendo MA, Malone KM, Ellis SP, Sackeim HA, Mann JJ.** Inadequacy of antidepressant treatment for patients with major depression who are at risk for suicidal behavior. Am J Psychiatry. 1999;156(2):190–4.
73. **Seong-Yi B, Bowers B, Oakley L, Susman J.** The recognition of depression: the primary care clinician's perspective. Ann Fam Med. 2005;3:31–7.
74. **Foa EB, Davidson JRT, Frances A.** The expert consensus guideline series: treatment of post-traumatic stress disorder. J Clin Psychiatry. 1999;60(suppl 16):1–76.
75. **Bradley R, Greene J, Russ E, Dutra L, Westen D.** A multidimensional meta-analysis of psychotherapy for PTSD. Am J Psychiatry. 2005;162 (2):214–27 (Feb).
76. **Institute of Medicine, ed.** Improving the Quality of Health Care for Mental and Substance Use Conditions: Quality Chasm Series. Washington, DC: The National Academies Press; 2006.
77. **Katzelnick DJ, Simon GE, Pearson SD, et al.** Randomized trial of a depression management program in high utilizers of medical care. Arch Fam Med. 2000;9(4):345–51 (Apr).
78. **Lin EH, Katon W, Von Korff M, et al.** Effect of improving depression care on pain and functional outcomes among older adults with arthritis: a randomized controlled trial. JAMA. 2003;290(18):2428–29 (Nov 12).
79. **Simon GE, Katon W, Rutter C, et al.** Impact of improved depression treatment in primary care on daily functioning and disability. Psychol Med. 1998;28(3):693–701 (May).
80. **Von Korff M, Balderson BH, Saunders K, et al.** A trial of an activating intervention for chronic back pain in primary care and physical therapy settings. Pain. 2005;113(3):323–30 (Feb).
81. **Lin EH, VonKorff M, Russo J, et al.** Can depression treatment in primary care reduce disability? A stepped care approach. Arch Fam Med. 2000;9(10):1052–8 (Nov–Dec).
82. **Newbold KB.** Self-rated health within the Canadian immigrant population: risk and the healthy immigrant effect. Soc Sci Med. 2005;60 (6):1359–70.
83. **Chen AW, Kazanjian A.** Rate of mental health service utilization by Chinese immigrants in British Columbia. Can J Public Health. 2005;96 (1):49–51.
84. **Marshall GN, Schell TL, Elliott MN, Berthold SM, Chun CA.** Mental health of Cambodian refugees 2 decades after resettlement in the United States. JAMA. 2005;294(5):571–9.
85. **Dobie DJ, Maynard C, Kivlahan DR, et al.** Posttraumatic stress disorder screening status is associated with increased VA medical and surgical utilization in women. J Gen Intern Med. 2006;21(suppl 3):S58–S64 (Mar).
86. **Hemenway D, Prothrow-Stith D, Browne A.** Report of the 2004 Boston Youth Survey. Boston: City of Boston; 2005 (Aug).
87. **Ahern J, Galea S, Fernandez WG, Koci B, Waldman R, Vlahov D.** Gender, social support, and posttraumatic stress in postwar Kosovo. J Nerv Ment Dis. 2004;192(11):762–70.
88. **Glynn SM, Asarnow JR, Asarnow R, et al.** The development of acute post-traumatic stress disorder after orofacial injury: a prospective study in a large urban hospital. J Oral Maxillofac Surg. 2003;61(7):785–92.
89. **Muller RT, Goebel-Fabbri AE, Diamond T, Dinklage D.** Social support and the relationship between family and community violence exposure and psychopathology among high risk adolescents. Child Abuse Negl. 2000;24(4):449–64.
90. **Viinamaki H, Kontula O, Niskanen L, Koskela K.** The association between economic and social factors and mental health in Finland. Acta Psychiatr Scand. 1995;92(3):208–13.
91. **Steptoe A, Marmot M.** Burden of psychosocial adversity and vulnerability in middle age: associations with biobehavioral risk factors and quality of life. Psychosom Med. 2003;65(6):1029–37.
92. **Sherbourne CD, Stewart AL.** The MOS social support survey. Soc Sci Med. 1991;32(6):705–14.
93. **Miller DK, Malmstrom TK, Joshi S, Andresen EM, Morley JE, Wolinsky FD.** Clinically relevant levels of depressive symptoms in community-dwelling middle-aged African Americans. J Am Geriatr Soc. 2004;52(5):741–8.
94. **Matheson FI, Moineddin R, Dunn JR, Creatore MI, Gozdyra P, Glazier RH.** Urban neighborhoods, chronic stress, gender and depression. Soc Sci Med. 2006;63(10):2604–16.
95. **O'Toole TP, Arbelaez JJ, Lawrence RS; Baltimore Community Health Consortium.** Medical debt and aggressive debt restitution practices: predatory billing among the urban poor. J Gen Intern Med. 2004;19 (7):772–8.
96. **Gaskin DJ, Hadley J.** Population characteristics of markets of safety-net and non-safety-net hospitals. J Urban Health. 1999;76(3):351–70.
97. **Bassuk EL, Buckner JC, Perloff JN, Bassuk SS.** Prevalence of mental health and substance use disorders among homeless and low-income housed mothers. Am J Psychiatry. 1998;155(11):1561–4.

BMC Medical Genetics

BioMed Central

Research

Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study

Kathryn L Lunetta[1,3], Ralph B D'Agostino Sr[1,5], David Karasik[4], Emelia J Benjamin[1,2], Chao-Yu Guo[1,2], Raju Govindaraju[1,2], Douglas P Kiel[4], Margaret Kelly-Hayes[1,2], Joseph M Massaro[1,5], Michael J Pencina[1,5], Sudha Seshadri[1,2] and Joanne M Murabito*[1,2]

Address: [1]The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA, [2]Section of General Internal Medicine and the Departments of Neurology, Cardiology, and Preventive Medicine, Boston University School of Medicine, Boston, MA, USA, [3]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA, [4]Hebrew Senior Life Institute for Aging Research and Harvard Medical School, Boston, MA, USA and [5]Statistics and Consulting Unit, Department of Mathematics, Boston University, Boston, MA, USA

Email: Kathryn L Lunetta - klunetta@bu.edu; Ralph B D'Agostino - ralph@bu.edu; David Karasik - karasik@hrca.harvard.edu; Emelia J Benjamin - emelia@bu.edu; Chao-Yu Guo - chaoyu@bu.edu; Raju Govindaraju - drgraju@bu.edu; Douglas P Kiel - kiel@hrca.harvard.edu; Margaret Kelly-Hayes - mkhayes@bu.edu; Joseph M Massaro - jmm@bu.edu; Michael J Pencina - mpencina@bu.edu; Sudha Seshadri - suseshad@bu.edu; Joanne M Murabito* - murabito@bu.edu

* Corresponding author

Published: 19 September 2007

BMC Medical Genetics 2007, **8**(Suppl 1):S13 doi:10.1186/1471-2350-8-S1-S13

This article is available from: http://www.biomedcentral.com/1471-2350/8/S1/S13

Abstract

Background: Family studies and heritability estimates provide evidence for a genetic contribution to variation in the human life span.

Methods: We conducted a genome wide association study (Affymetrix 100K SNP GeneChip) for longevity-related traits in a community-based sample. We report on 5 longevity and aging traits in up to 1345 Framingham Study participants from 330 families. Multivariable-adjusted residuals were computed using appropriate models (Cox proportional hazards, logistic, or linear regression) and the residuals from these models were used to test for association with qualifying SNPs (70, 987 autosomal SNPs with genotypic call rate ≥80%, minor allele frequency ≥10%, Hardy-Weinberg test p ≥ 0.001).

Results: In family-based association test (FBAT) models, 8 SNPs in two regions approximately 500 kb apart on chromosome 1 (physical positions 73,091,610 and 73, 527,652) were associated with age at death (p-value < 10^{-5}). The two sets of SNPs were in high linkage disequilibrium (minimum $r^2 = 0.58$). The top 30 SNPs for generalized estimating equation (GEE) tests of association with age at death included rs10507486 (p = 0.0001) and rs4943794 (p = 0.0002), SNPs intronic to *FOXO1A*, a gene implicated in lifespan extension in animal models. FBAT models identified 7 SNPs and GEE models identified 9 SNPs associated with both age at death and morbidity-free survival at age 65 including rs2374983 near *PON1*. In the analysis of selected candidate genes, SNP associations (FBAT or GEE p-value < 0.01) were identified for age at death in or near the following genes: *FOXO1A, GAPDH, KL, LEPR, PON1, PSEN1, SOD2, and WRN*. Top ranked SNP associations in the GEE model for age at natural menopause included rs6910534 (p = 0.00003) near *FOXO3a* and rs3751591 (p = 0.00006) in *CYP19A1*. Results of all longevity phenotype-genotype associations for all autosomal SNPs are web posted at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007.

Conclusion: Longevity and aging traits are associated with SNPs on the Affymetrix 100K GeneChip. None of the associations achieved genome-wide significance. These data generate hypotheses and serve as a resource for replication as more genes and biologic pathways are proposed as contributing to longevity and healthy aging.

Background

Genetic factors associated with human longevity and healthy aging remain largely unknown. Heritability estimates of longevity derived from twin registries and large population-based samples suggest a significant but modest genetic contribution to the human lifespan (heritability ~15 to 30%) [1-4]. However, genetic influences on lifespan may be greater once an individual achieves age 60 years [5]. Moreover, the reported magnitude of the genetic contribution to other important aspects of aging such as healthy physical aging (wellness)[6], physical performance [7,8], cognitive function [9], and bone aging [10] are much larger. Both exceptional longevity and a healthy aging phenotype have been linked to the same region on chromosome 4 [11,12], suggesting that although longevity per se and healthy aging are different phenotypes, they may share some common genetic pathways.

A number of potential candidate genes in a variety of biological pathways have been associated with longevity in model organisms. Genes involved in the regulation of DNA repair and genes in the evolutionarily conserved insulin/insulin-like growth factor signaling pathway [13,14] are emerging as holding great promise in the future elucidation of the underlying physiology controlling lifespan. Many of these genes have human homologs and thus have potential to provide insights into human longevity [15-20]. Although numerous candidate genes have been proposed, studies in humans are limited and initial findings often fail replication [21,22]. More recently genome-wide association studies (GWAS) have become feasible and offer a more comprehensive and untargeted approach to detect genes with modest phenotypic effects that underlie common complex conditions [23].

We had the opportunity to use the Framingham Heart Study (FHS) Affymetrix 100K SNP genotyping resource for a GWAS of longevity and aging-related phenotypes. The FHS offers the unique advantage of a longitudinal family-based community sample with participants who have been well-characterized throughout adulthood with respect to prospectively ascertained risk factors and diseases and continuously followed until death. We report several strategies for 100K SNP associations: 1) a simple low p-value SNP ranking strategy; 2) SNP selection due to associations with more than one related phenotype; and 3) SNP associations within candidate genes and regions previously reported to be associated with longevity in model organisms or humans.

Methods

Study sample

The genotyped study sample is comprised of 1345 Original cohort (n = 258) and Offspring (n = 1087) participants who are members of the 330 largest FHS families. The Overview [24] provides further details of this sample. With respect to aging and longevity traits, 149 deaths occurred at a mean age at death of 83 years (range 46 to 99 years) and 713 participants achieved age 65 years or greater. The Boston University Medical Center Institutional Review Board approved the examination content of Original Cohort and Offspring examinations. All participants provided written informed consent at every examination including consent for genetic studies.

Longevity and aging phenotype definitions and residual creation

Age at death

Both the Original Cohort and the Offspring Cohort remain under continuous surveillance and all deaths that occurred prior to January 1, 2005 were included in this study. Deaths were identified using multiple strategies including routine participant contact for research examinations or health history updates, surveillance at the local hospital, search of obituaries in the local newspaper, and if needed through use of the National Death Index. Death certificates were routinely obtained and all hospital and nursing home records prior to death and autopsy reports (if performed) were requested. In addition, if there was insufficient information to determine a cause of death, the next of kin were interviewed by a senior investigator. All records pertinent to the death were reviewed by an endpoint panel comprised of three senior investigators. The date and cause of death (classified as due to coronary heart disease, stroke, other cardiovascular disease [CVD], cancer, other causes, or unknown cause) was recorded.

Cox proportional hazards models were used to generate martingale residuals using the PHREG procedure in SAS to perform the regression analysis of survival time from age at study entry to age at death. Models were sex-specific and adjusted for 1) birth cohort and 2) birth cohort, education, current smoking status (yes/no), obesity (body mass index ≥30 kg/m^2), hypertension (blood pressure ≥140/90 mmHg or on antihypertensive treatment), elevated cholesterol (cholesterol > 239 mg/dL), diabetes

(fasting blood sugar ≥126 mg/dL, random blood sugar of ≥200 mg/dL, or use of insulin or oral hypoglycemic agents) and comorbidity defined as CVD and cancer. Birth cohort was defined as a categorical variable for all regression models with the following categories based on year of birth: birth year prior to 1900, 1900 to 1909, 1910 to 1919, 1920 to 1929, 1930 to 1939, 1940 to 1949, and 1950 and later. All covariates were measured at study entry. Residuals from Original Cohort and Offspring participants were pooled.

Morbidity-free survival at age 65 years

Morbidity-free survival was defined as achieving age 65 years free of CVD, dementia, and cancer. CVD events included angina pectoris, coronary insufficiency, myocardial infarction, heart failure, stroke, transient ischemic attack (TIA), intermittent claudication and coronary or CVD death. Suspected CVD events were reviewed by a panel of three investigators who adjudicated events using previously established criteria in place since study inception [25]. A separate panel of study neurologists determined the presence of stroke or TIA and a team of at least one neurologist and one neuropsychologist determined the presence of dementia. Two independent reviewers examined records for all cancers, and the vast majority of cancer cases were microscopically confirmed with pathology reports.

Logistic regression models were used to generate deviance residuals. Models were sex-specific and adjusted for 1) birth cohort and 2) birth cohort, education, current smoking status, obesity, hypertension, elevated cholesterol, and diabetes. Covariates were defined as above for age at death. All covariates were measured at the examination closest to the participant attaining age 65 years using a 5 year window around age 65 years. Residuals from Original Cohort and Offspring participants were pooled.

Age at natural menopause

Natural menopause occurred after a woman had ceased menstruating naturally for one year and the age at natural menopause was the self-reported age at last menstruation. Mean age at natural menopause was similar in Original Cohort and Offspring women and the distribution of naturally menopausal ages in women in the 330 FHS families was similar to that of women in all 1643 FHS families [26,27]. The mean age at natural menopause in women in the 100K sample was 50.2 years (range 38 to 57 years) in Original Cohort women and 49.1 years (range 29 to 60 years) in Offspring women.

Crude age at natural menopause and standardized residuals from multiple linear regressions in SAS [28] that adjusted age at natural menopause for covariates of interest were used as traits for analysis. Covariates were obtained at all attended examinations prior to the onset of menopause and included mean number of cigarettes smoked per day, mean body mass index, parity (0 versus 1 or more live births), and generation (Original Cohort vs. Offspring).

Walking speed

Walking speed was measured on Original Cohort participants at examination 27 (January 2002 through December 2003, mean age of Original Cohort at exam 27: 86.7 years) and Offspring participants attending an ancillary study to examination 7 (1999 to 2004, mean age at exam: 62.0 years). Trained technicians timed participants walking at their normal pace on a four meter course twice and subsequently asked participants to repeat the course walking at a rapid pace. The mean timed fast walk among Offspring participants in the 100K genotyping sample was 2.44 seconds (standard deviation 0.89). The timed fast walk was used for analysis. Sex-specific linear regression was used to generate residuals adjusted for age and height measured at the time of the walk.

Biologic age by osseographic scoring system

An osseographic scoring system (OSS) was applied to hand radiographs obtained on original cohort (1967 to 1969, mean age 58.7 years) and offspring participants (1992 to 1993, mean age 51.6 years) [10]. Biologic age was then defined as the standardized residual between the OSS predicted age and the actual age. Biologic age defined by this system predicted mortality [10,29], was very heritable ($h^2 = 0.57 \pm 0.06$), and a genome-wide linkage analysis was performed with LOD scores >1.8 present on chromosomes 3q, 11p, 16q, and 21q [10]. Sex- and cohort-specific ranked residuals generated from linear regression of age on log-OSS adjusted for height, body mass index, menopause, and estrogen therapy, were used for analysis.

Genotyping

Affymetrix 100K SNP GeneChip genotyping and the Marshfield STR genotyping performed by the Mammalian Genotyping Service http://research.marshfieldclinic.org/genetics are described in the Overview paper [24].

Statistical analysis

The statistical methods for genome-wide linkage and association analyses are described in the Overview [24].

Association

All residual traits described above as well as the additional traits listed in Table 1 were computed using Cox proportional hazards with martingale residuals for survival traits, logistic regression with deviance residuals for dichotomous traits, and linear regression with standard residuals for quantitative traits. The full set of FHS participants with

the phenotype were used to create the residuals. The residuals were used to test for association between the genotyped subset of individuals and the SNPs using additive family-based association test (FBAT) and generalized estimating equations (GEE) models as described in the Overview [24]. A total of 70,987 autosomal SNPs met the criteria of genotypic call rate ≥80%, minor allele frequency ≥10%, Hardy-Weinberg test $p \geq 0.001$, and ≥10 informative families for FBAT. The number of tests with an FBAT $p < 0.001$, $p < 0.0001$, and $p < 0.00001$ for all phenotypes was similar to what would be expected under the assumptions that the 70,987 tested SNPs were independent and there were no true associations. The GEE tests tended to give an excess of very small p-values over what would be expected under these assumptions.

SNP prioritization

We used several strategies to prioritize SNPs associated with longevity and aging traits. First, we used an untargeted approach whereby the top 50 SNP associations ranked according to the strength of the p-value for each trait were examined. Next, we explored the consistency of SNP associations across related sets of traits chosen a priori (trait set one: age at death and morbidity-free survival at age 65 years; trait set two: biologic age and walking speed). Trait set one was chosen based upon linkage data in humans demonstrating that both longevity and a healthy aging trait were linked to the same region on chromosome 4 raising the hypothesis that the two phenotypes may share common genetic pathways [11,12]. The traits in set two reflect aging with good physical functioning and thus we postulated that biologic age and walking speed may have genetic variants in common. We also investigated SNP associations in candidate genes and regions reported to be associated with longevity identified from established databases including NCBI [14] using the search term "longevity" and the Science of Aging Knowledge Environment genes/intervention database http://sageke.sciencemag.org/cgi/genesdb[30] choosing genes potentially related to lifespan in humans.

The SNPs were annotated using the UCSC genome browser tables using the May 2004 assembly http://genome.ucsc.edu/[31,32]. All genes within 60 kb of the top ranked SNPs were identified.

Results

The longevity and aging traits available in the FHS 100K SNP resource are listed in Table 1. In this report, we consider only five of the traits listed in Table 1: multivariable-adjusted age at death, morbidity-free survival at age 65 years, age at natural menopause, walking speed, and biologic age by OSS. These traits include a pooled sample of Original Cohort and Offspring participants, with the exception of walking speed, which is reported in Offspring participants only. Details of the sample size and covariate adjustment for each trait are provided in Table 1.

For each of the five phenotypes, Table 2a and 2b provides the top five SNPs ranked in order by lowest p-value for the GEE and FBAT models (all associations can be viewed on the web http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007). If multiple SNPs in linkage disequilibrium (LD $r^2 > 0.80$) were included in the top 5, additional SNPs were included until a set of 5 independent associations were listed. Eight SNPs on chromosome 1 were associated with age at death in the FBAT analysis; all with p-value < 10^{-4} and two with p-value < 10^{-5}. The 8 SNPs consisted of two sets of SNPs (rs10493513, rs10493514, rs6689491, rs6657082, rs1405051) and (rs10493515, rs10493518, rs10493517), clustered in two regions approximately 500 kb apart. There was exceptionally high LD across this 500 kb region: the minimum r^2 between pairs of the eight SNPs was 0.58. The nearest genes in this region existing in public databases were >500 kb from any of these SNPs [31,32].

There were several additional associations not listed in Table 2a and 2b that were of interest. For age at death in the GEE analysis, SNP associations ranked numbers 9 and 13 were rs10507486 (p-value 0.000128) and rs4943794 (p-value 0.000277), both are intronic *FOXO1A* SNPs. For age at natural menopause, top ranked SNP associations in the GEE model included number 11, rs6910534 ($p = 0.00003$) near *FOXO3a* and number 18, rs3751591 ($p = 0.00006$) in *CYP19A1*.

Table 2c presents the LOD scores ≥2.0 and the corresponding 1.5-LOD support interval from genome-wide linkage for the three quantitative aging traits. None of the regions overlapped with SNPs associated with these aging traits in the FBAT and GEE analyses. Of note for biologic age by OSS the linkage peak on chromosome 21 confirmed a prior Framingham Study report using a genome-wide scan with 401 microsatellite markers [10].

Table 3 provides all SNP associations with a GEE or FBAT $p < 0.01$ for both traits within the two pairs of related traits. For age at death and morbidity-free survival at age 65 years, FBAT models identified 7 SNPs and GEE models identified 9 SNPs associated with both traits including rs2374983 near *PON1* (Tables 3a and 3b). For biologic age by OSS and walking speed, 13 SNPs in FBAT models and 6 SNPs in GEE models were associated with both traits (Tables 3c and 3d).

We identified from the literature 79 potential candidate genes and regions associated with longevity (see Additional file 1 for listing). Of these, 12 genes had no SNPs and 67 genes had 1 to 45 SNPs within 60 kb of the gene

Table 1: Aging and Longevity Phenotypes for Framingham Heart Study 100K Project

		Exam cycle(s)		
Phenotype Subgroup **• Trait (variable name on the website*)**	**Number of Traits**	**N (MV**)**	**Offspring / Original Cohort**	**Adjustment**
Survival Traits: Cox regression				
Survival • Age at death (1. deathageX, 2. deathageMV)	2	1345 (1166)	Cohort & Offspring pooled	Cox regression Sex-specific 1. birth cohort 2. multivariable adjusted for birth cohort, education, smoking, obesity (BMI ≥ 30), CVD risk factors, co-morbidity measured at exam 1
Categorical traits: Logistic regression				
• Survival past the ALE (1. deathpastALEX, 2. deathpastALEMV)	2	1345 (1166)	Cohort & Offspring pooled	Logistic regression Sex-specific 1. birth cohort 2. multivariable adjusted for birth cohort, education, smoking, obesity (BMI ≥ 30), CVD risk factors, co-morbidity measured at exam 1
Morbidity-free survival (free of CVD, cancer and dementia) • At age 65 years (1. morbidityfree65X, 2. morbidityfree65MVX)	2	558 (558)	Cohort & Offspring pooled, exams closest to age 65 years	Logistic regression Sex-specific 1. birth cohort 2. multivariable adjusted for birth cohort, education, smoking, obesity, CVD risk factors measured at exam closest to age 65 years (within a 5 year horizon)
Quantitative Traits: Linear regression				
Reproductive Aging • Age at natural menopause (1. menoageX, 2. menoageMVX)	2	438 (378)	Cohort & Offspring pooled, women only	Linear regression 1. crude 2. multivariable adjusted for smoking, BMI, parity, generation (measured at exams prior to menopause)
Cognitive function • MMSE at age 65 years (1. MMSE65X, 2. MMSE65MVX) • MMSE at the specified Offspring exam (1. MMSE5X, 2. MMSE5MVX, 1. MMSE7X, 2. MMSE7MV, 1. MMSE5to7X, 2. MMSE5to7MVX)	2	593 (462)	Cohort & Offspring pooled, exams at age 65	Linear regression Sex-specific 1. birth cohort 2. multivariable adjusted for birth cohort, education, FSRP measured at exam closest to age 65 years (5 year horizon)
	6	1038 (913)	Exam 5 Exam 7 Exam 5 & 7 average score	Linear regression Sex-specific 1. birth cohort 2. multivariable adjusted for birth cohort, education, FSRP; covariates measured at the specified exam
Physical Performance • Hand grip (2. handgrip7x, 2. handgrips727x) • Walking speed (2. walkingspeed7x, 2. walkingspeed727x)	6	764	Exam 7 Exam 7 and Exam 27	Linear regression Sex-specific‡ 1. age 2. multivariable adjusted for age, height, weight at the specified exam
Biologic Age by Osseographic Scoring System (1. deltaOSSr, delta OSSrf, deltaOSSrm)	3	714	Offspring and Cohort pooled exam 6/7 and exam 22	Linear regression Sex- and cohort-specific ranked residuals§ 1. multivariable adjusted for age, height, BMI, menopause, estrogen use

Residuals from these models were used as traits to test for association with SNP genotypes.
* The number preceding the variable name refers to the covariate adjustment in the last column of the table. The website with all results is found at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007; ** MV = N for multivariable trait
‡ cohort- and sex-specific residuals for traits that included both cohort and offspring; §cohort-specific for traits limited to one sex
ALE = average life expectancy, BMI = body mass index, Co-morbidity = cardiovascular disease and cancer, CVD = cardiovascular disease, FSRP = Framingham stroke risk profile, MMSE = mini-mental state exam, Risk factors = hypertension, diabetes, elevated cholesterol

on the 100K Affymetrix GeneChip. There were 2036 SNPs in the LGV1 region on chromosome 4 previously linked to exceptional longevity [11]. Table 4 shows the candidate genes with SNPs associated with an FBAT or GEE p-value < 0.01 for age at death including: *FOXO1a, GAPDH, KL, LEPR, PON1, PSEN1, SOD2, and WRN* and for morbidity-free survival at 65 years including:*GHR, LEPR, MORF4L1, PON1, PTH, and WRN*. Biologic age by OSS shared 2 SNPs in common with age at death: rs4943794 intronic to *FOXO1a* and rs911847 near *SOD2*.

Discussion

To our knowledge, this is the first dense GWAS of longevity and aging traits in a community-based sample of adults from two generations of the same families. Over 1300 men and women have detailed longevity and aging-

Table 2: Aging and Longevity Phenotypes† for FHS 100K Project: Results of Association and Linkage Analyses

2a. GEE, Top 5 p-values by Phenotype*

Trait	SNP	Chromosome	Physical location	GEE p-value	FBAT p-value	Gene Region (within 60 kb)
Age at death						
	rs1528753	11	90,523,987	**8.1×10^{-8}**	0.024	
	rs2371208	7	81,982,510	**2.6×10^{-6}**	0.031	
	rs10496799	2	139,261,401	**1.4×10^{-5}**	0.735	NXPH2
	rs10489006	4	31,444,987	**3.6×10^{-5}**	0.078	
	rs3757354	6	16,235,386	**6.4×10^{-5}**	0.316	MYLIP
Morbidity-free survival at age 65						
	rs1412337	1	165,350,299	**1.8×10^{-9}**	0.505	DPT
	rs32566	5	5,845,507	**1.9×10^{-9}**	0.323	
	rs10484246	6	9,559,183	**8.4×10^{-8}**	0.928	
	rs4831837	8	12,756,234	**4.7×10^{-7}**	0.182	
	rs2639889	16	59,680,648	**9.4×10^{-7}**	0.903	
Age at natural menopause‡						
	rs10496265	2	81,580,466	**1.1×10^{-8}**	0.001	
	rs10496262*	2	81,662,782	**3.3×10^{-7}**	0.005	
	rs958672	2	154,896,075	**1.9×10^{-6}**	0.087	GALNT13
	rs291353	1	232,046,939	**5.5×10^{-6}**	0.035	GNG4
	rs726336	5	163,911,906	**1.1×10^{-5}**	0.125	
Walking speed exam 7						
	rs7137869	12	118,452,366	**6.3×10^{-7}**	0.009	CCDC60
	rs7662116	4	154,375,569	**1.9×10^{-5}**	0.016	
	rs7972859*	12	118,452,765	**2.5×10^{-5}**	0.005	CCDC60
	rs9318312	13	74,489,506	**5.8×10^{-5}**	0.266	
	rs1994854	4	78,124,824	**9.4×10^{-5}**	0.280	
	rs7718104	5	122,183,258	**1.2×10^{-4}**	0.011	SNX2
Biologic age by osseographic scoring system						
	rs1463605	12	30,005,150	**7.0×10^{-8}**	5.3×10^{-4}	
	rs7176093	15	84,170,434	**7.4×10^{-6}**	0.005	KLHL25
	rs3772255	3	157,585,436	**8.2×10^{-6}**	0.085	KCNAB1
	rs726846	5	136,099,953	**1.1×10^{-5}**	0.003	
	rs646983	13	29,413,553	**1.2×10^{-5}**	0.003	

2b. FBAT, Top 5 p-values by Phenotype*

Trait	SNP	Chromosome	Physical location	GEE p-value	FBAT p-value	Gene Region (within 60 kb)
Age at death						
	rs10493513	1	73,091,610	0.640	**1.5×10^{-6}**	
	rs10493514*	1	73,092,533	0.623	**2.8×10^{-6}**	
	rs6689491*	1	73,064,050	0.205	**2.0×10^{-5}**	
	rs10493515	1	73,527,652	0.225	**2.3×10^{-5}**	
	rs10493518*	1	73,572,652	0.191	**3.6×10^{-5}**	
	rs10493517*	1	73,570,372	0.215	**4.2×10^{-5}**	
	rs6657082*	1	73,065,349	0.224	**5.5×10^{-5}**	
	rs10498263	14	19,285,288	0.310	**8.3×10^{-5}**	OR4Q3 OR4M1
	rs1915501	4	28,612,632	0.383	**1.1×10^{-4}**	
	rs1405051*	1	73,060,505	0.176	**1.4×10^{-4}**	
	rs6459623	6	18,634,791	0.604	**1.5×10^{-4}**	IBRDC2

Table 2: Aging and Longevity Phenotypes† for FHS 100K Project: Results of Association and Linkage Analyses *(Continued)*

Morbidity-free survival at age 65					
rs10509200	10	65,296,567	0.613	**7.0×10^{-5}**	
rs965036	6	20,099,022	0.550	**8.6×10^{-5}**	
rs720565	6	136,834,657	0.014	**9.8×10^{-5}**	MAP7
rs1192372	2	84,923,204	0.094	**9.9×10^{-5}**	
rs10505239	8	115,976,403	0.141	**1.3×10^{-4}**	
Age at natural menopause†					
rs959702	10	2,139,260	0.003	**1.4×10^{-5}**	
rs7165378	15	69,478,558	0.006	**6.4×10^{-5}**	
rs997161	10	130,876,127	0.074	**8.6×10^{-5}**	
rs165284	1	91,235,574	0.006	**8.6×10^{-5}**	ZNF644
rs2280585	3	64,882,324	0.884	**1.1×10^{-4}**	
Walking speed exam 7					
rs4471448	11	86,760,972	0.175	**3.8×10^{-6}**	TMEM135
rs336963	5	83,005,460	0.570	**2.6×10^{-5}**	HAPLN1
rs7862683	9	18,257,947	0.001	**4.9×10^{-5}**	
rs9317757	13	68,458,430	0.252	**7.8×10^{-5}**	
rs10501636*	11	86,789,967	0.355	**1.1×10^{-4}**	
rs2340392	3	80,440,990	0.002	**1.7×10^{-4}**	
Biologic age by osseographic scoring system					
rs1380703	2	57,852,938	0.008	**1.1×10^{-5}**	
rs324702	4	77,094,969	0.390	**3.3×10^{-5}**	PPEF2
rs324735	4	77,062,193	0.096	**7.6×10^{-5}**	
rs1106184	2	10,914,125	0.006	**8.8×10^{-5}**	PDIA6
rs604578	18	30,923,438	0.004	**9.5×10^{-5}**	MAPRE2

2c Linkage¶ Peaks with LOD scores ≥ 2.0

Trait	SNP closest to linkage peak	Chromosome	Physical location	1.5 – LOD support interval start	1.5 – LOD support interval end	LOD score
Age at natural menopause						
	rs1371217	**4**	182,890,808	178,671,796	186,905,362	**2.08**
	rs10509024	**10**	56,567,832	36,084,470	70,573,011	**2.39**
	rs4793513	**17**	66,429,892	60,635,492	69,512,021	**2.48**
Walking speed, exam 7						
	rs2769261	**1**	113,278,949	107,080,550	144,332,709	**2.30**
	rs921055	**2**	233,522,842	229,328,008	242,141,304	**2.13**
	rs2602044	**3**	109,427,883	102,255,525	111,604,019	**3.38**
	rs8011773	**14**	96,927,485	96,342,135	100,389,787	**2.69**
	rs1362626	**16**	4,489,227	205,160	10,344,522	**2.05**
Biologic age by osseographic scoring system						
	rs353810	**9**	86,258,745	81,441,976	92,220,168	**3.26**
	rs1203981	**16**	205,160	205,160	7,431,239	**2.49**
	rs2248383	**21**	35,151,811	27,412,716	40,940,879	**2.22**

SNP criteria: Autosomal SNPs with genotypic call rate ≥ 80%, minor allele frequency ≥ 10%, Hardy-Weinberg test p > 0.001, and ≥10 informative families for FBAT

* For each phenotype SNPs are ranked by p-value. A SNP in LD ($r^2 > 0.8$) with a higher ranked SNP, is identified with an asterisk. All SNPs for a phenotype are listed until 5 independent SNPs are identified. Thus, for some phenotypes more than 5 SNPs are listed. For the age at death trait, the FBAT analysis identified two areas on chromosome 1 in LD, with $r^2 = .5–.6$ between the two regions and r^2 of nearly 1.0 within the region.

† Multivariable-adjusted trait results are presented

‡Trait had <500 participants in the sample.

¶Results limited to traits presented

Table 3: All Significant SNP Associations (GEE or FBAT p-value < 0.01) for at least Two Traits

3a. FBAT: Age at Death and Morbidity-Free Survival at 65 years

Trait 1	Trait 2	SNP	Chr	Physical Position	Gene	Trait 1 GEE p-value	Trait 1 FBAT p-value	Trait 2 GEE p-value	Trait 2 FBAT p-value
Age at Death	**Morbidity-free at 65**	rs6682403	1	234,743,324		0.849	**0.004**	0.106	**0.004**
		rs10488907	4	113,669,709	ALPK1	0.452	**0.008**	0.336	**0.009**
		rs17190837	9	13,391,548		0.010	**0.009**	0.097	**0.004**
		rs4752977	11	47,257,005	**MADD**	0.736	**0.008**	0.002	**0.009**
		rs10506274	12	80,103,932		0.531	**0.001**	0.989	**0.001**
		rs2831154	21	28,059,331		0.323	**0.008**	0.358	**0.006**
		rs243725*	21	28,060,803		0.264	**0.007**	0.359	**0.008**

*r^2 > 0.80 with the preceding SNP

3b. GEE: Age at Death and Morbidity-Free Survival at 65 years

Trait 1	Trait 2	SNP	Chr	Physical Position	Gene	Trait 1 GEE p-value	Trait 1 FBAT p-value	Trait 2 GEE p-value	Trait 2 FBAT p-value
Age at Death	**Morbidity-free at 65**	rs9308261	1	113,603,160	MAGI3	**0.009**	0.156	**0.002**	0.900
		rs10490518	2	31,223,004	GALNT14	**0.009**	0.373	**0.010**	0.948
		rs2374983	7	94,516,375	**PPP1R9A/ PON1**	**0.006**	0.980	**0.007**	0.727
		rs655883	11	98,994,584	CNTN5	**0.006**	0.390	**0.001**	0.293
		rs1368850	11	130,433,518		**0.004**	0.387	**0.005**	0.292
		rs4943116	13	32,995,650	STARD13	**0.006**	0.205	**0.008**	0.049
		rs2254191	13	45,344,403		**0.007**	0.425	**0.004**	0.116
		rs1620210	13	45,759,488	C13orf18	**0.001**	0.161	**0.004**	0.068
		rs2823322	21	15,814,903		**0.0004**	0.045	**0.006**	0.029

3c. FBAT: Biologic Age and Walking Speed

Trait 1	Trait 2	SNP	Chr	Physical Position	Gene	Trait 1 GEE p-value	Trait 1 FBAT p-value	Trait 2 GEE p-value	Trait 2 FBAT p-value
Biologic age	Walking speed	rs873348	4	178,246,509		0.135	**0.004**	0.114	**0.006**
Biologic age	Walking speed	rs10520361*	4	178,247,037		0.074	**0.006**	0.054	**0.005**
Biologic age	Walking speed	rs31564	5	135,258,152	IL9	0.015	**0.001**	**0.008**	**0.002**
Biologic age	Walking speed	rs1862345	5	148,018,498	HTR4	0.172	**0.0002**	0.399	**0.008**
Biologic age	Walking speed	rs7844834	8	11,323,556	C8orf12|C8orf13	0.017	**0.004**	0.011	**0.004**
Biologic age	Walking speed	rs952658	12	20,756,568	SLCO1C1	0.024	**0.008**	0.935	**0.006**
Biologic age	Walking speed	rs6487366	12	23,994,617	SOX5	0.017	**0.003**	0.959	**0.004**
Biologic age	Walking speed	rs7135493	12	28,134,847		0.020	**0.006**	0.033	**0.004**
Biologic age	Walking speed	rs10492036	12	124,728,934		0.308	**0.005**	0.169	**0.009**
Biologic age	Walking speed	rs1978945	13	105,641,257		0.093	**0.009**	0.031	**0.006**
Biologic age	Walking speed	rs2165723*	13	105,641,610		0.046	**0.010**	0.027	**0.003**
Biologic age	Walking speed	rs10492651*	13	105,641,634		0.083	**0.009**	0.023	**0.001**
Biologic age	Walking speed	rs9301112*	13	105,642,018		0.097	**0.004**	0.053	**0.003**

* r^2 > 0.8 with the preceding SNP (calculated if the distance is <250,000 base pairs)

3d. GEE: Biologic Age and Walking Speed

Trait 1	Trait 2	SNP	Chr	Physical Position	Gene	Trait 1 GEE p-value	Trait 1 FBAT p-value	Trait 2 GEE p-value	Trait 2 FBAT p-value
Biologic Age	Walking speed	rs1474827	6	134,886,011		**0.007**	0.746	**0.000**	0.233
Biologic Age	Walking speed	rs10231641	7	119,166,342		**0.004**	0.077	**0.009**	0.278
Biologic Age	Walking speed	rs310575	8	51,603,257	SNTG1	**0.008**	0.209	**0.003**	**0.007**
Biologic Age	Walking speed	rs10520603	15	84,170,955		**0.009**	**0.008**	**0.003**	0.050
Biologic Age	Walking speed	rs7166323*	15	84,171,745		**0.009**	0.012	**0.003**	0.065
Biologic Age	Walking speed	rs2215921	16	9,604,834		**0.007**	0.049	**0.001**	0.316

* r^2 > 0.8 with the preceding SNP (calculated if the distance is <250,000 base pairs)

Table 4: All Significant SNP Associations with Selected Longevity Candidate Genes* (FBAT or GEE p-value < 0.01)

Trait	Gene	SNP	Chr	Physical Position	FBAT p-value	GEE p-value	SNP function	SNP position relative to gene (up to 60 kb)
Age at death	**FOXO1a**	**rs4943794**	13	40,071,408	0.068	**0.00028**	Intron	in
		rs10507486	13	40,084,501	0.043	**0.00013**	Intron	in
	GAPDH†	rs4764600	12	6,472,241	0.833	**0.005**	Locus/intron	near
	KL	rs683907	13	32,522,175	**0.009**	0.507	Intron	in
		rs687045	13	32,522,889	**0.007**	0.712	Intron	in
	LEPR	rs1475398	1	65,695,278	0.069	**0.005**	Untranslated	in
		rs1343981	1	65,757,349	0.031	**0.006**	Intron	in
		rs10493379	1	65,757,948	0.015	**0.004**	Intron	in
		rs2154380	1	65,769,462	**0.004**	**0.003**	Intron	in
		rs6669117	1	65,773,093	0.050	**0.007**	Intron	in
	PON1	**rs2374983**	7	94,516,375	0.980	**0.006**	Intron	near
	PSEN1	rs362356	14	72,708,382	**0.005**	0.130	Intron	in
	SOD2	**rs911847**	6	160,039,379	0.358	**0.005**	Unknown	near
	WRN‡	rs2543600	8	30,969,282	0.182	**4.2×10^{-6}**	Unknown	near
Morbidity-free survival at age 65	**GHR**	rs719756	5	42,761,386	**0.003**	0.676	Unknown	near
	LEPR	rs1171278	1	65,700,167	0.042	**0.003**	Untranslated	in
		rs3790426	1	65,755,040	0.460	**0.002**	Intron	in
	MORF4L1	rs1383636	15	76,893,275	0.458	**0.007**	Unknown	near
	PON1	**rs2374983**	7	94,516,375	0.727	**0.007**	Intron	near
		rs854523	7	94,542,884	0.850	**0.007**	Intron	in
	PTH	rs10500784	11	13,530,401	**0.010**	0.990	Unknown	near
	WRN‡	rs2725369	8	30,970,566	0.113	**0.003**	Unknown	near
Biologic Age by OSS	**FOXO1a**	rs1923249	13	40,041,881	**0.006**	**0.004**	Intron	in
		rs4943794	13	40,071,408	**0.009**	0.016	Intron	in
	HSPA9	rs256014	5	137,930,983	0.101	**0.005**	Intron	in
	LASS6	rs1002666	2	169,303,525	**0.001**	**0.008**	Intron	in
	SOD2	**rs911847**	6	160,039,379	0.024	**0.009**	Unknown	near
	TLR4	rs1927914	9	117,544,279	**0.007**	0.401	Locus	near
Walking speed	**ESR1**	rs9322361	6	152,551,257	0.124	**0.0089**	Intron	in
	LASS6	rs6433083	2	169,324,821	0.232	**0.006**	Intron	in
	NR3C1	rs2918418	5	142,703,566	**0.005**	0.081	Intron	in
		rs10515522	5	142,738,587	**0.004**	0.084	Intron	in
	SOD1	rs2833485	21	32,000,796	**0.008**	0.507	Locus/intron	in
	TERF2	rs728546	16	68,013,029	**0.0045**	0.533	Unknown	near
	FASLG	rs6700734	1	169,362,468	**0.003**	0.029	Intron	in

*79 genes identified from NCBI, SAGE ke, and GenAge databases; 12 genes with no SNPs on 100K chip; 67 genes with 1–45 SNPs on 100K chip; LGV1 2036 SNPs on 100K chip, results for this region available on the web
†The most strongly associated SNP near GAPDH is actually closer to MRPL51
‡The most strongly associated SNP near WRN is actually closer to PURG

related phenotypes and 100K SNP genotyping results available on the web. This resource has the potential to detect novel susceptibility genes for human longevity and aging and to examine the relevance of promising candidate gene associations reported in animal models to human aging. We describe several strategies to prioritize SNP associations in this unique resource to enhance the discovery of various genes and pathways that contribute to the control of human longevity. Furthermore, FHS investigators are part of the NIA sponsored Longevity Consortium http://www.longevityconsortium.org which offers the opportunity of collaboration with other investigators to replicate important findings in additional cohorts.

In our untargeted approach of ranking SNP associations by the strength of the p-value, 2 intronic *FOXO1a* SNPs were associated with age at death. One of these SNPs (rs4943794) also was associated with biologic age by OSS in our a priori evaluation of select candidate genes. FOXO (forkhead box group O) transcription factors are targets of insulin-like signaling and are involved in a diverse set of physiological functions including DNA repair and resistance to oxidative stress [33,34]. Further, FOXO plays a role in lifespan extension in *C. elegans* and *Drosophila* [35]. Studies of this gene in humans are limited; two case-control studies have not identified an association between *FOXO1a* and longevity [36,37]. However, the prospective population-based Leiden 85-plus Study found that *FOXO1a* was associated with increased mortality attributable to diabetes related deaths in participants aged 85 years and older [38]. The Leiden 85-plus Study also reported that genetic variation causing a reduction in insulin/IGF-1 signaling resulted in improved old age survival among women [20]. However, that report examined other genes in the insulin/insulin-like signaling pathway and did not specifically examine *FOXO1a*. Finally, the untargeted approach to SNP selection also identified a SNP near *FOXO3a* associated with age at natural menopause. This gene has been implicated in oocyte death,

depletion of functioning ovarian follicles, and infertility in mice [39,40] and thus represents a plausible candidate gene for menopause. Most positive common gene variant-disease association studies have failed replication [41] including reports on exceptional longevity. Haplotype-based fine mapping of the region on chromosome 4 linked to human longevity initially suggested the *MTP* gene, a gene important in lipoprotein synthesis, was associated with longevity [21]. However, this association failed replication in a French cohort of long-lived individuals and subsequent case-control studies of nonagenarians [22,42]. Beekman, *et al* [43] found neither linkage to chromosome 4 nor association with the MTP gene and longevity among nonagenarians in the Leiden Longevity Study. Meta-analyses implicated admixture of the control sample in the original report as an explanation for the presumed false-positive association. Thus, our findings are hypothesis generating and their importance can not be determined without evidence of consistent replication in other populations.

We examined pleiotropic effects by identifying SNP associations across two pairs of related traits. One SNP near *PON1* emerged as associated with both age at death and morbidity-free survival. Surprisingly, there were relatively few SNPs associated with both traits; prior work had suggested that longevity per se and healthy aging may share common genetic pathways [11,12]. However, morbidity-free survival was measured at age 65 years, it is possible that as our participants age morbidity-free survival defined at age 75 or 85 years will share additional SNP associations with our longevity trait, age at death. A SNP near *SOX5*, a gene potentially related to musculoskeletal function was associated with both biologic age by OSS and walking speed.

Our strategy of selecting SNPs in candidate genes and regions previously reported to be associated with longevity yielded interesting findings. For age at death, we identified SNPs in or near several genes including *KL*, *LEPR*, *PON1*, *SOD2*, and *WRN*. Defects in the *WRN* gene are the cause of Werner Syndrome, an autosomal recessive disorder characterized by premature aging. A longitudinal study of ageing Danish twins recently reported a possible association between a successful aging trait and 3 SNPs in the *WRN* gene [44]. We were unable to determine if our SNP (rs2725369) was in LD with the SNPs in the prior report because the SNPs were not included in HapMap. Mutations in the *KL* (Klotho) gene in the mouse lead to a syndrome resembling human aging [45-47]. There has been one report linking a functional variant of the *KL* gene to human longevity [15]. Thus, results from this GWAS may direct resources to the most relevant candidate genes and pathways for further investigation in humans. Several important limitations merit comment. First, we acknowledge that there may be a survival bias as participants in this sample had to survive to provide DNA (first systematic DNA collection began 1995) and hence are likely healthier than the full FHS sample. To ameliorate this issue, we adjusted for covariates using the full Framingham sample, and used the residual traits for the subset of individuals genotyped using the 100K Affymetrix GeneChip to test for association with the SNPs using linear regression models. Residual traits from Cox and logistic models typically are not ideally distributed for linear regression models, but our adjustment method using the full sample precludes the testing of SNP associations with age at death and morbidity-free survival using Cox and logistic models. Second, the 100K Affymetrix GeneChip provides limited coverage of the genome; many of our a priori candidate genes did not have any SNP coverage on the chip. For example, several genes that have been studied in model organisms or even in humans such as *ACE*, *Lamin A*, *SIRT2* and *SIRT3*, had no SNPs within 60 kb of the gene on the 100K Affymetrix GeneChip. However, genotyping is near-complete for the NHLBI funded 550 K genome-wide scan on all FHS participants. This will enable deeper exploration of our initial 100K SNP associations in a larger sample with denser coverage of the genome. Third, in this analysis we did not examine epistasis or gene-environment interactions which may modify the associations in this study. Importantly, this study is hypothesis generating. Our findings need to be replicated in other samples.

Conclusion

In summary, the untargeted genome-wide approach to detect genetic associations with longevity and aging traits provides an opportunity to identify novel biologic pathways related to lifespan control. GWAS also have the potential to direct investigators of human aging to the most promising candidate gene associations and biologic pathways reported to regulate lifespan in animal models. Enhancing our understanding of the mechanisms responsible for aging may in turn identify directions for health promotion and disease prevention efforts in middle-aged and older adults so that older persons can enjoy more time in good health. These data generate hypotheses regarding novel biologic pathways contributing to longevity and healthy aging and serve as a resource for replication of findings from other population-based samples.

Abbreviations

CVD = cardiovascular disease; FBAT = family-based association test; FHS = Framingham Heart Study; GEE = generalized estimating equations; GWAS = genome-wide association study; LD = linkage disequilibrium; LOD = logarithm of the odds; NCBI = National Center Biotechnology Information; OSS = osseographic scoring system;

SNP = single nucleotide polymorphism; TIA = transient ischemic attack.

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

Authors' contributions
All authors have made substantial contributions to conception and design or acquisition of phenotypic data. JMM, KL, EJB, CG, DK, DPK, JMM, MJP, RBD contributed to the analysis and interpretation of data. JMM, KL, EJB, DK, DPK, SS have been involved in drafting the manuscript or revising it critically for important intellectual content. All authors read and approved the final manuscript.

Additional material

Additional file 1
Candidate Gene List for FHS 100K Longevity and Aging Traits
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2350-8-S1-S13-S1.doc]

Acknowledgements
NHLBI's Framingham Heart Study is supported by contract number N01-HC-25195. This work was also supported in part by R01 AR/AG 41398 and AG028321.

FHS 100K analyses were conducted using the Boston University.

Linux Cluster for Genetic Analysis (LinGA) funded by the NIH NCRR (National Center for Research Resources) Shared Instrumentation grant 1S10RR163736-01A1 http://www.bu.edu/dbin/sph/departments/biostatistics/linga_publications.php.

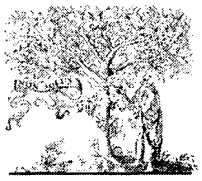
We thank the FHS participants for their ongoing participation and dedication to the study making this work possible.

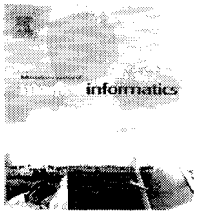
This article has been published as part of *BMC Medical Genetics* Volume 8 Supplement 1, 2007: The Framingham Heart Study 100,000 single nucleotide polymorphisms resource. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2350/8?issue=S1.

References
1. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW: **The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900.** *HumGenet* 1996, **97(3):**319.
2. Iachine IA, Holm NV, Harris JR, Begun AZ, Iachina MK, Laitinen M, Kaprio J, Yashin AI: **How heritable is individual susceptibility to death? The results of an analysis of survival data on Danish, Swedish and Finnish twins.** *TwinRes* 1998, **1(4):**196.
3. Ljungquist B, Berg S, Lanke J, McClearn GE, Pedersen NL: **The effect of genetic factors for longevity: a comparison of identical and fraternal twins in the Swedish Twin Registry.** *JGerontolA BiolSciMedSci* 1998, **53(6):**M441.
4. McGue M, Vaupel JW, Holm N, Harvald B: **Longevity is moderately heritable in a sample of Danish twins born 1870–1880.** *JGerontol* 1993, **48(6):**B237.
5. Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, Kaprio J, Pedersen NL, Christensen K: **Genetic influence on human lifespan and longevity.** *HumGenet* 2006, **119(3):**312.
6. Reed T, Dick DM: **Heritability and validity of healthy physical aging (wellness) in elderly male twins.** *TwinRes* 2003, **6(3):**227.
7. Carmelli D, Kelly-Hayes M, Wolf PA, Swan GE, Jack LM, Reed T, Guralnik JM: **The contribution of genetic influences to measures of lower-extremity function in older male twins.** *JGerontolA BiolSciMedSci* 2000, **55(1):**B49.
8. Reed T, Fabsitz RR, Selby JV, Carmelli D: **Genetic influences and grip strength norms in the NHLBI twin study males aged 59–69.** *AnnHumBiol* 1991, **18(5):**425.
9. Swan GE, Carmelli D, Reed T, Harshfield GA, Fabsitz RR, Eslinger PJ: **Heritability of cognitive performance in aging twins. The National Heart, Lung, and Blood Institute Twin Study.** *ArchNeurol* 1990, **47(3):**259.
10. Karasik D, Hannan MT, Cupples LA, Felson DT, Kiel DP: **Genetic contribution to biological aging: the Framingham Study.** *JGerontolA BiolSciMedSci* 2004, **59(3):**218.
11. Puca AA, Daly MJ, Brewster SJ, Matise TC, Barrett J, Shea-Drinkwater M, Kang S, Joyce E, Nicoli J, Benson E, Kunkel LM, Perls T: **A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4.** *ProcNatlAcadSciUSA* 2001, **98(18):**10505.
12. Reed T, Dick DM, Uniacke SK, Foroud T, Nichols WC: **Genome-wide scan for a healthy aging phenotype provides support for a locus near D4S1564 promoting healthy aging.** *JGerontolA BiolSciMedSci* 2004, **59(3):**227.
13. Cheng CL, Gao TQ, Wang Z, Li DD: **Role of insulin/insulin-like growth factor 1 signaling pathway in longevity.** *World JGastroenterol* 2005, **11(13):**1891.
14. Katic M, Kahn CR: **The role of insulin and IGF-1 signaling in longevity.** *Cell MolLife Sci* 2005, **62(3):**320.
15. Arking DE, Atzmon G, Arking A, Barzilai N, Dietz HC: **Association between a functional variant of the KLOTHO gene and high-density lipoprotein cholesterol, blood pressure, stroke, and longevity.** *CircRes* 2005, **96(4):**412.
16. Bellizzi D, Rose G, Cavalcante P, Covello G, Dato S, De Rango F, Greco V, Maggiolini M, Feraco E, Mari V, Franceschi C, Passarino G, De Benedictis G: **A novel VNTR enhancer within the SIRT3 gene, a human homologue of SIR2, is associated with survival at oldest ages.** *Genomics* 2005, **85(2):**258.
17. Browner WS, Kahn AJ, Ziv E, Reiner AP, Oshima J, Cawthon RM, Hsueh WC, Cummings SR: **The genetics of human longevity.** *AmJMed* 2004, **117(11):**851.
18. Hadley EC, Rossi WK: **Exceptional survival in human populations: National Institute on Aging perspectives and programs.** *MechAgeing Dev* 2005, **126(2):**231.
19. Rose G, Dato S, Altomare K, Bellizzi D, Garasto S, Greco V, Passarino G, Feraco E, Mari V, Barbi C, Bonafe M, Franceschi C, Tan Q, Boiko S, Yashin AI, De Benedictis G: **Variability of the SIRT3 gene, human silent information regulator Sir2 homologue, and survivorship in the elderly.** *ExpGerontol* 2003, **38(10):**1065.
20. van Heemst D, Beekman M, Mooijaart SP, Heijmans BT, Brandt BW, Zwaan BJ, Slagboom PE, Westendorp RG: **Reduced insulin/IGF-1 signalling and human longevity.** *Aging Cell* 2005, **4(2):**79.
21. Geesaman BJ, Benson E, Brewster SJ, Kunkel LM, Blanche H, Thomas G, Perls TT, Daly MJ, Puca AA: **Haplotype-based identification of a microsomal transfer protein marker associated with the human lifespan.** *ProcNatlAcadSciUSA* 2003, **100(24):**14115.
22. Nebel A, Croucher PJ, Stiegeler R, Nikolaus S, Krawczak M, Schreiber S: **No association between microsomal triglyceride transfer protein (MTP) haplotype and longevity in humans.** *ProcNatlAcadSciUSA* 2005, **102(22):**7906.
23. Hirschhorn JN, Daly MJ: **Genome-wide association studies for common diseases and complex traits.** *NatRevGenet* 2005, **6(2):**95.
24. Cupples LA, Arruda H, Benjamin EJ, D'Agostino RB Sr, Demissie S, DeStefano AL, Dupuis J, Falls K, Fox CS, Gottlieb DJ, Govindaraju DR, Guo CY, Heard-Costa N, Hwang SJ, Kathiresan S, Kiel DP, Laramie JM, Larson MG, Levy D, Liu CY, Lunetta KL, Mailman MD, Manning AK, Meigs JB, Murabito JM, Newton-Cheh C, O'Connor GT, O'Donnell CJ, Pandey MA, Seshadri S, Vasan RS, Wang ZY, Wilk JB, Wolf PA, Yang Q, Atwood LD: **The Framingham Heart Study 100K SNP genome-wide association study resource: Overview of**

17 phenotype working group reports. *BMC Med Genet* 2007, **8(Suppl 1):**S1.
25. **Survival following initial cardiovascular disease events: 30 year follow-up.** In *The Framingham Heart Study: An epidemiological investigation of cardiovascular disease* Edited by: *Kannel WB, Wolf PA, Garrison RJ. Bethseda, MD: NHLBI, NIH*; 1988.
26. Murabito JM, Yang Q, Fox C, Wilson PW, Cupples LA: **Heritability of age at natural menopause in the Framingham Heart Study.** *J Clin Endocrinol Metab* 2005, **90(6):**3427-3430.
27. Murabito JM, Yang Q, Fox CS, Cupples LA: **Genome-wide linkage analysis to age at natural menopause in a community-based sample: the Framingham Heart Study.** *Fertil Steril* 2005, **84(6):**1674-1679.
28. **SAS/STAT software, Version 8.2. SAS System for Windows. Copyright 1999–2001.** *SAS Institute Inc. Cary, NC, USA*; 1999.
29. Karasik D, Demissie S, Cupples LA, Kiel DP: **Disentangling the genetic determinants of human aging: biological age as an alternative to the use of survival measures.** *JGerontolA BiolSciMedSci* 2005, **60(5):**574.
30. 2006 [http://sageke.sciencemag.org/cgi/genesdb].
31. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ: **The UCSC Genome Browser Database.** *Nucleic Acids Res* 2003, **31(1):**51.
32. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12(6):**996.
33. Barthel A, Schmoll D, Unterman TG: **FoxO proteins in insulin action and metabolism.** *Trends EndocrinolMetab* 2005, **16(4):**183.
34. Lam EW, Francis RE, Petkovic M: **FOXO transcription factors: key regulators of cell fate.** *BiochemSoc Trans* 2006, **34(Pt 5):**722.
35. Hwangbo DS, Gershman B, Tu MP, Palmer M, Tatar M: **Drosophila dFOXO controls lifespan and regulates insulin signalling in brain and fat body.** *Nature* 2004, **429(6991):**562.
36. Bonafe M, Barbieri M, Marchegiani F, Olivieri F, Ragno E, Giampieri C, Mugianesi E, Centurelli M, Franceschi C, Paolisso G: **Polymorphic variants of insulin-like growth factor I (IGF-I) receptor and phosphoinositide 3-kinase genes affect IGF-I plasma levels and human longevity: cues for an evolutionarily conserved mechanism of life span control.** *JClinEndocrinolMetab* 2003, **88(7):**3299.
37. Kojima T, Kamei H, Aizu T, Arai Y, Takayama M, Nakazawa S, Ebihara Y, Inagaki H, Masui Y, Gondo Y, Sakaki Y, Hirose N: **Association analysis between longevity in the Japanese population and polymorphic variants of genes involved in insulin and insulin-like growth factor I signaling pathways.** *ExpGerontol* 2004, **39(11–12):**1595.
38. Kuningas M, Magi R, Westendorp RG, Slagboom PE, Remm M, van Heemst D: **Haplotypes in the human Foxo1a and Foxo3a genes; impact on disease and mortality at old age.** *EurJHumGenet* 2007, **15(3):**294.
39. Brenkman AB, Burgering BM: **FoxO3a eggs on fertility and aging.** *Trends Mol Med* 2003, **9(11):**464-467.
40. Castrillon DH, Miao L, Kollipara R, Horner JW, DePinho RA: **Suppression of ovarian follicle activation in mice by the transcription factor Foxo3a.** *Science* 2003, **301(5630):**215-218.
41. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K: **A comprehensive review of genetic association studies.** *Genet Med* 2002, **4(2):**45-61.
42. Bathum L, Christiansen L, Tan Q, Vaupel J, Jeune B, Christensen K: **No evidence for an association between extreme longevity and microsomal transfer protein polymorphisms in a longitudinal study of 1651 nonagenarians.** *EurJHumGenet* 2005, **13(10):**1154.
43. Beekman M, Blauw GJ, Houwing-Duistermaat JJ, Brandt BW, Westendorp RG, Slagboom PE: **Chromosome 4q25, microsomal transfer protein gene, and human longevity: novel data and a meta-analysis of association studies.** *JGerontolA BiolSciMedSci* 2006, **61(4):**355.
44. Sild M, Koca C, Bendixen MH, Frederiksen H, McGue M, Kolvraa S, Christensen K, Nexo B: **Possible associations between successful aging and polymorphic markers in the Werner gene region.** *AnnNYAcadSci* 2006, **1067:**309.
45. Kuro-o M, Matsumura Y, Aizawa H, Kawaguchi H, Suga T, Utsugi T, Ohyama Y, Kurabayashi M, Kaname T, Kume E, Iwasaki H, Iida A, Shiraki-Iida T, Nishikawa S, Nagai R, Nabeshima YI: **Mutation of the mouse klotho gene leads to a syndrome resembling ageing.** *Nature* 1997, **390(6655):**45.
46. Kurosu H, Yamamoto M, Clark JD, Pastor JV, Nandi A, Gurnani P, McGuinness OP, Chikuda H, Yamaguchi M, Kawaguchi H, Shimomura I, Takayama Y, Herz J, Kahn CR, Rosenblatt KP, Kuro-o M: **Suppression of aging in mice by the hormone Klotho.** *Science* 2005, **309(5742):**1829.
47. Masuda H, Chikuda H, Suga T, Kawaguchi H, Kuro O: **Regulation of multiple ageing-like phenotypes by inducible klotho gene expression in klotho mutant mice.** *MechAgeing Dev* 2005, **126(12):**1274.

INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS 76 (2007) 836–842





journal homepage: www.intl.elsevierhealth.com/journals/ijmi

The story behind the story: Physician skepticism about relying on clinical information technologies to reduce medical errors

Ann Scheck McAlearney[a,b,*], *Deena J. Chisolm*[b,c], *Sharon Schweikhart*[a], *Mitchell A. Medow*[d], *Kelly Kelleher*[b,c]

[a] *Division of Health Services Management and Policy, School of Public Health, The Ohio State University, 1841 Millikin Road, Cunz Hall 476, Columbus, OH 43210, United States*
[b] *Department of Pediatrics, The Ohio State University College of Medicine, United States*
[c] *Columbus Children's Research Institute, Columbus, OH, United States*
[d] *Division of General Internal Medicine, The Ohio State University College of Medicine, United States*

ARTICLE INFO

Article history:
Received 18 November 2005
Received in revised form
30 June 2006
Accepted 27 September 2006

Keywords:
Medical informatics
Attitude to computers
Information systems
Computerized physician order entry
Computers handheld
Medical errors

ABSTRACT

Purpose: In order to better understand physicians' perspectives about the use of clinical information technology (CIT) to reduce medical errors, we asked physicians about opportunities and issues around clinical use of computerized physician order entry (CPOE) systems, order sets within CPOE, and handheld computers (HHCs).
Methods: We conducted 10 focus groups including 71 physicians involved in technology implementation efforts across the US between April 2002 and February 2005.
Results: Two major themes emerged across focus groups around reliance on CIT to reduce errors: (1) can it work? and (2) at what cost to the medical profession? Within the first theme, physicians expressed concern about the appropriateness of physician-directed CIT as a solution for medical errors, concerns regarding the current technical capabilities and level of technical support for CIT solutions, and concern about the introduction of new errors. Within the second theme, physicians were particularly concerned about time efficiency and workload redistribution associated with the introduction of CIT. Across focus groups, physicians tended to generalize about the role of all IT in their lives, potentially biasing opinions about specific technologies.
Conclusions: Health care organizations attempting to promote physician use of CIT are advised to deepen consideration of physicians' perspectives about technology adoption and use in order to address their concerns, reduce skepticism, and increase the likelihood of implementation success.

1. Introduction

Given daunting estimates about the prevalence of medical errors in clinical medicine [1], the potential for CIT to prevent such errors is undeniably attractive. In particular, recommendations from the industry-based Leapfrog Group sparked tremendous interest in the application of CPOE systems [2–5], which have proved effective in reducing prescribing

* *Corresponding author at*: Division of Health Services Management and Policy, School of Public Health, The Ohio State University, 1841 Millikin Road, Cunz Hall 476, Columbus, OH 43210, United States. Tel.: +1 614 292 0662; fax: +1 614 438 6859.
E-mail address: mcalearney.1@osu.edu (A.S. McAlearney).
1386-5056/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2006.09.021

errors in hospital settings [6,7]. New physician-directed CIT, including CPOE, computerized order sets, and even handheld computers (HHCs), can help improve legibility and provide information to help with clinical decision making at the point of care [6,8–10]. For example, the CPOE process permits physicians to enter medical orders directly into a computer [11,12], providing the opportunity for monitoring medications, improving legibility, and checking for drug-drug interactions, among other functions. Such CPOE systems also introduce the potential to practice evidence-based care by providing aggregated order sets or embedded clinical guidelines [13], thereby reducing the likelihood of errors of omission. The use of HHCs in medicine offers potential to reduce medical errors by enabling physicians to check for drug-drug interactions and providing access to clinical data that can help physicians make medical decisions [9,10].

Yet while the potential for CIT to reduce the prevalence of medical errors appears great, the proliferation of such technologies in clinical practice is not particularly fast [14,15], nor without controversy [16–20]. All of these technologies challenge healthcare organizations on several fronts, including security, confidentiality, compatibility, capital, and coordination issues [9,16,17,21–26]. Several recent studies have highlighted new sources of errors associated with the introduction of CPOE [16,18–20], while the introduction of HHCs has raised concerns about information security, data accuracy, and physician dependency on the technologies themselves [9,10].

Most of these issues appear resolvable, and the benefits of physician-directed CIT innovations in reducing costs and errors are likely to accrue if these technologies can be successfully introduced and integrated into clinical practice, e.g. [29,30]. Yet within most stories about CIT implementation lurk lists of potential barriers, risks, and challenges associated with getting physicians to adopt and use the new technologies, e.g. [9,12,21,25–29,31]. We were interested in exploring these issues, and particularly the largely unnamed problem of physician skepticism, from the perspectives of physicians themselves. Focusing on three information technologies in particular, CPOE systems, order sets within CPOE systems, and HHCs, we asked both user and non-user physicians about their expectations for these technologies. While user perspectives may vary considerably across different technologies, we were interested in open discussion about CIT-related issues to uncover any consistent themes or issues. Our investigation was enabled by a multiple focus group methodology.

2. Methods

2.1. Study design

We used focus groups [32] to collect physician opinions about the use of physician-directed CIT across a range of CIT interventions. A multidisciplinary team of investigators gathered data from ten focus groups held between April 2002 and February 2005. These included focus groups discussing CPOE, computerized order sets, and HHCs. Consistent with the techniques of rigorous qualitative research [33,34], semi-structured focus group guides with open-ended questions were used to facilitate discussion of topics among participants. Our exploratory study was framed by our research question, "What do physicians think about the potential for clinical information technologies including CPOE systems, computerized order sets, and HHCs to reduce medical errors?" Focus group questions were designed to explore the issues of how, why, and when physicians did or would use CPOE, HHCs, or computerized order sets within a CPOE system.

2.2. Focus group process

Physicians were recruited by electronic mail messages inviting them to participate in our study, and were offered a token incentive of a gift certificate ($50 or $25) in appreciation for their willingness to participate. Sessions lasted 60–90 min and were conducted over a meal. Typically, focus group sessions were moderated by one study investigator with a co-moderator available to assist. Each session was audiotaped and transcribed verbatim to facilitate data analysis. We received ethical approval from the appropriate Institutional Review Boards of The Ohio State University and Columbus Children's Hospital.

2.3. Population studied

Combined, focus groups included 71 physicians purposely sampled so that informants represented both users and non-users of the different information technologies, across diverse clinical specialties and practice settings. Focus groups were conducted across the US in academic medical centers (2), community hospital systems (3), a children's hospital (2), an independent practice association (1), and at a regional meeting of physicians (2), the Society for General Internal Medicine annual meeting in Chicago, IL. Across focus groups, 30% of participants were women, 20% were specialists, and 32% were clinicians in full-time practice (68% were residents or fellows affiliated with a teaching institution).

2.4. Analysis

Nearly 250 pages of single-spaced transcripts were produced for analysis during the course of this research. Our analysis applied a combination of deductive and inductive methods [35], including a grounded theory approach [36,37]. During the study, we iteratively read transcripts and discussed findings, using the constant comparative method of analysis [34] to permit us to explore new ideas and themes that emerged in subsequent focus groups. We used common techniques to code data [35,38], identifying broad themes and patterns that emerged from the data. Discussions among study investigators led to consensus around the major themes and patterns that emerged across focus groups that we describe in this paper. We present the two major themes that emerged across all focus groups, along with representative patterns in the data that emerged in a majority of focus groups. We used the qualitative data analysis software program Atlas.ti (Version 4.2) [39] to support our analysis process.

3. Results

Two major themes emerged around physicians' expectations about the studied information technologies and their poten-

tial to impact medical errors: (1) skepticism regarding appropriateness and capability of CIT to substantially reduce medical errors; and (2) concern about the impacts of CIT on their work and lives. Simply, these concerns can be summarized as "Can it work?" and, "At what cost to the medical profession?" These two themes are each presented below, followed by our discussion and a conclusion to the article.

3.1. *Theme 1: can it work?*

Across focus groups, a strong theme that emerged from the data was that physicians remain skeptical about the ability of physician-directed CIT to reduce medical errors. Three patterns emerged around this theme: (1) concerns regarding the appropriateness of physician-directed CIT as a solution for medical errors; (2) concerns regarding the current technical capabilities and level of technical support for CIT solutions; and, (3) concern about the introduction of new errors.

3.1.1. *Appropriateness*

First, in most focus groups, several respondents noted that it appeared unclear that the CIT solutions implemented were appropriately matched to the problems of medical errors. As one physician explained, "Frankly most docs have to be convinced that CPOE specifically addresses patient safety issues because there is still some decent data, at least our data locally and some of the stuff I have read, that still seem to show that a heck of a lot of errors occur in the actual administration of the medication." Another physician describing order sets within CPOE noted how, "having never been on the other side of the order, it is still a real big mystery there that could lead to errors." Similarly, physician participants were, for the most part unconvinced that HHCs could have a measurable impact on medical errors, even though there was consensus that "it should." Other physicians offered stories to illustrate their skepticism. As one explained, "In 20 years as a clinician I can count ... on fewer than five fingers the numbers of times that verbal orders have been taken down wrongly that caused anything even close to a significant ill effect on a patient. ... So you can say verbal orders have to be changed—that it is a patient safety issue. And from a doc's standpoint, what they are going to say is, 'Prove to me that verbal orders have caused bad patient outcomes.'

3.1.2. *Technical capability*

A second pattern emerged around the technical capability of current CIT implementations to facilitate error reductions. Across focus groups, both hardware and software applications were soundly criticized for their limitations. Consequently, many physicians reported that they were waiting for proof that change is necessary, or that a new technology will actually help. As one commented, "My philosophy towards technology has always been that the role of technology is to serve me and not the other way around." This shared sentiment was reflected across focus groups when respondents described resistance to, rather than wholehearted endorsement of, the different CIT innovations. Consensus among focus group participants showed that many existing technologies are not yet sufficiently useful, and that many physicians are not yet sufficiently comfortable with the IT to ensure any appreciable benefit from IT use. Further, with the exception of the perspectives of self-described "technophiles" or "power users"[10], many technology limitations proved insurmountable for physicians frustrated by their inability to do what they want to do with the technology both efficiently and effectively.

Skepticism about technology capability was particularly apparent in two main areas: (1) system accessibility; and, (2) system design. First, the ability to access CIT systems quickly and easily was repeatedly noted as an issue. In the words of one physician, "You have med students, fellows, occasional attending physicians, seniors, and interns that are doing things in the morning, and there's often not enough computers." Another summarized, "if you try to do things at shift change, it's a nightmare." These limitations could create problems with work-arounds, as physicians searching for available computers make compromises. For example, one physician explained, "And every once in a while if you have just one order that you need to put in you may ask someone else to just put in for you because they're already logged in. But, then again that's not ideal either because if they're not following that patient and they get a call from pharmacy asking, 'do you really want that dosing?' it's hard to say, 'well, actually, someone else wanted me to put in that order." Similarly, physicians typically reliant on HHCs to look up medications reported tending to "guess" or "cross my fingers" when they could not access information on the device rather than make the effort to find the appropriate reference manual. Limitations on battery life for laptops and HHCs and time required for complex sign-on processes were also noted as challenges in accessing existing systems, consistent with prior research [8,9,40–43].

In addition, accessibility problems associated with system down time were a serious concern. Across groups, participants commented about the problems created when the CIT device or system failed. As one participant described a CPOE system failure, "you have nurses who don't know how to do charting on paper when the system is down and they have to go back and talk to the older nurses or other people about how did you chart this information?" Another explained how "the system was turned off last night to update and it was updating this morning for 2½ h and it was chaos." Similarly, multiple physician respondents provided examples of how they "panicked" when their HHC crashed or "had to go home" when they had forgotten the device, contributing to skepticism about reliance on CIT.

A second issue regarding the technical capacity for CIT to reduce errors emerged around system design and included comments about both the content design and maintenance within CIT systems. With computerized order sets, physicians expressed concern that the evidence used in developing order sets and alerts could be out of date, incomplete, or incorrect, and reported that this fear has led some clinicians to limit use of decision support, one of the key elements in medical error reduction. As one participant explained of the order sets, "Some of them are really good and some are just very incomplete." With CPOE, one physician noted how, "all of our medications are dependent on the weight. And there's nothing that warns you that the weight has not been done." Others expressed frustration with not really understanding the CPOE system itself, and reported learning about system

enhancements from peers and colleagues more frequently than through any other avenue. With HHCs, several physicians reported that they did not always trust the information available in ePocrates (the common drug reference database), and many made comments about the fact that "things are always changing."

3.1.3. *New errors*

Finally, in several focus groups, participants provided examples of how the new technologies could actually lead to new types of errors. As one participant explained of a CPOE system, "Sometimes there are defaults so that if you're entering an order quickly, you can put in a number like 6 but it will be bumped to 60 because of the way the order is written. So you actually have to type 06." Another example was provided in explaining the use of the comments field in the CPOE system. As one physician explained, "Those comments are often ignored ... if you put in a lab for 2 days from now and you write "please draw 2 days from now," it gets drawn that night." One physician particularly resistant to CPOE adoption predicted problems associated with "the classic law of unintended consequences," and explained how by eliminating the option of verbal orders, "you are going to have a delay in care." Discussion of the sources and potential for new errors consistently mentioned both CPOE systems and order sets, with less discussion about observed errors using HHCs. Across groups, consensus appeared that these CIT innovations had undeniable potential, but, as one participant summarized, "I think it decreases the errors we know about and it increases the errors we don't know about."

3.2. *Theme 2: at what cost to the medical profession?*

A second major theme emerged across focus groups around physicians' concerns about the impact of CIT on their lives as physicians, thus tempering their enthusiasm about CIT as a medical error reduction strategy. Two patterns of concern involved time efficiency and workload distribution.

First, many physicians in our focus groups were concerned about the fact that the adoption and use of new technologies actually took more time than familiar manual processes. With order sets and CPOE systems, these changes were particularly apparent. Specifically, when the order set or system was limited, this was frustrating to physicians. As one explained, "if there are things significantly absent, it's not even quicker. There are just too many things to add." In contrast, several physicians noted that "if it's a good order set, it significantly saves time." Especially in the early stages of technology introduction, physicians noted how the IT involved more time. As one explained, "you still have a tendency to want to call downstairs" to verify if a result is unavailable "because you still don't believe." Participants also reported an association between problems with system access and efficiency gains. As one commented, "The efficiency that we've gained with these computers just goes down the tubes when we can't find a computer to enter orders in." Perspectives about HHC use were particularly mixed, with many physicians reporting efficiency gains and others reporting none. Consensus across groups and technologies emerged as one physician summarized: "I think it changes time. But I don't think it's faster in terms where we're getting to see more. I mean we're talking minutes."

The availability of timely technical support was also noted as a threat to physicians' efficiency, and physicians' comments about the lack of immediate technical support were fairly common. As one frustrated physician complained, "Because you're putting that order set in and you're trying to admit the patient and you have three more to admit after that. And you have to wait for 10 minutes for them to call you back ... I know 10 minutes is a short amount of time for an IS department, but for us, I mean we're already doing two things after that." Another participant explained how, "to call someone and make the complaint and get it in just extends the amount of time it takes." This frustration thus potentially limits the ability of the organization itself to respond to and correct problems that may remain unreported.

Another pattern of concern about these technologies involved a shift in physicians' workloads. Across all three information technologies, physicians noted that their use of the technologies resulted in physicians becoming more responsible for actual data entry than had been the case in the past. Instead of relying on nurses, clerks, or other ancillary personnel to support their clinical practice, physicians using these new technologies found themselves doing direct data entry. As one physician explained, "Where is the payback for us? We are trying to be cooperative. We are trying to be good soldiers, but where is the payback for us? That is kind of the theme it seems to me as to why a lot of CPOE and other initiatives are frowned upon. I don't think it is because of the old cliché that docs are techno-phobic. I don't think that is the truth. How does it at least make my life time-neutral?" However, not all respondents were completely negative. Instead, several commented about how with CPOE systems, "now you have the more medically trained individuals doing the orders instead of the unit clerk saying, well, I think it might be this and I think those might be this." Yet while electronic data entry processes clearly have the potential to improve both the legibility and the accuracy of data, the experienced shift in workload was not well-appreciated by the majority of our physician participants.

4. Discussion

Across focus groups, many physicians remain skeptical about the ability of new information technologies to reduce medical errors and improve their lives. For each technology, a myriad of different factors can result in a personal non-adoption or abandonment decision, or can lead to collective resistance toward a technology, at worst. Among these physicians, their own prior experiences with technology, both good and bad, seemed to weigh disproportionately. Especially when judged on the basis of time savings or improvements in care and service to patients, many of our physician informants remained unconvinced about the potential of new CIT. Further, given the difficulties already associated with convincing physicians to adopt CIT, e.g. [9,11,12,44], this remaining skepticism can create even greater challenges, potentially affecting the ability of organizations to both successfully implement new technologies and to foster organizational culture change [20,45] around the issue of reducing medical errors.

Unintended consequences of various information technology applications have been documented in the areas of both data entry and data retrieval, and these represent a source of new errors in medicine that become new problems to solve [18,19,46]. In our focus groups, physicians noted opportunities for these new sources of errors, and expressed concern about the inherent trade-offs. Thus consistent with recent research [18,19,23,44–50], it remains critical to carefully study the results of IT implementation, considering both intentional and unintentional solutions and problems.

Our finding about increased time requirements was consistent with other published research reporting that physician usage of new information technologies requires changes in practice patterns and workflow, and may increase the time providers typically spend on documentation activities [12,24,50,51]. Further, despite increasing evidence that time savings might appear in the future due to these new technologies, the reported skepticism among many of our physician respondents appeared to limit their willingness to experiment with the new technologies, instead causing them to stick to the basics required or what they actually know will work. Consequently, in practical CIT implementations where physicians are not mandated to use new technologies, physician resistance to adoption and use may not be surprising.

Perhaps more troubling is our perception that physicians tend to generalize about the role of electronic information in their lives. Instead of evaluating each CIT on its own merits, we found that physicians resistant to the new technologies were resistant to all new technologies—not only the individual CIT being introduced. In practice, this "lumping" phenomenon could have positive or negative consequences, depending on their perspectives about the CIT. For instance, for physicians coming out of medical schools and residency programs where CIT innovations had been prevalent, resistance to adopting or using a new technology affected only the particular version of the order set, CPOE system, or HHC they were using. They fully expected CIT to be part of their clinical practice, and focused complaints on the specifics of a technology. In contrast, physicians who were more techno-phobic or who had negative experiences with past technology implementations tended to report more frustration, complain more, and appear less open to change, despite the apparent inevitability of CIT in clinical practice.

Given physicians' apparent tendency to generalize across technologies, it may be advantageous for health care organizations planning CIT implementations to consider the inherent variability across target physician users. For instance, it may be possible to tailor implementation strategies, making a distinction between physicians with positive pre-conceptions and those with negative pre-conceptions. By segmenting target users, information systems departments could work to have appropriate responses ready for the different user populations, especially with respect to anticipated concerns about time requirements, CIT impact, and availability of support. A customized approach in practice could streamline the CIT implementation process for many users already "convinced" about the value of IT, but provide additional information and feedback to address the remaining skepticism among "unconvinced" physicians.

On a more positive note, the healthy dose of skepticism remaining among current and potential users of new CIT may help drive both product and process improvements around CIT that can, in turn, improve the quality of care supported by the new technologies. Across focus groups, physicians' willingness to provide input about what technology features are most helpful and what capabilities they would like was striking. Including interested physicians in technology-aided process improvements and in developing decision support tools such as order sets or HHC databases will likely continue to help increase physicians' usage and acceptance of such technologies. Given that CIT implementation appears to proceed in a never-ending cycle of product enhancements, exploring and accommodating physicians' perspectives and interests could conceivably help both product developers and organizations involved in technology implementation in order to ensure that physicians' needs are best met and to maximize the opportunity to reduce physician skepticism.

Finally, our data highlight the importance of considering physicians' perspectives about CIT and its role in their lives. While brief user satisfaction surveys may satisfy vendors and health care organizations, our focus groups reveal additional concerns about the various technologies that our deeper exploration of CIT adoption and use uncovered. Health care administrators, and information systems personnel in particular, would be advised to take into account these perspectives when attempting to introduce and implement new or enhanced CIT within physicians' clinical practices. If not, our study suggests the potential for downstream problems with adoption and use that are currently reflected in physician skepticism about the role of CIT in reducing medical errors.

4.1. *Limitations*

Several important limitations of this research must be acknowledged. First, we realize that combining the study of different technologies across focus groups risks losing the nuances associated with findings about the individual technologies. While we have attempted to report areas of consensus and dissent within and across focus groups, we are aware that these individual technologies are markedly different, and generalizations about CIT as a category may be difficult to support. Second, it is possible that our purposeful sampling methodology has led to bias in our results. While consistent with the standards of rigorous qualitative research, this methodology limits our ability to generalize our findings, also admittedly a limitation of much qualitative research. Third, the design of focus groups introduces the possibility that respondents can influence each other in their discussion of a particular topic. The themes we discuss in this paper emerged from issues raised in the focus group discussions, but it is possible that the apparent salience of a particular issue was influenced by the discussion process itself. In addition, it is possible that some participants did not feel comfortable sharing their opinions given the non-confidential nature of the focus group discussions. Finally, our study of individual technologies at distinct points in time is not ideal. Especially in the context of a rapidly changing health care industry flooded with technology innovation, capturing physicians' perspectives at one point in time does not permit us to consider changes

Summary points

What is already known on this topic:

- While considerable research supports the important potential for clinical information technologies (CIT) to reduce medical errors, little research has examined physicians' perspectives about these expectations for CIT in general, or for specific types of CIT.
- No prior research has specifically explored the areas of concern physicians may have about the important role for CIT in error reduction.

What this study adds:

- Physicians who use CIT appear skeptical about the ability of such technologies to reduce medical errors without introducing new problems.
- Concerns about new CIT include the realities of shifting workload and the limited abilities of technologies to improve practice efficiency.
- Given physicians' tendencies to generalize their opinions about CIT across all technologies, it may be especially important for organizations to address physicians' concerns as they work to improve CIT implementation success.

in viewpoints that may occur over time and that might be researchable using a longitudinal study design.

5. Conclusions

Policymakers, organizations, and providers relying on the promise of CIT to solve problems and reduce medical errors must be aware of the limitations of such technologies, and remain cognizant of the many issues associated with adoption and implementation of such technologies. At this time in the evolution of various CIT, there is still a profound risk that exaggerated promises about the potential of new technologies may be misdirecting institutional focus toward fairly trivial tactics that only minimally address the non-trivial problem of medical errors. Further, if existing physician skepticism is not recognized and addressed, the true potential for CIT to reduce medical errors through widespread adoption and usage will remain limited.

Acknowledgements

The authors are extremely grateful to the Center for Health Management Research and to Columbus Children's Research Institute which both funded portions of the study. We also thank our research associates, Robynn Young and Tracy Bryan Mullis, for their help with transcribing.

Preliminary findings about physician skepticism were presented at the INFORMS Conference, Denver, Colorado, 2004, the AcademyHealth Annual Research Meeting in Boston, Massachusetts, June 2005, and the Academy of Management National Meeting in Honolulu, Hawaii, August 2005. There is no conflict of interest associated with this manuscript.

REFERENCES

[1] L. Kohn, J. Corrigan, M. Donaldson (Eds.), To Err is Human: Building a Safer Health System, National Academy Press, Washington, DC, 1999.
[2] The Leapfrog Group, Fact Sheets. Available at: http://www.leapfroggroup.org/FactSheets.htm. Accessed January 20, 2005.
[3] G. Schiff, T. Rucker, Computerized prescribing: building the electronic infrastructure for better medication usage, JAMA 279 (1998) 1024–1029.
[4] B. Martinez, Business consortium to launch effort seeking higher standards at hospitals, Wall Street J. (2000) A3.
[5] B. Martinez, Group pushes to cut medical errors, Wall Street J. (2002) A4.
[6] D. Bates, L. Leape, D. Cullen, N. Laird, L.A. Petersen, J.M. Teich, E. Burdick, M. Hickey, S. Kleefield, B. Shea, M. Vander Vliet, D.L. Seger, Effect of computerized physician order entry and a team intervention on prevention of serious medication errors, JAMA 280 (1998) 1311–1316.
[7] D. Bates, J. Teich, J. Lee, D. Seger, G.J. Kuperman, N. Ma'Luf, D. Boyle, L. Leape, The impact of computerized physician order entry on medication error prevention, J. Am. Med. Inform. Assoc. 6 (1999) 313–321.
[8] A.S. McAlearney, D. Chisolm, S. Veneris, D. Rich, K. Kelleher, Utilization of evidence-based computerized order sets, Int. J. Med. Inform. 75 (2007) 501–502.
[9] A.S. McAlearney, S.B. Schweikhart, M.A. Medow, Organizational and physician perspectives about facilitating use of handheld computers in clinical practice: results of a cross-site qualitative study, J. Am. Med. Inform. Assoc. 12 (2005) 568–575, doi:10.1197/jamia.M1816.
[10] A.S. McAlearney, S.B. Schweikhart, M.A. Medow, Doctors' experience with handheld computers in clinical practice: qualitative study, BMJ 328 (2004) 1162–1165.
[11] J. Ash, P. Gorman, M. Lavelle, T.H. Payne, T.A. Massaro, G.L. Frantz, J. Lyman, A cross-site qualitative study of physician order entry, J. Am. Med. Inform. Assoc. 10 (2003) 188–200.
[12] R.A. Foster, P.J. Antonelli, Computerized physician-order entry: are we there yet? Otolaryng. Clin. N. Am. 35 (2002) 1237–1243.
[13] T. Payne, P. Hoey, P. Nichol, C. Lovis, Preparation and use of preconstructed orders, order sets, and order menus in a computerized prescriber order entry system, J. Am. Med. Inform. Assoc. 10 (2003) 322–329.
[14] D. Berwick, Disseminating innovations in health care, JAMA 289 (2003) 1969–1975.
[15] Institute of Medicine, Crossing the Quality Chasm, National Academy Press, Washington, DC, 2001.
[16] R. Berger, J. Kichak, Computerized physician order entry: helpful or harmful? J. Am. Med. Inform. Assoc. 11 (2004) 100–103.
[17] G.J. Kuperman, R.F. Gibson, Computer physician order entry: benefits, costs, and issues, Ann. Intern. Med. 139 (2003) 31–39.
[18] J.S. Ash, M. Berg, E. Coiera, Some unintended consequences of information technology in health care: the nature of patient care information system-related errors, J. Am. Med. Inform. Assoc. 11 (2004) 104–112.
[19] R. Koppel, J.P. Metlay, A. Cohen, B. Abaluck, A.R. Localio, S.E. Kimmel, B.L. Strom, Role of computerized physician order

entry systems in facilitating medical errors, JAMA 293 (2005) 1197–1203.
[20] R. Wears, M. Berg, Computer technology and clinical work: still waiting for Godot, JAMA 293 (2005) 1261–1263.
[21] J. Ash, P. Gorman, V. Seshadri, W. Hersch, Perspectives on CPOE and patient care, J. Am. Med. Inform. Assoc. 11 (2004) 104–112.
[22] R. Kaushal, K. Shojania, D.W. Bates, Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review, Arch. Intern. Med. 163 (2003) 1409–1416.
[23] A. Bobb, K. Gleason, M. Husch, J. Feinglass, P. Yarnold, G. Noshkin, The epidemiology of prescribing errors, Arch. Intern. Med. 164 (2004) 785–792.
[24] F. Lee, J. Teich, C. Spurr, D. Bates, Implementation of physician order entry: user satisfaction and self-reported usage patterns, J. Am. Med. Inform. Assoc. 3 (1996) 42–55.
[25] R. Ferner, Computer aided prescribing leaves holes in the safety net, BMJ 328 (2004) 1172–1173.
[26] R. Shane, Computerized physician order entry: challenges and opportunities, Am. J. Health Syst. Pharm. 59 (2002) 286–288.
[27] E.G. Poon, D. Blumenthal, T. Jaggi, M.M. Honour, D.W. Bates DW, R. Kaushal, Overcoming barriers to adopting and implementing computerized physician order entry systems in U.S. hospitals, Health Aff. 23 (2004) 184–190.
[28] R.H. Miller, I. Sim, Physicians' use of electronic medical records: barriers and solutions, Health Aff. 32 (2004) 116–126.
[29] D.W. Bates, M. Cohen, L.L. Leape, J.M. Overhage, M.M. Shabot, T. Sheridan, Reducing the frequency of errors in medicine using information technology, J. Am. Med. Inform. Assoc. 8 (2001) 299–308.
[30] D.W. Bates, A. Gawande, Improving safety with information technology, NEJM 348 (2003) 2526–2534.
[31] S. Maviglia, R. Zielstorff, M. Paterno, J. Teich, D. Bates, G. Kuperman, Automating complex guidelines for chronic disease: lessons learned, J. Am. Med. Inform. Assoc. 10 (2003) 154–165.
[32] D.L. Morgan, Focus Groups as Qualitative Research, Sage, Thousand Oaks, CA, 1996.
[33] J. Maxwell, Qualitative Research Design, Sage, Thousand Oaks, CA, 1996.
[34] F. Crabtree, W. Miller, Doing Qualitative Research, Sage, Thousand Oaks, CA, 1999.
[35] M. Miles, A. Huberman, Qualitative Data Analysis, Sage, Thousand Oaks, CA, 1994.
[36] B. Glaser, A. Strauss, The Discovery of Grounded Theory: Strategies for Qualitative Research, Aldine de Gruyter, New York, 1967.
[37] A. Strauss, J. Corbin, Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, Sage, Thousand Oaks, CA, 1998.
[38] M. Constas, Qualitative analysis as a public event: the documentation of category development procedures, Am. Educ. Res. J. 29 (1992) 253–266.
[39] Scientific Software Development, Atlas.ti, in. 4.2 ed. Scientific Software Development, Berlin, 1998.
[40] S. Fischer, T. Stewart, S. Mehta, R. Wax, S. Lapinsky, Handheld computing in medicine, J. Am. Med. Inform. Assoc. 10 (2003) 139–149.
[41] S. Schneider, R. Kostecke, J. Tokazewski, Buying your first PDA, Fam. Pract. Manag. 8 (2001) 50–51.
[42] P. Embi, Information at hand: using handheld computers in medicine, Cleve Clin. J. Med. 68 (2001) 840–849.
[43] P. Wright, C. Bartram, N. Rogers, H. Emslie, J. Evans, B. Wilson, S. Belt, Text entry on handheld computers by older users, Ergonomics 43 (2000) 702–716.
[44] T. Payne, The transition to automated practitioner order entry in a teaching hospital: the VA Puget Sound experience, Proc. AMIA Symp. (1999) 589–593.
[45] N. Lorenzi, R. Riley, Managing change: an overview, J. Am. Med. Inform. Assoc. 7 (2000) 116–124.
[46] C.J. McDonald, J.M. Overhage, B.W. Mamlin, P.D. Dexter, W.M. Tierney, Physicians, information technology, and health care systems: a journey, not a destination, J. Am. Med. Inform. Assoc. 11 (2004) 121–124.
[47] E.S. Patterson, R.I. Cook, M.L. Render, Improving patient safety by identifying side effects from introducing bar coding in medication administration, J. Am. Med. Inform. Assoc. 9 (2002) 540–543.
[48] R. Cook, M. Render, D. Woods, Gaps: learning how practitioners create safety, BMJ 320 (2000) 791–794.
[49] A. Tucker, A. Edmondson, Why hospitals don't learn from failures: organizational and psychological dynamics that inhibit system change, Calif. Manage. Rev. 45 (2003) 55–72.
[50] J.M. Overhage, S. Perkins, W.M. Tierney, C.J. McDonald, Controlled trial of direct physician order entry: effects on physicians' time utilization in ambulatory primary care internal medicine practices, J. Am. Med. Inform. Assoc. 8 (2001) 361–371.
[51] D. Bates, K. Shu, D. Narasimhan, J. Horsky, Comparing time spent writing orders on paper and physician computer order entry, Proc. AMIA Symp. (2000) 265.

ORIGINAL ARTICLES

Sources of Variation in Physician Adherence with Clinical Guidelines: Results from a Factorial Experiment

J. B. McKinlay[1,3], *C. L. Link*[1], *K. M. Freund*[2], *L. D. Marceau*[1], *A. B. O'Donnell*[1], *and K. L. Lutfey*[1]

[1]New England Research Institutes, 9 Galen Street, Watertown, MA 02472, USA; [2]Women's Health Unit, Evans Department of Medicine, and Women's Health Interdisciplinary Research Center,Boston University School of Medicine, Boston, MA, USA; [3]Institute for Community Health Studies,New England Research Institutes, 9 Galen Street, Watertown, MA 02472, USA.

BACKGROUND: Health services research has documented the magnitude of health care variations. Few studies focus on provider level sources of variation in clinical decision making-for example, which primary care providers are likely to follow clinical guidelines, with which types of patient.

OBJECTIVES: To estimate: (1) the extent of primary care provider adherence to practice guidelines and the unconfounded influence of (2) patient attributes and (3) physician characteristics on adherence with clinical practice guidelines.

DESIGN: In a factorial experiment, primary care providers were shown clinically authentic video vignettes with actors portrayed different "patients" with identical signs of coronary heart disease (CHD). Different types of providers were asked how they would manage the different "patients" with identical CHD symptoms. Measures were taken to protect external validity.

RESULTS: Adherence to some guidelines is high (over 50% of physicians would follow a third of the recommended actions), yet there is low adherence to many of them (less than 20% would follow another third). Female patients are less likely than males to receive 4 of 5 types of physical examination ($p<.03$); older patients are less likely to be advised to stop smoking ($p<.03$). Race and SES of patients had no effect on provider adherence to guidelines. A physicians' level of experience (age) appears to be important with certain patients.

CONCLUSIONS: Physician adherence with guidelines varies with different types of "patient" and with the length of clinical experience. With this evidence it is possible to appropriately target interventions to reduce health care variations by improving physician adherence with clinical guidelines.

Received February 22, 2006
Revised June 19, 2006
Accepted November 10, 2006
Published online January 9, 2007

KEY WORDS: clinical decision making; guidelines; disparities.
DOI: 10.1007/s11606-006-0075-2
 2007;22:289–296

Social and behavioral scientists have begun to identify different nonmedical influences on provider clinical decision making. For a range of common medical problems, clinical decision making appears to be influenced as much by who the patient is (their age, race, or gender), which provider they encounter, and the type of organizational setting in which medical care is provided as it is by the signs and symptoms of the problem itself [1–3]. Some studies show that only about half of those who utilize the health system actually receive the recommended processes of medical care [4]. Recently, attention has turned to the quality of medical decisions–that is, the extent to which specific providers adhere to appropriate medical practice guidelines (test ordering, prescribing behavior, life style recommendations, referrals, and patient follow-up) for common medical problems presented by specific types of patients [5]. This paper focuses on specific sources of variation in physician adherence to clinical guidelines: that is, which primary care providers are likely to evidence appropriate clinical practice, with which types of patients? Evidence on these specific sources of provider variation is required to ensure that interventions to improve the quality of care and reduce health variations are appropriately targeted. First, we sought to identify the extent to which providers adhere to clinical guidelines in their clinical decision making; second, we examine the influence of patient and provider factors on clinical decisions making. Is variation associated with specific patients? Is it associated with particular providers?

METHODS

The objective of this research was to estimate the unconfounded influence (either singly or in combination) of: (a) patient attributes (age, gender, race, and socioeconomic status) and (b) provider characteristics (physician gender and years of clinical experience) on medical decision making when providers are presented "patients" who show identical signs and symptoms strongly suggestive of coronary heart disease (CHD) (a common medical problem). We conducted a factorial experiment that permits estimation of unconfounded main effects and interactions of any

2 of the variables listed above. The experiment focused on a range of clinical decisions as they relate to a case of CHD [6, 7]. Research methods are summarized below (see also McKinlay et al 2006, [8].

A version of the videotaped CHD condition (varying by age, race, or gender) was shown to each of the physicians recruited as subjects for the experiment. We recruited a total of 128 Massachusetts licensed, randomly sampled internists and family physicians. We stratified subjects according to gender and level of clinical experience, including eligible physicians until each cell was complete. The "patient" (reluctantly made the appointment) presents with a complaint of "indigestion" and features of chronic atypical chest pain. For the estimation of main effects, a total sample of 128 physicians gives 80% power to detect an absolute difference in means of 25%. For 2-way interactions, it provides 80% power to detect an effect size of .25. The effect size is a ratio of the variability of the hypothesized means divided by the variability of the observations. For 2 means with a difference Δ, standard deviation of subjects s, the effect size $\Delta/2s$ [9]. Immediately after viewing the selected video for the experiment, the experimental subjects completed a semistructured interview. This interview included questions concerning how they would manage the case of CHD depicted in the video in their everyday clinical practice, including their most likely diagnoses, their certainty with respect to the diagnosis, test ordering, prescriptions, lifestyle recommendations, and what other information they might seek. Qualitative techniques were employed to elicit the reasons why decisions were made.

Experimental Stimuli (Scenarios)

Professional actors and actresses were recruited and trained (under experienced physician supervision) to realistically portray a "patient" presenting to a primary care provider with the signs and symptoms of CHD. Sixteen versions of the scenario were videotaped, systematically varying the "patient's" age (55 vs. 75 years to get some separation between the middle aged and older patients), race (white vs. black), gender and socioeconomic status (lower vs. higher social class—a janitor vs. a teacher). Potentially relevant nonverbal indicators were embedded in the script, such as the "Levine fist." Each videotaped encounter simulated an initial interview with either an internist or family practitioner and was of 7–8 minutes in duration, reflecting the average length of a consultation (face time) with a primary care physician (not including a physical exam [10]).

CHD was selected because: (a) it is among the most common and costly problems presented by older patients to primary care providers [11]; (b) it is a relatively well-defined organic medical condition; and (c) it can result in a range of possible diagnostic, therapeutic, and life style actions. A script for the case of CHD was developed from tape-recorded role-playing sessions with experienced, clinically active advisors. "Patients" in the CHD vignette presented with symptoms suggestive, although not pathophenomic, of CHD (including, for example, heartburn, indigestion unrelieved with antacids, new substernal discomfort, which is exertional and resolves after several minutes rest, pain in the back between the shoulder blades, stress, and elevated blood pressure).

Experimental Subjects (Physicians)

To be eligible for selection, an equal number of male and female physicians had to: (a) be internists or family practitioners; (b) have ≤12 years clinical experience (graduated between 1989 and 1996) or ≥22 years experience (graduated between 1965 and 1979) to get clear separation by level of experience; (c) be trained at an accredited medical school in the US; and (d) be currently providing clinical care at least half time. Eligible physicians were randomly sampled from throughout Massachusetts to fill 4 design cells (gender by level of experience). Screening telephone calls were conducted to identify eligible subjects and an hour-long, in-person interview was scheduled (at which time informed consent was obtained). Each physician subject was provided a modest stipend ($100) to partially offset lost revenue and to tangibly acknowledge participation. The response rate was 64.9%.

Assessing the Quality of Medical Care

Assessment of the quality of decision making requires some gold standard against which physician behavior can be compared. We originally planned to derive this standard from 2 main sources: (a) official clinical guidelines promulgated by, for example, AHRQ and AHA; and (b) the recommendations of a respected group of local clinical peers as to what any minimally competent provider should do when encountering the videotaped "patient." This approach was designed to accommodate the competing interests of different groups by developing a consensus view triangulated from these 2 sources on the most appropriate management of the presenting "patients" on the videotape. Table 1 depicts considerable divergence between the actions listed in the clinical guidelines

Table 1. The Concordance or Discordance Between 6 Boston Area Clinical Experts and Clinical Guidelines on Key Aspects of Care for CHD

Guidelines	Clinician Agreement
Information seeking	
Quality of pain	-
Duration of pain	-
Provoking factors	∅
Reliving factors	-
Patient medical history	√
Family history	-
Physical examination	
Heart	-
Lungs	-
Abdomen	∅
Peripheral extremities	-
Vascular neck exam	-
Test ordering	
ECG/EKG	√
Hemoglobin	-
Glucose	-
Lipids	-
Chest x-ray	-
Stress test	√
Drug treatments	
Aspirin	√
Beta blocker	∅
Short acting nitrates	√
Lifestyle recommendations	
Diet or weight	-
Exercise	-
Smoking	-

∅ =3 physicians in agreement with guidelines
√ =>3 physicians in agreement with guidelines
= <3 physicians in agreement with guidelines
Sources: Gibbons RJ, et al. [12] and Gibbons RJ, et al. [13]

and the recommendations made by clinically active peers. Furthermore, there was little consensus among the clinical peers as a group about the recommended action for the clinical case. This divergence may partly explain the fact that when asked whether their knowledge of guidelines contributed to their decisions with respect to the "patient" in the videotape, 75% of the physician subjects, said "no": There were no significant differences in the use of guidelines depending on physician gender (p=1.0) level of experience (p=.69), or their interaction (p=.69). For the purposes of this paper, clinical guidelines developed by ACC/AHA/ACP-ASIM (Guidelines for the Management of Patients with Chronic Stable Angina [12, 13] for CHD were used as 1 useful gold standard (recognizing there are others) against which physician decision making could be assessed. The clinical actions recommended in these guidelines were grouped into 5 categories, following the logical order of the encounter: information seeking, physical examination, test ordering, drug prescriptions, and life style recommendations. Whereas chest x-ray is not required for all patients with CHD, the guidelines strongly recommend it if there is evidence of congestive heart failure, valvular heart disease, pericardial disease, aortic dissection, or pulmonary disease.

Validity of the Experimental Approach

With every study there is some trade off between internal and external validity. The present experimental study has excellent internal validity, but its external validity can be questioned. Four precautionary steps were taken to enhance external validity, that is, whether the responses subjects gave represent the care they would truly provide. First, to achieve clinical authenticity of the scenario, physicians provided expertise during script development and were present during filming, where professional actors played the patient roles. Second, when physician subjects were asked how typical the "patient" viewed on the videotape was compared with patients in their everyday practice, 92% considered them very typical or reasonably typical. Third, the doctors viewed the tapes in the context of their practice day (not at a professional meeting, a course update, or in their home). In other words, it was likely they saw real patients before and after they viewed the "patient" scenario. Fourth, the doctors were specifically instructed to view the "patient" as 1 of their own patients and to respond as they would in their own practice.

RESULTS

We present primarily main effects results. There were a number of significant 2-way interactions, but there was no consistent pattern. The significant 2-way interactions are discussed below when they modify significant main effects.

Results from the experiment are divided into 3 main groups. First, we examined the proportion of physicians who would follow the clinical guidelines for the management of CHD. Second, we focused on the influence of 4 patient attributes (gender, age, race, SES) and their relation to guideline adherence. Third, we examined the influence of 2 provider characteristics (physician gender, age/years of clinical experience). Thus, our experimental approach permits unconfounded estimation of 5 different influences on physician behavior.

1. The Extent of Guideline Adherence
   Figure 1 summarizes the proportion of Massachusetts primary care physicians randomly sampled by design cell who would follow clinical guidelines when encountering the "patient" presenting with symptoms of CHD. A high percentage would obtain a medical history (83%), order an ECG/EKG (88%), examine the heart (74%), lungs (76%), and abdomen (78%), and order a stress test (75%). Compared to our guidelines standard, many physicians would obtain incomplete information on pain (duration and relieving factors), would not examine peripheral extremities and the neck, and would not order a glucose test or chest x-ray. Only a small proportion (6%) would recommend an increase in physical activity. Whereas there is high adherence to some guideline proscriptions (over 50% of primary providers would follow a third of them) there is low adherence to many of them (less than 20% of providers would follow a third of actions) recommended in the guidelines. There was high adherence (>50%) concerning the acquisition of information regarding provoking factors and patient history; the examination of the heart, lungs and abdomen; ordering EGG, hemoglobin, and stress tests; and making dietary recommendations. There was lower adherence (<20%) regarding the acquisition of information regarding the duration of pain, factors that relieve it, and family history; the examination of the extremities and neck; ordering glucose tests and x-rays; and making recommendations to increase exercise level.
2. Influence of Patient Attributes
   The influence of the "patient's" gender on CHD guideline adherence is depicted in Fig. 2. Female patients are significantly less likely to receive 4 of the 5 components of the physical examination: cardiac (heart; 84% of men vs 64% of women), pulmonary (lungs) (87% vs 64%), peripheral vascular (extremities) (18% vs 6%), and vascular neck exam (26% vs 9%). Whereas many of the patient gender differences in CHD guideline adherence do not reach statistical significance, there is general consistency in the results. The female "patients" received less than males on 2/3 of the specific clinical actions suggested in the guidelines (16 of the 23 listed).

   Figure 3 summarizes results concerning the effect of "patient" age (55 vs 75 years) on CHD guideline adherence. The only significant difference concerned the recommendation that the "patient" stop smoking-younger "patients" were twice as likely (40%) to receive such advice as older "patients" (23%, p<.04).

   No significant differences or consistent patterns were evident with respect to either the race (White vs Black) or socioeconomic status (lower vs higher) of the "patient" and physician adherence to guidelines.
3. The influence of Physician Characteristics
   Figure 4 summarizes the overall (main) effects of physician level of experience as to whether they follow guidelines when encountering a "patient" with CHD. The gender of the physician in the study did not appear related to his or her adherence to the CHD guidelines. A physician's level of clinical experience, however, did produce several significant findings. Table 2 shows that older/more experienced providers were more likely to inquire about factors provoking chest pain (for older patients), to order a glucose test (for male patients) and to prescribe short-acting

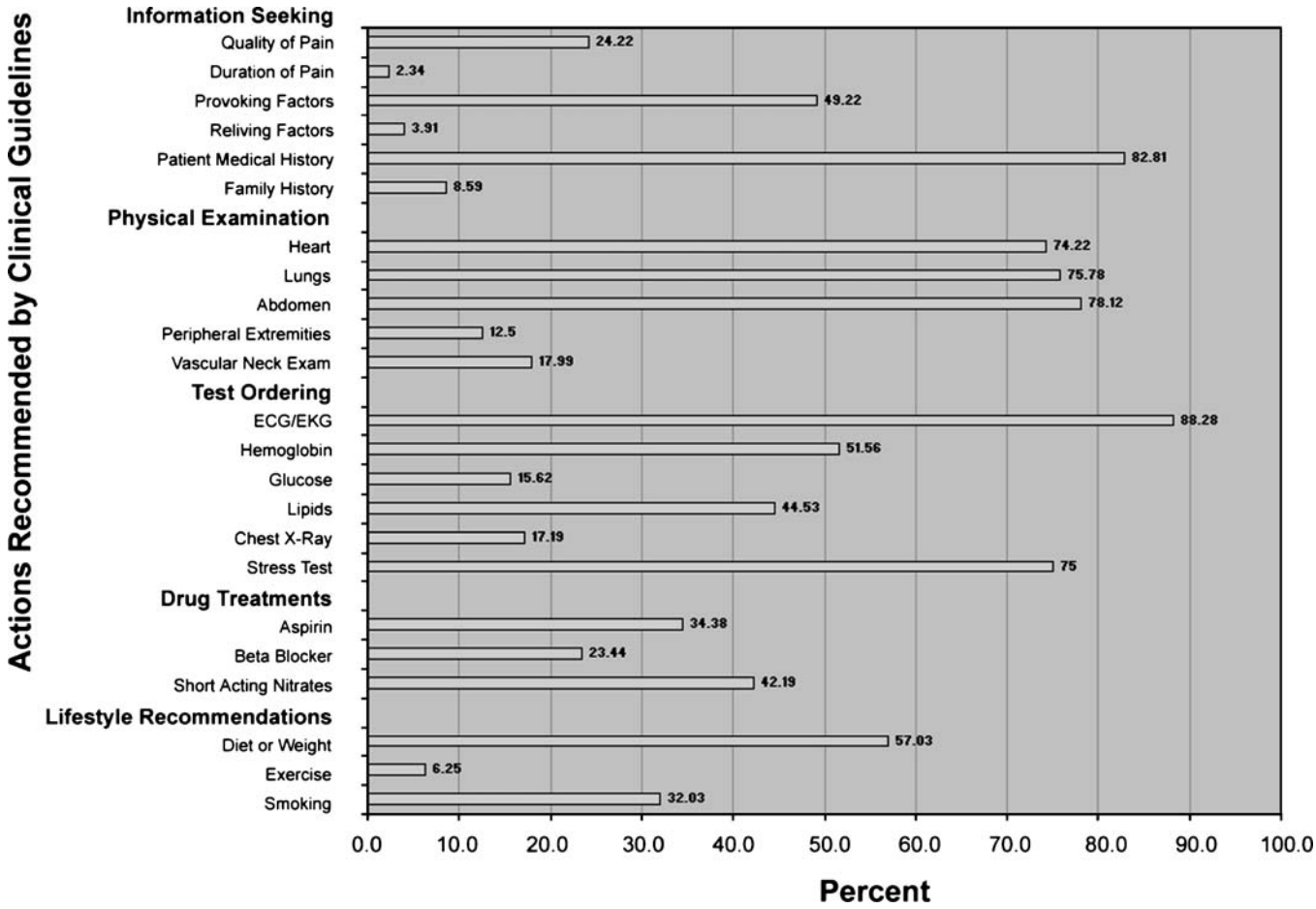


Figure 1. The percentage of primary care physicians who follow clinical guidelines when encountering a "patient" with coronary heart disease. *X* axis = percent of physicians who follow guidelines. *Y* axis = actions recommended by clinical guidelines.

nitrates (62.5% of more experienced physicians vs 21.9% of less experienced physicians, $p<.0001$). There was a consistent pattern (although results did not achieve statistical significance) of younger/less experienced doctors conducting all 5 components of the physical examination, whereas the older/more experienced physicians were more likely to order all 6 of the tests listed.

DISCUSSION

In this experiment, we provided physicians with a standardized clinical scenario to examine their clinical decision making, and compared this against 1 recognized standard of clinical guidelines for the evaluation of possible CHD. We found that whereas there was high physician adherence to some of the clinical guidelines for management of CHD, many physicians would not follow others—less than 50% would adhere to 2/3 of the specific recommendations. Moreover, characteristics of both the provider and the patient appear to play a role in adherence to guidelines. The gender of the "patient" appears to be influential: female "patients" received fewer of the actions recommended by clinical guidelines for the diagnostic evaluation of CHD. A "patient's" age significantly affected a physician's recommendation to quit smoking. An explanation for this may be found in the reaction of a physician colleague who stated, "I understand this result. I've got older patients and I've been going on for years about their smoking. They're never going to quit. I've really tried and I've given up." A physician's years of clinical experience were also associated with significant differences: less experienced doctors conducted more components of the physical examination, whereas more experienced physicians were more likely to recommend diagnostic testing.

The use of clinical guidelines as the gold standard for clinical care was chosen for several reasons: they are a) thought to reflect a consensus opinion based upon the current medical evidence (panels of experts from professional societies), and b) a useful standard because they include history of symptoms, clinical examination, diagnostic testing, and lifestyle and pharmacologic intervention recommendations. It is noteworthy that several parts of the clinical history and physical examination are inadequately addressed, as these are inexpensive and provide potentially critical information to elucidate the problem. The majority of physicians included the two diagnostic tests directly related to diagnosis of CHD, namely EKG and stress testing, whereas there was greater variability on tests for related or alternative diagnoses, such as glucose testing for undiagnosed diabetes, or chest x-ray for alternative diagnosis.

Considering those guidelines directly concerned with diagnosing and treating CHD and unstable or new angina, we find that providers generally follow the testing guidelines well, but are less adherent to those pertaining to historical information, and early treatments with aspirin, beta blockers or short-

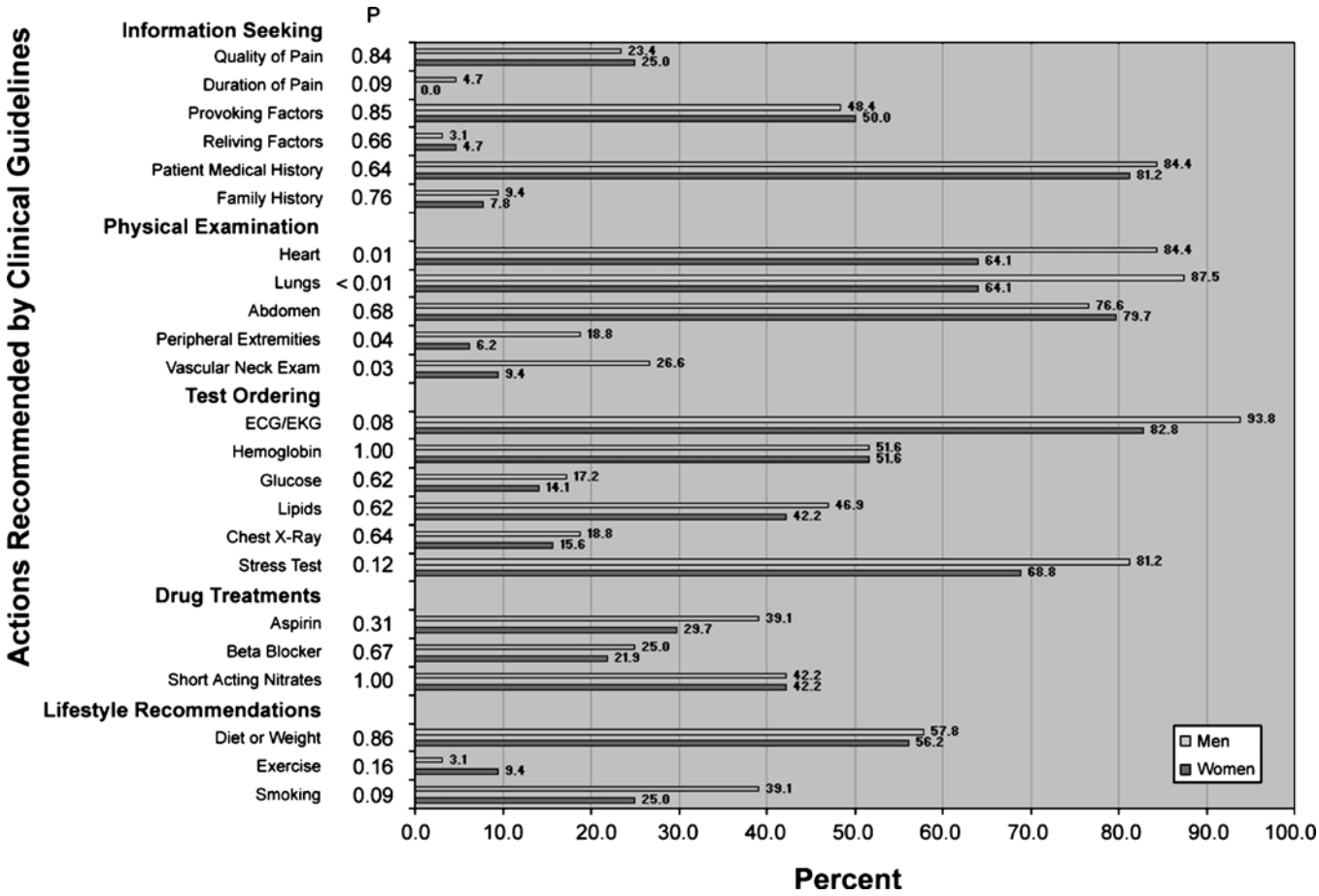


Figure 2. The percentage of primary care physicians who follow clinical guidelines when encountering a "patient" with coronary heart disease: patient gender. *X* axis = percent of physicians who follow guidelines. *Y* axis = actions recommended by clinical guidelines. Red bars = female. Blue bars = male.

activating nitrates. Likewise, the general differences in eliciting historic information suggests that these differences are important and could account for some of the gender disparities observed. Guidelines that relate to primary and secondary prevention, or assessment of important comorbidities may be considered less important at the first evaluation, and thus their lower use by physicians is perhaps understandable.

Some of the variability from the guidelines may reflect timing issues, where providers would postpone beta blocker therapy until after EKG and exercise stress testing, when the extent of disease and potential risk of unstable angina is understood. However, the prescription of short-acting nitrates or aspirin are recommended for their potential protection against cardiac events, even in this case of "heartburn" where upper gastrointestinal pathology may account for symptoms. The lack of exercise counseling could reflect caution until the diagnosis is established, but the same would not hold for the 2/3 of providers who would not recommend smoking cessation.

Our findings corroborate and extend the work of others, and indicate that even when there are evidence-based guidelines for the management of a condition as common as suspected CHD, some physicians' use of guideline activities continues to lag. This has been identified in other conditions as well. For example, regular monitoring of blood sugar is considered essential to the effective treatment of diabetes and prevent complications [14]. However, Saaddine and colleagues [15] found that only 29% of diabetic patients reported having their blood sugar tested during the previous year. Another study by McGlynn and colleagues [4] found that 24% of diabetic patients received 3 or more glycosylated hemoglobin tests over a 2-year period. Grant and colleagues [5] found fewer than half of all diabetics with elevated glycosylated hemoglobins had a change in medication, and only 10% of patients with elevated blood pressure readings had a change in management.

Our results show gender differences in the initial evaluation of possible CHD in women. Prior literature suggests reduced rates of revascularization, or delay in care for acute coronary events in women compared to men, as an explanation for higher morbidity and mortality for CHD in women [16–22]. However, our findings suggest that gender disparities in evaluation may begin even earlier in the clinical history and physical examination for CHD.

Our findings did not show a main or interaction effect by physician gender. That is, female physicians did not provide more guideline-based care to either male or female patients than did their male counterparts. Similarly, we found no differences based on patient race in recommendations for guideline appropriate CHD care.

The findings based on physician experience are mixed. Whereas more experienced providers ordered more diagnostic

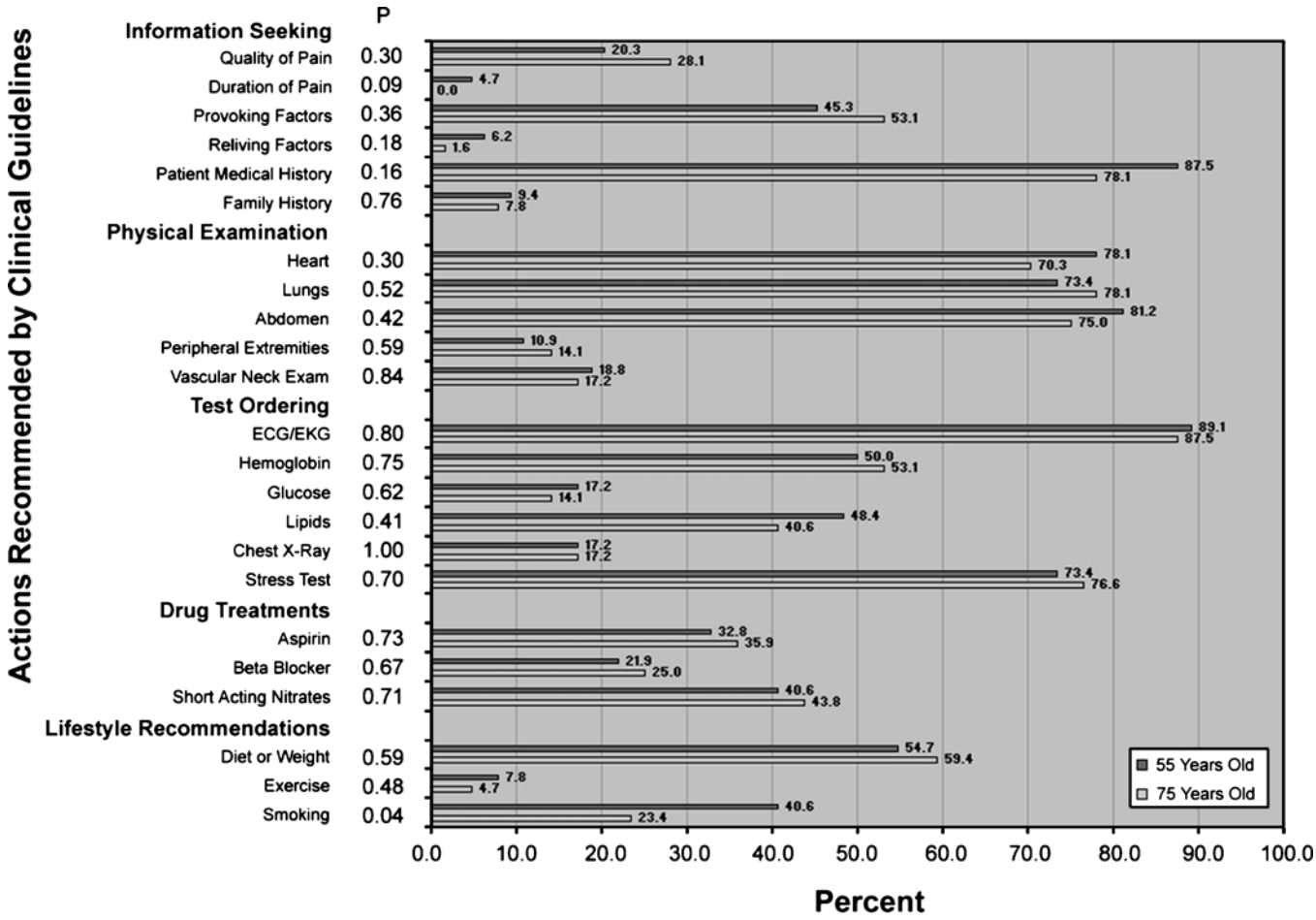


Figure 3. The percentage of primary care physicians who follow clinical guidelines when encountering a "patient" with coronary heart disease: patient age. *X* axis = percent of physicians who follow guidelines. *Y* axis = actions recommended by clinical guidelines. Red bars = 55 years old. Blue bars = 75 years old.

testing when the "patient" was older, their less experienced counterparts conducted more thorough physical examinations. The reduced behavioral counsel to older patients in the vignette may reflect a bias that older patients are *less* likely to adopt smoking cessation and other behavioral change. However, some evidence suggests that older patients are *as or more likely* to adopt change [23].

Limitations

Several limitations of this study should be noted. First, whereas the rigorous experimental design permits excellent internal validity, external validity remains a threat. Four precautionary steps were taken to hopefully minimize this threat (physicians were involved in script development; study subjects [physicians] were specifically asked how typical the "patient" was compared with patients in everyday practice; subjects viewed the tapes in the context of their practice day; subjects were specifically instructed to view the "patient" as 1 of their own patients). Second, the response rate of 64.9& (while high for a study of US physicians in the present climate) means over a third of those eligible and selected did not agree to participate. This is an unavoidable consequence of the decision to randomly sample in an attempt to increase the generalizability (external validity) of the research findings. Third, the level of adherence to guidelines may depend on which guidelines are selected as the gold standard. The guidelines used in this research were recommended by clinical colleagues as promulgated by a reputable professional organization and considered to have wide visibility among providers. They are clearly only 1 set among many different guidelines developed by numerous groups. Future research could investigate whether any variability in physician adherence is guideline specific.

Implications

The implications of this study rest on the assumption that physicians need to adopt and adhere to evidence-based guidelines in their everyday practice. Evidence-based disease management strategies, including early use of aspirin and short-acting nitrates, cannot reach their full potential if not incorporated into clinical care. The widely reported gender inequalities in coronary heart disease are unlikely to be reduced if improvements in the illness behavior of women (personal risk assessment, symptom recognition, and earlier help-seeking) are not matched by appropriate diagnoses, test ordering, and lifestyle recommendations by providers [24]. The

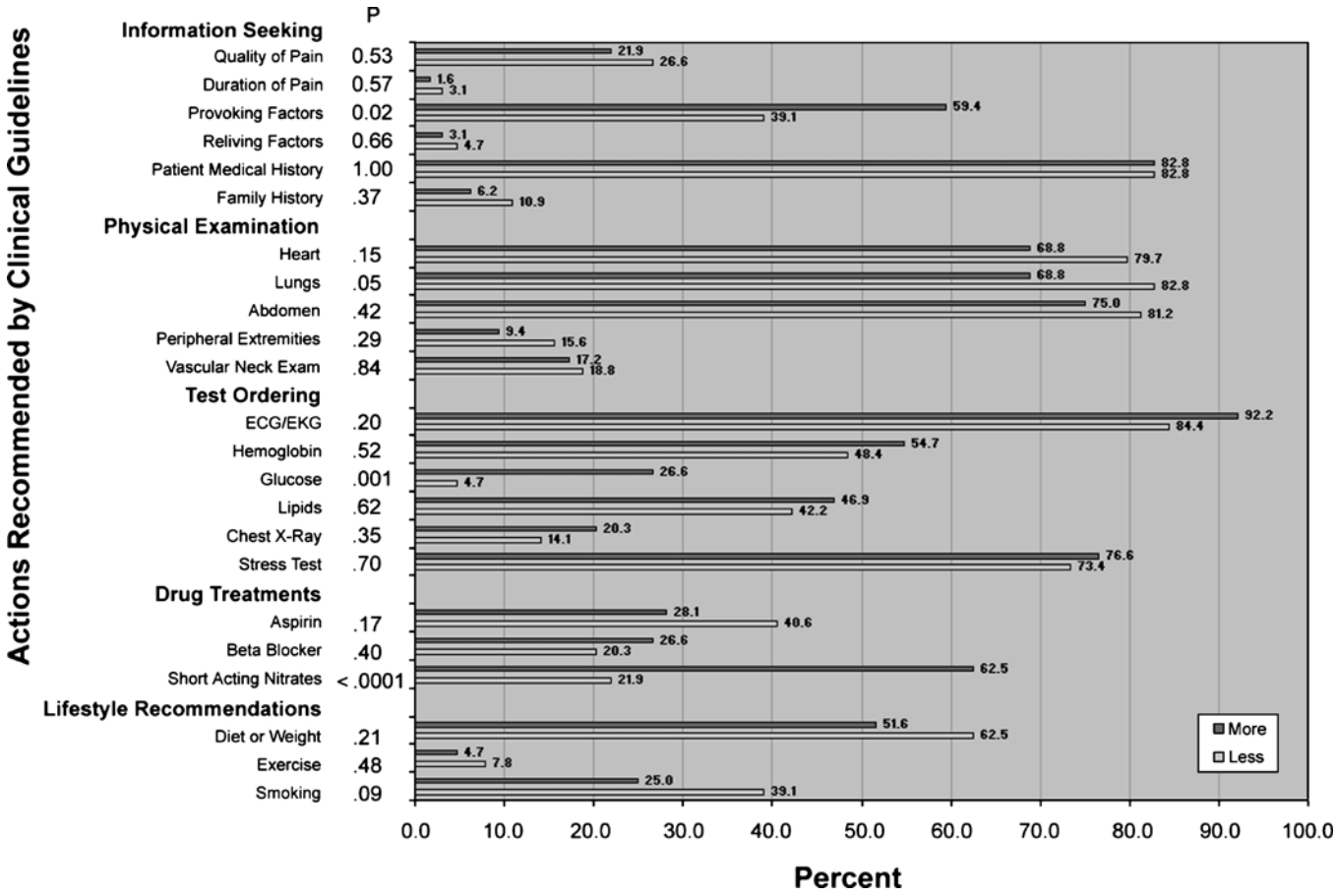


Figure 4. The percentage of primary care physicians who follow clinical guidelines when encountering a "patient" with coronary heart disease: physician level of experience. *X* axis = percent of physicians who follow guidelines. *Y* axis = actions recommended by clinical guidelines. Red bars = more experience. Blue bars = less experience.

question is how to ensure the incorporation of clinical guidelines into everyday practice: that is, how to institutionally support and reimburse physicians' adherence to their own, or their colleagues' own, recommendations (possibly pay-for-performance).

Whereas there is considerable literature addressing patient barriers to adherence to treatment recommendations, less is written on the barriers in clinical practice to physician adherence to recommendations. Some authors cite inadequate professional training [25], especially the limitations of training of primary care providers to address so many complex medical conditions. Others cite reimbursement policies that reward procedures over evaluation and management. Other barriers to adherence to guidelines include provider concerns about "cookbook medicine", perceived regulatory intrusion into practice and unwillingness to buy into the concept of management guidelines. Other barriers may be a lack of systems support, such as electronic medical records, which incorporate guideline management recommendations to ensure uniformity in care; or lack of patient participation in decision support [26]. Lastly, the number of guidelines propagated and the lack of concurrence among them make it difficult for providers to find a clear consensus on the best management practices [27].

Table 2. The Way in Which Physician Experience Interacts with Patient Age and Gender

		Level of Physician Experience (%)		*p* value
Variable	Patient Age (years)	Less	More	
Information Seeking	75	34.4	71.9	.0452
Provoking factors	55	43.8	46.9	
	Patient Gender			
Test Ordering	Male	0.0	34.4	.0497
Glucose	Female	9.4	18.8	

Acknowledgments: *This work was supported by National Institutes of Health, National Institute on Aging (Grant #AG16747). For further information on this work contact Dr. John B. McKinlay (jmckinlay@neriscience.com).*

Potential Financial Conflicts of Interest: *All authors agree that they have: • participated sufficiently in the work to take public responsibility for the content; • have made substantial contributions to the conception, design, or analysis and interpretation of the data and have approved the manuscript; • certify that the manuscript represents valid work and neither this nor a similar manuscript has been published or is being considered for publication elsewhere. All*

authors attest that they have no financial interest conflicting with complete and accurate reporting of the study findings.

Corresponding Author: *J. B. McKinlay, Institute for Community Health Studies,New England Research Institutes, 9 Galen Street, Watertown, MA 02472, USA (e-mail: jmckinlay@neriscience.com).*

REFERENCES

1. **Feldman HA, McKinlay JB, Potter DA, Freund KM, Burns RB, Moskowitz MA.** Non-medical influences on medical decision making: an experimental technique using videotapes, factorial design, and survey Sampling. Health Serv Res. 1997;32(3):343–65.
2. **McKinlay JB, Burns R, Feldman HA, et al.** Physician variably and uncertainty in the management of breast cancer. Med Care. 1998;36 (3):385–96.
3. **McKinlay JB, Lin T, Freund K, Moskowitz M.** The unexpected influence of physician attributes on clinical decisions: results of an experiment. J Health Soc Behav. 2002;43(1):92–106.
4. **McGlynn EA, Asch SM, Adams J, et al.** The quality of health care delivered to adults in the United States. N Engl J Med. 2003;348 (26):2635–45
5. **Grant RW, Buse JB, Meigs JB.** University HealthSystem Consortium (UHC) diabetes benchmarking project team. Diabetes Care. 2005; 28:337–442.
6. **Fisher RA.** Statistical Methods, Experimental Design and Scientific Inference. New York: Oxford University Press; 1990.
7. **Cochran WG, Cox CM.** Experimental Designs. New York: Wiley & Sons, Inc; 1957.
8. **McKinlay JB, Link CL, Arber S, Marceau L, O'Donnell AB, Adams A.** How do doctors in different countries manage the same patient? Results of a factorial experiment. Health Serv Res. 2006;41(6):2182–200.
9. PASS (Power Analysis and Sample Size). Kaysville, UT: NCSS; 2005.
10. **Rhoades DR, McFarland KF, Finch WH, Johnson AO.** Speaking and interruptions during primary care office visits. Fam Med. 2001;33 (7):528–32
11. **Cohen JW, Krauss NA.** Spending and service use among people with the fifteen most costly medical conditions. Health Affs. 2003;22(2):129–38.
12. **Gibbons RJ, Chatterjee K, Daley J, et al.** ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients With Chronic Stable Angina). J Am Coll Cardiol. 1999;33 (7):2092–197.
13. **Gibbons RJ, Chatterjee K, Daley J, et al.** ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: executive summary and recommendations. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients with Chronic Stable Angina). Circulation. 1999;99(21):2829–48.
14. **UKPDS Study Group.** Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. Lancet. 1998;352:837–53.
15. **Saaddine JB, Engelgau MM, Beckles GL, Gregg EW, Thompson TJ, Narayan KM.** A diabetes report card for the United States: quality of care in the 1990s. Ann Intern Med. 2002;136(8):565–74.
16. **Udvarhelyi IS, Gatsonis C, Epstein AM, Pashos CL, Newhouse JP, McNeil BJ.** Acute myocardial infarction in the Medicare population. JAMA. 1992;268:2530–6.
17. **Kostis JB, Wilson AC, O'Dowd K, et al.** Coronary heart disease/ myocardial infarction: sex differences in the management and long-term outcome of acute myocardial infarction: a statewide study. Circulation. 1994;90:1715–30.
18. **Weitzman S, Cooper L, Chambless L, et al.** Sex, racial, and geographic differences in the performance of cardiac diagnostic and therapeutic procedures for hospitalized acute myocardial infarction in four states. Am J Cardiol. 1997;79:722–6.
19. **Coronado BE, Griffith JL, Beshansky JR, Selker HP.** Hospital mortality in women and men with acute cardiac ischemia: a prospective multicenter study. JACC. 1997;29(7):1490–6.
20. **Maynard C, Every N, Martin JS, Kudenchuk PJ, Weaver WD.** Association of sex and survival in patients with acute myocardial infarction. Arch Intern Med. 1997;157(12):1379–84.
21. **Kober L, Torp-Pedersen C, Ottesen M, Rasmussen S, Lessing M, Skagen K.** Influence of sex on short- and long-term mortality after acute myocardial infarction. Am J Cardiol. 1996;77(12):1052–6.
22. **Wilkinson P, Laji K, Ranjadayalan K, Parsons L, Timmis A.** Acute myocardial infarction in women: survival analysis in first six months. BMJ. 1994;309(6954):566–9.
23. **Freund KM, D'Agostino RB, Belanger AJ, Kannel WB, Stokes J 3rd**. Predictors of smoking cessation: the Framingham Study. Am J Epidemiol. 1992;135(9):957–64.
24. **McKinlay J.** Some contributions from the social system to gender inequalities in heart disease. J Health Soc Behav. 1996;37:1–26.
25. **Hirsch IB.** The Burden of Diabetes Care. Diab Care. 2003;26:1613–4.
26. **Silliman, RA**. Crossing the quality chasm in breast cancer care. J Clin Onc. 2004;22(18):3654–6
27. **Broedl UC, Geiss HC, Parhofer KG**. Comparison of current guidelines for primary prevention of coronary heart disease: risk assessment and lipid-lowering therapy. J Gen Intern Med. 2003;8(3):190–5.

Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor

Emmanuel Munguia Tapia[1], Stephen S. Intille[1],
William Haskell[3], Kent Larson[1], Julie Wright[2], Abby King[3], and Robert Friedman[2]
[1]*House_n, Massachusetts Institute of Technology, Cambridge, MA, USA*
[2]*Boston Medical Center, Boston, MA USA;* [3] *Stanford Medical School, Palo Alto, CA, USA*
emunguia | intille @mit.edu

Abstract

In this paper, we present a real-time algorithm for automatic recognition of not only physical activities, but also, in some cases, their intensities, using five triaxial wireless accelerometers and a wireless heart rate monitor. The algorithm has been evaluated using datasets consisting of 30 physical gymnasium activities collected from a total of 21 people at two different labs. On these activities, we have obtained a recognition accuracy performance of 94.6% using subject-dependent training and 56.3% using subject-independent training. The addition of heart rate data improves subject-dependent recognition accuracy only by 1.2% and subject-independent recognition only by 2.1%. When recognizing activity type without differentiating intensity levels, we obtain a subject-independent performance of 80.6%. We discuss why heart rate data has such little discriminatory power.

1. Introduction

Automatic detection of physical activity (PA) might enable new types of health assessment and intervention tools that help people maintain their energy balance and stay physically fit and healthy. Recent research has shown that wearable accelerometers can be used to reliably detect some physical activity types when tested on small datasets (e.g.[1-4]). We are unaware; however, of work showing the same algorithm can detect not only activity type but also, in some cases, the same activity at different intensities. Furthermore, most work with accelerometers has either used cumbersome wired sensors [3] or sensors that store data locally for off-line processing [1, 4, 5]. Here we show how wireless sensors transmitting raw data in real-time (and thus susceptible to signal loss) could be used for automatic PA and PA-intensity recognition.

Past work in recognizing activities from accelerometer data has used computationally intensive supervised classification algorithms that typically require offline analysis [2, 3, 6]. In this work, we utilize fast Decision Tree (DT) classifiers (as in [1, 4, 5]) with a set of efficiently computable features to achieve real-time performance on current PCs. DTs can overfit to the data if sufficiently diverse training sets are not used. Therefore, we test on a relatively large dataset (compared with prior work) consisting of 30 gymnasium activities (see Table 1) collected from 21 participants by two different teams.

We also study the usefulness of heart rate (HR) data in discriminating the intensity of activities. HR may be useful since it correlates with energy expenditure for aerobic exercise; however, alone it provides little information about activity type, and it is influenced by other factors such as emotional states, ambient temperature, and fitness level. HR also responds and stabilizes slowly. In this work we explore the "best-case" scenario of how well activity type and, for some activities, intensity can be recognized when we place one accelerometer at each limb and the hip, and a HR monitor on the chest.

2. System overview and data collection

The PA recognition system consists of five triaxial wireless wearable accelerometers sensors sampling at 30Hz, a wireless HR monitor based on the Polar chest strap (Wearlink), and a laptop computer with a wireless receiver. These sensors allow data to be collected from multiple body points simultaneously without constraining movement [7] .

To acquire training data for the PA recognition system, a total of 21 participants between 18 and 65 years old and with varying levels of physical fitness were recruited at two separate medical labs: (1) The Boston Medical Center and (2) Stanford Medical School. Using cotton elastic sweat bands or non-restrictive adhesive bandages, researchers placed the accelerometers on each subject, with one at each of the following locations: top of the dominant wrist just behind the wrist joint, side of the dominant ankle just above the ankle joint, outside part of the dominant upper arm just below the shoulder joint, on the upper part of the dominant thigh, and on the dominant hip, as

indicated in Figure 1b. All the accelerometers were ±10G except the accelerometer on the hip, which was ±2G. The HR monitor was worn on the chest.

After the sensors were placed, each participant was asked to sit still and, after a stabilization period, resting HR was measured by measuring pulse for one minute. The participant's age-predicted maximum HR (MHR=220-age) was calculated. A combined 21 participants each performed 30 gymnasium activities for 2min each, with 12 and 9 datasets being collected from each site, respectively. The list of gymnasium activities, broken down by type and intensity differences, is shown in Table 1.

The activities with different intensity levels are *walking*, *cycling*, and *rowing*. For *walking*, we varied intensity by changing the treadmill speed (e.g. 2, 3, and 4 mph) and inclination (e.g. 4, 8, and 12 degrees). For *cycling*, we varied the cycle speed (e.g. 60, 80, and 100 rpm) and the cycle resistance level to settings that participants subjectively considered equivalent to light, moderate, and hard. Finally, for *rowing*, we kept the rowing speed constant at 30 spm and varied the resistance until reaching levels that participants considered light, moderate, and hard.

In total, 16.6 hours of usable annotated data were collected. In the remainder of this work, G1 refers to the gym dataset collected from site one and G2 refers to the dataset collected from site two. Dataset G2 differs slightly from dataset G1; due to lab constraints, data for the *move weight*, and *calisthenics (cali.)* activities were not collected, and the *rowing* activity was substituted by *arm ergometry*. Nevertheless, dataset G2 contains the same number of activities with different intensity levels as G1 (*walking*, *cycling*, and *arm ergometry*). Researchers interested in using these datasets should contact the authors.

3. Activity recognition algorithm

In the training step, data segments not labeled as one of the target activities listed in Table 1 are discarded. The 15 acceleration data streams (x, y, and z axes) were then broken into 50% overlapping sliding windows of length 4.2s and independently interpolated using cubic spline interpolation to fill out missing sensor values lost during wireless transmission. In Section 4, we describe how 4.2s was selected as the window size. If the percentage of samples lost inside a given window of length 4.2s was greater than 20% for any of the accelerometer axis streams, the window was discarded. To smooth out the noisy HR data, we applied a running average filter over the past 30s of data. The HR data is then segmented by accumulating the data over 30s windows from the end time of each acceleration window going backwards in time. HR windows are discarded when no samples are available for a given window. Overall, only 3.2% of the data (32.2m of 16.6h) collected were discarded due to accelerometer signal loss and 1.9% due to HR signal loss.

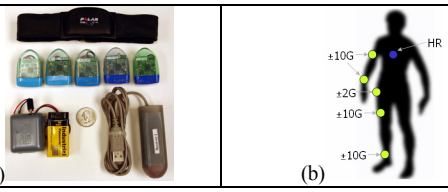


Figure 1. (a) Five 3-axis wireless accelerometers, a heart rate monitor, and USB wireless receiver, and (b) Placement of the sensors on the body.

Type	Intensity	Type	Intensity
Lying down	N/A	Cycling	Moderate at 80 rpm
Standing	N/A	Cycling	Light at 80 rpm
Sitting	N/A	Cycling	Light at 60 rpm
Sitting	Fidget feet legs	Cycling	Light at 100 rpm
Sitting	Fidget hands arms	Rowing	Light at 30 spm
Walking	2mph 0% grade	Rowing	Hard at 30 spm
Walking	3mph 0% grade	Rowing	Moderate 30 spm
Walking	3mph at 4% grade	Carry weight 2mph	N/A
Walking	3mph at 8% grade	Move weight high	N/A
Walking	3mph at 12% grade	Move weight low	N/A
Walking	4mph at 0% grade	Move weight side	N/A
Running	5mph at 0% grade	Cali. Bicep curls	N/A
Ascend stairs	N/A	Cali. Jumping jacks	N/A
Descend stairs	N/A	Cali. Push ups	N/A
Cycling	Hard at 80 rpm	Cali. Sit ups	N/A

Table 1. Activities studied in this work.

Time domain and frequency domain features are then computed for each 4.2s window. Here we use the area under curve (AUC) and variance to capture signal variability, mean distances between axes and mean to capture sensor orientation with respect to ground for postures, entropy to differentiate activity type, correlation coefficients to capture simultaneous motion of limbs, and FFT peaks and energy to discriminate between intensities. All the features are computed over each acceleration axis. The only feature computed over the HR data was the number of heart beats above the resting HR value (BPM-RHR). Finally, we used the WEKA toolkit [8] to evaluate the performance of the C4.5 DT [9] (pruned) and the NB classifier using one Gaussian distribution per feature per class.

4. Evaluation

To evaluate the performance of the recognition algorithm, we computed the true positive rate, false positive rate, precision, recall, and F-Measure over the segmented classes using subject-dependent and subject-independent training. In subject-dependent training, we performed 10-fold cross-validation over each subject's data and averaged the results over all the subjects. In subject-independent training, we trained the algorithms with the data of all the subjects but one and tested the performance on the left-out subject. We repeated this procedure for as many subjects as we had and averaged the results. To better

understand the performance of the algorithms, the results are clustered into three categories based on activity type: (1) postures (e.g. *lying down*, *standing*, and *sitting*), (2) activities with multiple intensities (*walking*, *rowing/arm ergometry*, and *cycling*), (3) and other activities (*running*, *calisthenics*, *move weight*, and *using stairs*).

4.1. Subject dependent analysis

We first determined the most appropriate window length to use (4.2s) by varying the window length from 0.5 to 17 seconds and measuring the performance of the C4.5 classifier over the datasets. A window of 4.2s is long enough to obtain a good accuracy (74-86%) while minimizing real-time classification delay. Using a similar strategy, we determined that using only two FFT peaks provided good performance. We then performed feature selection over subsets of all the features using the wrapper method and the C4.5 classifier. The most powerful features found, in decreasing order of importance, were the area under curve (93.1% accuracy using only this feature), mean distances between axes (92.1%), mean (91.3%), variance (88.7%), FFT peaks (86.1%), and correlation coefficients (74.8%). Consequently, we measured the performance of the C4.5 DT and the Naïve Bayes (NB) classifier over the best performing subset of these features that we call *variant* features: area under curve (15 values), mean distances between axes (15), mean (15), variance (15), FFT peaks (60), correlation coefficients (105), energy (15), and entropy (15) for a total feature vector with 255 values. Table 2 shows the performance measures using these features over both datasets (4.2s windows).

Table 2 shows that the performance is comparable using both classifiers. As a result, from this point on, we present results only for the C4.5 decision tree classifier. From Table 2, we can also observe that the performance is higher for dataset G2. We believe that this is because G2 contains 8 fewer activities than dataset G1 as explained in Section 2. Another important result is that we have achieved an average false positive rate of only 0.15% over both datasets.

A problem we encountered was that features with the highest discriminant power, such as AUC and the mean, are strongly sensitive to the acceleration signal magnitude and thus dependent on sensor orientation, and calibration. Consequently, we considered utilizing only features invariant to the signal magnitude. After evaluating the performance of subsets of these features using the C4.5 DT classifier over the datasets, we found that the best subset of features was: mean distances between axes, variance, energy, FFT peaks, and correlation coefficients, (225 values in total).

Dataset	Classifier	Activity			Total Accuracy (%)
		Postures (%)	Other (%)	Intensity (%)	
G1	C4.5	FP: 0.04 P: 98.7 R: 98.6 F: 98.7	FP: 0.20 P: 93.9 R: 93.8 F: 93.8	FP: 0.28 P: 92.2 R: 92.2 F: 92.2	93.7 ± 1.5
G2	C4.5	FP:0.04 P: 98.9 R: 99.1 F: 99.0	FP: 0.19 P: 96.3 R: 96.1 F: 96.2	FP: 0.19 P: 96.0 R: 96.0 F:96.0	96.0 ± 0.9
G1	NB	FP: 0.06 P: 98.1 R: 99.2 F: 98.7	FP: 0.22 P: 93.7 R: 93.6 F: 93.5	FP: 0.30 P: 92.1 R: 92.3 F: 92.1	93.3 ± 2.2
G2	NB	FP: 0.09 P: 97.7 R: 98.5 F: 98.1	FP: 0.05 P: 99.0 R: 95.3 F: 97.1	FP: 0.13 P: 97.2 R: 98.0 F:97.6	97.6 ± 0.8

Table 2. False positives (FP), precision (P), recall (R) and F-measure (F) for subject-dependent analysis using the *variant* features

Dataset	Features	Activity			Total Accuracy (%)
		Postures (%)	Other (%)	Intensity (%)	
G1	Invariant	FP: 0.06 P: 98.1 R: 98.8 F: 98.4	FP: 0.23 P: 93.5 R: 93.1 F: 93.1	FP: 0.32 P: 91.5 R: 92.0 F: 91.6	92.8 ± 2.5
G2	Invariant	FP: 0.09 P: 97.5 R: 98.4 F: 98.0	FP: 0.05 P: 99.1 R: 94.6 F: 96.8	FP: 0.16 P: 96.6 R: 97.57 F: 97.0	97.1 ± 1.2
G1	Invariant + HR	FP: 0.07 P: 98.0 R: 98.4 F: 98.2	FP: 0.19 P: 93.8 R: 93.7 F:93.8	FP: 0.20 P: 94.3 R: 94.2 F: 94.3	94.8 ± 1.7
G2	Invariant + HR	FP: 0.06 P: 98.3 R: 98.5 F:98.4	FP: 0.29 P: 94.8 R: 95.0 F: 94.9	FP: 0.14 P: 97.1 R: 96.9 F:97.0	96.9 ± 0.7

Table 3. False positives (FP), precision (P), recall (R) and F-measure (F) for subject-dependent analysis using the C4.5 DT and *invariant* features.

Table 3 presents the performance using these features we call invariant. Overall, the C4.5 classifier achieved an average accuracy of 94.9% on both datasets, an accuracy as good as the one obtained using the non-invariant features (94.9%).

After analyzing Table 3 and the confusion matrices, we observed that most of the errors were occurring when the classifier was trying to discriminate between the different *intensity* levels of the same activity. To further explore this, we created a new dataset that we call the *no-intensities* dataset where activities with different intensities, such as all the walking activities, were merged into one class. When we trained the C4.5 classifier using the *invariant* features over this new dataset, we found an improved performance of 97.3 ± 0.7 on G1 and 98.7 ± 0.4 on G2, or an average improvement of 3.3%.

The next step was to investigate if HR data could improve the discrimination among the intensity levels of an activity. To test this, we added the number of heart beats above resting HR (BPM-RHR) to the *invariant* features. Table 2 shows the result. The average performance over both datasets is 95.8%, an improvement of 1.2%.

In order to investigate why the HR feature has such a low impact on improving the discrimination between intensity levels, we trained the C4.5 classifier using only the HR. The recognition performance obtained was 34.0 ± 6.0 for G1 and 49.2 ± 6.7 for G2 using subject-dependent training. Overall, the results are higher for G2 because it contains fewer activities (8) than G1. After plotting misclassification histograms, we observed that the errors were concentrated at the beginning and end of activities. This is because HR lags physical activity and remains altered once the activity has finished (errors at the end of activity or beginning of the next one). Furthermore, for vigorous activities of short duration such as walking up stairs, HR increases constantly, resulting in classifications errors all across the activity.

4.2. Subject independent analysis

For the subject-independent analysis, we repeated the same procedure as the one followed in the previous section. Table 4 shows the results over the invariant features with and without incorporating HR.

The overall performance is relatively low, with an average accuracy of 56.3% (FP: 1.5%) using the C4.5 classifier on both datasets. When the HR feature is added, the average performance improves only 2.1%. To better understand why the HR feature has such a low impact on improving the discrimination between intensity levels, we trained the C4.5 classifier using only the HR feature. The recognition performance is as follows: 12.3 ± 1.7 for G1 and 14.4 ± 2.7 for G2 using subject-independent training. Consequently, we believe that HR does not improve discrimination in subject-independent training because subjects have different fitness levels and the number of beats above resting HR (BPM-RHR) is different for two subjects performing the same activity but with different levels of physical fitness. To minimize the effects of the physical fitness level of each individual, we repeated the experiment when HR (BMP) is normalized to lie between resting HR (RHR) and maximum HR (MHR) for each individual (MHR estimated as 220-age). Using this normalization, two individuals with different fitness level performing the same activity could have different BMP values, but relative to their MHR, they could be performing in the same intensity zone. Unfortunately, the results were similar to those obtained when not scaling the HR data. This may be because the MHR was estimated rather than measured.

After analyzing the confusion matrices we also observed that most of the errors were occurring when the classifier was trying to discriminate between the different intensity levels of an activity. Furthermore, when we train the C4.5 classifier (subject-independent) using the no-intensities dataset and the invariant features, we found an improved performance of 81.1 ± 11.9 for G1 and 80.1 ± 19.4 for G2. This means that (1) the classifier is indeed confusing between intensity levels and (2) that the subject-independent performance when no intensity levels are present is reasonable.

Dataset	Features	Activity			Total Accuracy (%)
		Postures (%)	Other (%)	Intensity (%)	
G1	Invariant	FP: 1.21 P: 66.2 R: 65.0 F: 63.9	FP: 1.02 P: 67.8 R: 68.9 F: 67.8	FP: 2.05 P: 46.4 R: 47.0 F: 46.1	55.6 ± 8.8
G2	Invariant	FP: 0.66 P: 83.7 R: 78.7 F: 80.8	FP: 1.97 P: 61.3 R: 55.5 F: 55.1	FP: 2.55 P: 48.8 R: 53.3 F: 50.3	57.0 ± 13.3
G1	Invariant + HR	FP: 1.15 P: 68.1 R: 67.5 F: 66.5	FP: 1.0 P: 68.8 R: 69.4 F: 68.7	FP: 1.91 P: 49.0 R: 49.0 F: 48.5	58.2 ± 11.0
G2	Invariant + HR	FP: 0.55 P: 86.1 R: 82.3 F: 83.8	FP: 1.99 P: 63.08 R: 46.5 F:51.9	FP: 2.47 P: 50.0 R: 57.6 F: 52.4	58.6 ± 14.9

Table 4. False positives (FP), precision (P), recall (R) and F-measure (F) for subject-independent analysis using the C4.5 DT and *invariant* features.

5. Acknowledgements

Funded by NIH R21 grant CA106745-02. Sensor development supported by NSF grant #0313065.

6. References

[1]. Bao, L. and S.S. Intille, *Activity Recognition from User-Annotated Acceleration Data*, in *Proc. Int'l Conf. Pervasive Comp*. 2004: p. 1-17.

[2]. Lester, J., et al., *A Hybrid Discriminative/Generative Approach for Modeling Human Activities*, in *Proc. Int'l Joint Conf. on Artificial Intelligence*. 2005: p. 766-772.

[3]. Lukowicz, P., et al., *Recognizing Workshop Activity Using Body Worn Microphone and Accelerometers*, in *Proc. Int'l Conf. Pervasive Comp*. 2004: p. 18-32.

[4]. Maurer, U., et al., *Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions*, in *Proc. Int'l Workshop on Wearable and Implantable Body Sensor Networks*. 2006, p. 113-116.

[5]. Parkka, J., et al., *Activity Classification Using Realistic Data From Wearable Sensors.* IEEE Transactions on inf. Tech. in Biomedicine, 2006. **10**(1): p. 119-128.

[6]. Subramanya, A., et al., *Hierarchical Models for Activity Recognition*, in *Proc. of the IEEE Conference in Multimedia Signal Processing.* 2007: p. 233-237.

[7]. E. Munguia Tapia, et al., *The Design of a Portable Kit of Wireless Sensors for Naturalistic Data Collection*, in *Proc. Int'l Conf. Pervasive Comp*. 2006: p. 117-134.

[8]. Witten, I.H. and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. 2005: Morgan Kaufmann.

[9]. Quinlan, J.R., *C4.5: Programs for machine learning*. 1993, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 302.

BMC Medical Genetics

BioMed Central

Research

A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study

Joanne M Murabito*[1,2], Carol L Rosenberg[2], Daniel Finger[2], Bernard E Kreger[1,2], Daniel Levy[1,3], Greta Lee Splansky[1], Karen Antman[2] and Shih-Jen Hwang[1,3]

Address: [1]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA, [2]Section of General Internal Medicine and the Sections of Hematology/Oncology, Department of Medicine, Boston University School of Medicine, Boston, MA, USA and [3]National Heart, Lung, and Blood Institute, Bethesda, MD, USA

Email: Joanne M Murabito* - murabito@bu.edu; Carol L Rosenberg - crosenbe@bu.edu; Daniel Finger - Daniel.finger@bmc.org; Bernard E Kreger - Bernard.kreger@bmc.org; Daniel Levy - levyd@nhlbi.nih.gov; Greta Lee Splansky - gretalee@bu.edu; Karen Antman - kha4@bu.edu; Shih-Jen Hwang - hwangs2@nhlbi.nih.gov

* Corresponding author

Published: 19 September 2007

BMC Medical Genetics 2007, **8**(Suppl 1):S6 doi:10.1186/1471-2350-8-S1-S6

This article is available from: http://www.biomedcentral.com/1471-2350/8/S1/S6

Abstract

Background: Breast and prostate cancer are two commonly diagnosed cancers in the United States. Prior work suggests that cancer causing genes and cancer susceptibility genes can be identified.

Methods: We conducted a genome-wide association study (Affymetrix 100K SNP GeneChip) of cancer in the community-based Framingham Heart Study. We report on 2 cancer traits – prostate cancer and breast cancer – in up to 1335 participants from 330 families (54% women, mean entry age 33 years). Multivariable-adjusted residuals, computed using Cox proportional hazards models, were tested for association with qualifying SNPs (70, 987 autosomal SNPs with genotypic call rate ≥80%, minor allele frequency ≥10%, Hardy-Weinberg test p ≥ 0.001) using generalized estimating equations (GEE) models and family based association tests (FBAT).

Results: There were 58 women with breast cancer and 59 men with prostate cancer. No SNP associations attained genome-wide significance. The top SNP associations in GEE models for each trait were as follows: breast cancer, rs2075555, $p = 8.0 \times 10^{-8}$ in *COL1A1*; and prostate cancer, rs9311171, $p = 1.75 \times 10^{-6}$ in *CTDSPL*. In analysis of selected candidate cancer susceptibility genes, two *MSR1* SNPs (rs9325782, GEE p = 0.008 and rs2410373, FBAT p = 0.021) were associated with prostate cancer and three *ERBB4* SNPs (rs905883 GEE p = 0.0002, rs7564590 GEE p = 0.003, rs7558615 GEE p = 0.0078) were associated with breast cancer. The previously reported risk SNP for prostate cancer, rs1447295, was not included on the 100K chip. Results of cancer phenotype-genotype associations for all autosomal SNPs are web posted at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007.

Conclusion: Although no association attained genome-wide significance, several interesting associations emerged for breast and prostate cancer. These findings can serve as a resource for replication in other populations to identify novel biologic pathways contributing to cancer susceptibility.

Background

Breast and prostate cancer are the most frequently diagnosed cancers in women and men respectively with over 200,000 cases each of new breast and prostate cancer estimated for 2006 in the United States [1]. Furthermore, prostate cancer is the second leading cause of cancer-related deaths in men and breast cancer is the second leading cause of cancer-related deaths in women. Family history is a well established risk factor for both breast and prostate cancer providing evidence for underlying genetic factors contributing to cancer occurrence. Accumulating research has identified a number of candidate genes and biologic pathways associated with increased susceptibility to cancer. However, even the most penetrant mutations, such as in *BRCA1* and *BRCA2*, account for only 5–10% of cases and are present in <1% of the general population. Genome-wide association studies (GWAS) provide a comprehensive approach to identification of genetic variants associated with cancer risk unconstrained by existing knowledge and may permit detection of common genetic variants each with small associated cancer risk but great public health impact. Reports from two recent GWAS demonstrated the importance of this approach with the discovery of novel loci for breast cancer susceptibility [2,3]. Four SNPs in the *FGFR2* gene were strongly associated with breast cancer and the association was confirmed in a sample of cases and controls derived from three additional studies [3].

We used the Framingham Heart Study (FHS) Affymetrix 100K SNP genotyping resource for GWAS of breast and prostate cancer phenotypes. The FHS offers the advantage of a prospective longitudinal family-based community sample with participants who have been well-characterized throughout adulthood with respect to risk factors and diseases, including cancer. We report results of two complementary strategies to identify genome-wide associations with cancer phenotypes: 1) a simple low p-value SNP ranking strategy; and 2) 100K SNP associations within candidate genes and regions previously reported to be associated with these cancers in humans.

Methods

Study sample

The genotyped study sample comprised 1345 Original cohort (n = 258) and Offspring (n = 1087) participants from the 330 largest FHS families. The Overview [4] provides further details of this sample. There were 250 participants in the sample with cancer (excluding non-melanoma skin cancer) including 58 women with breast cancer, and 59 men with prostate cancer. The Boston University Medical Center Institutional Review Board approved the examination content of Original Cohort and Offspring examinations. All participants provided written informed consent including consent for genetic studies.

Cancer phenotype definitions and residual creation

The 5209 Original Cohort participants have been examined biennially since study inception in 1948 and the 5124 Offspring Cohort participants (children of the Original Cohort and spouses of the children) have been examined approximately every 4 years since enrollment in 1971. Cancer cases were identified at routine examinations or by health-history updates for participants who did not attend an examination. Medical records were reviewed by two independent reviewers (BEK, GLS). The vast majority of cancers were confirmed by pathology reports; <3.4% of cancer cases were based on death certificate or clinical diagnosis alone. The 1976 World Health Organization ICD-O coding was used to classify all primary cancers. Hence, topography, location (subdivision of site), histology or morphology (cell histopathology), behavior (degree of malignancy), and grade (histological grading & differentiation) were recorded along with date of diagnosis. Cancer cases reviewed through December 31, 2005 were included in this study. The proportion of women and men in the study sample with breast (8%) and prostate cancer (9%) respectively was similar to that in the full FHS sample.

Cox proportional hazards models were used to generate martingale residuals using the PHREG procedure in SAS to perform the regression analysis of time from study entry to cancer diagnosis or last contact free of cancer. Breast cancer was examined in women only and models were cohort-specific and adjusted for 1) age at entry and 2) age, parity, and body mass index at study entry. For prostate cancer, in men only, models were cohort-specific and adjusted for age at entry.

Genotyping

Affymetrix 100K SNP GeneChip genotyping and the Marshfield STR genotyping performed by the Mammalian Genotyping Service http://research.marshfieldclinic.org/genetics are described in the Overview [4]. SNPs were excluded if minor allele frequency <0.10 (n = 38062); genotypic call rate <0.80 (n = 2346); Hardy Weinberg equilibrium test $p < 0.001$ (n = 1595). There were 70,987 autosomal SNPs available for analysis after the exclusions.

Statistical Analysis

The statistical methods for genome-wide association analyses are described in detail in the Overview [4]. While there are various suggested methods for interpretation of genome-wide significance, we chose to use a conservative ($p < 0.05/10^{-6} = 5 \times 10^{-8}$) threshold to define genome-wide significance for this report.

Association

All cancer residual traits listed in Table 1 were computed using Cox proportional hazards models. The full set of FHS participants with the phenotype were used to create the residuals. The residuals were used to test for association between the genotyped subset of participants and the SNPs using family-based association test (FBAT) and generalized estimating equation (GEE) models. FBAT analyses were restricted to at least 10 informative families. The GEE tests tended to give an excess of very small p-values over what would be expected (see Overview [4]).

SNP prioritization

We used several strategies to prioritize SNPs associated with cancer traits. First, we used an untargeted approach whereby SNP associations were ranked according to the strength of the p-value for each trait. Next we identified candidate genes reported to be associated with each cancer trait from review of the literature. Candidate genes were selected by searching PubMed (using susceptibility, gene, cancer and (breast or prostate) as keywords, last accessed 08-15-06), and the Entrez Gene and Online Mendelian Inheritance in Man resources, as well as recent text books. All available 100K SNPs in or near the a priori selected candidate genes were investigated for association with cancer traits. Finally for prostate cancer, we also examined SNP associations in the region on chromosome 8 (8q24) previously reported to be associated with prostate cancer in Icelandic families and confirmed in three case-control series [5] and African American men [6]. Further, for prostate cancer we examined the overlap in SNP associations in our study and the top 500 ranked SNPs from the Cancer Genetics Markers of Susceptibility (CGEMS) project sponsored by the National Cancer Institute. Because CGEMS used an Illumina platform for genotyping and the genotyping used in this study was performed with an Affymetrix platform, the gene_symbol from the UCSC annotation was used to link with CGEMS top 500 SNP list. Using this method, 1487 SNPs in FHS 100K (including SNPs with MAF < 0.1) correlated with the CGEMS top 500 SNPs related to a known gene.

SNPs were annotated with the UCSC genome browser tables using the May 2004 assembly http://genome.ucsc.edu/[7,8]. All genes within 60 kb of the top ranked SNPs were identified. The physical location of the SNPs was based on Build 35 of Genome for this report; however, the 100K web browser was based on Build 36.

Results

The cancer phenotypes available in the FHS 100K SNP resource, including details of the sample size, number of cancer events, and covariate adjustment for each trait are listed in Table 1. In this report, we consider only two phenotypes: breast cancer in women (multivariable-adjusted) and prostate cancer in men. Among participants in the 100K sample the mean age at breast cancer diagnosis was 59 years (range 35 to 83 years) in Offspring Cohort women and 70 years (range 35 to 97 years) in Original Cohort women; the mean age at prostate cancer diagnosis was 66 years (range 43 to 85 years) in Offspring Cohort men and 76 years (range 53 to 95 years) in Original Cohort men.

For each of the cancer phenotypes, Table 2 provides the top 15 SNPs ranked in order by lowest p-value for the GEE models and for the FBAT models (all SNP associations can be viewed on the web) [9]. None of the SNP associations achieved genome-wide significance ($p < 5 \times 10^{-8}$) [4]. However, for prostate cancer, the top SNP in GEE models, rs9311171, is in *CTDSPL* (CTD {carboxy-terminal domain, RNA polymerase II, polypeptide A} small phosphatase-like), a gene that may play a role in tumor suppression [10].

Table 1: Cancer Phenotypes for the Framingham Heart Study 100K Analyses

Phenotype*	Traits	Number of participants	Events	Adjustment (database phenotype variable name)
All Cancer* (excluding non-melanoma skin cancer)	2	1335	250	Cohort- and sex-specific 1. age at study entry (allcancer1) 2. multivariable adjusted for age, body mass index, and cigarette smoking at study entry (allcancer2)
Breast Cancer (women)	2	723	58	Cohort-specific 1. age at study entry (breastcancer1) 2. multivariable adjusted for age, parity, and body mass index at study entry (breastcancer2)
Prostate Cancer (men)	1	617	59	Cohort-specific 1. age at study entry (prostatecancer1)

Residuals from these models were used as traits to test for association with SNP genotypes. All residuals created using Cox Proportional Hazards regression. Cancers ascertained through December 31, 2005

*All SNP associations for all phenotypes in the table including all cancer are available on the web http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007

There were several additional associations not listed in Table 2 that were of interest. For prostate cancer, in GEE models rs906304 (rank 27, p = 0.000067), is in *NCOR2* also known as *SMRT*. SMRT levels have been reported to be elevated in prostate cancer cells, and result in suppression of anti-proliferative target gene actions for the vitamin D receptor [11,12]. In FBAT models, for prostate cancer SNP rs255561 (rank 17, p = 0.00039), is near *XRCC4*, a gene that plays a role in DNA repair and rs1897676 (rank 50, p = 0.0012), is in *PTPRD*. Protein tyrosine phosphatases are signaling molecules involved in the regulation of a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation [13]. For breast cancer in GEE models rs4146372 (rank 31, p = 0.00007), is near *RAB21*, SNP rs9307561 (rank 40, p = 0.0001), is near *FAT4*, and rs10512849 (rank 46, p = 0.00014), is in *FGF10*. These genes appear to play biologic roles in a variety of processes including tumor growth and suppression [13,14]. In FBAT models, for breast cancer rs2836391 (rank 46, p = 0.0012), is in *ERG*, an oncogene important in the development of prostate cancer [15,16].

Our second strategy was to identify from the literature candidate genes implicated in breast and prostate cancer susceptibility (see Additional data file 1). For prostate cancer, we identified 63 candidate genes. Twenty of these candidate genes had from 1 to 20 SNPs on the 100K chip whereas the remaining genes had no SNP coverage on the chip. For breast cancer, 75 potential candidate genes were identified, 28 of these genes had between 1 and 35 SNPs on the 100K chip and the remaining candidate genes were not covered on the chip. Two SNPs in *MSR1* (rs9325782, GEE p = 0.008 and rs2410373, FBAT p = 0.021) were associated with prostate cancer and three SNPs in *ERBB4* (rs905883 GEE p = 0.0002, rs7564590 GEE p = 0.003, rs7558615 GEE p = 0.0078) were associated with breast cancer (Table 3). For prostate cancer, a region on chromosome 8q24 was recently reported to be associated with prostate cancer risk in Icelandic men and confirmed in three case-control series of men of European ancestry and African American men [5,6]. There were a total of 64 SNPs on the 100K chip in this 8q24 region (128 to 129.3 Mb interval). However, the reported risk SNP, rs1447295, was not included on the 100K chip and none of the 64 available SNPs were in linkage disequilibrium with the risk SNP. Five other SNPs in this region were associated with prostate cancer with a GEE or FBAT p-value < 0.01 (Table 4).

The National Cancer Institute commenced the CGEMS [17] initiative to conduct genome-wide association studies to identify genetic factors related to prostate and breast cancer. We examined overlap between the top 500 ranked SNPs for prostate cancer in CGEMS phase 1a [18] and the results of the FHS 100K GWAS analysis for prostate cancer. The physical position of the SNP was used to detect overlapping associations and the results are shown in Table 5. Of note, many of the associations in Table 5 are in SNPs with very low minor allele frequencies and the results are presented according to minor allele frequency. *WWOX* gene, a tumor suppressor gene, that has been reported to play a role in prostate cancer [19], showed evidence of association (rs3751832, p = 0.0009) in our study sample.

Discussion

Breast and prostate cancer are the two most frequently diagnosed cancers in the United States and result in substantial morbidity and mortality [1]. A number of breast and prostate cancer susceptibility genes and chromosomal regions have been identified [2,3,5,20-34]. However, currently known genes account for only a fraction of the familial aggregation of breast cancer [25] and few prostate cancer susceptibility genes have even been identified. Risk for these cancers is likely mediated through variation in many genes, each conferring a relatively small risk for the disease. Genome-wide association studies provide an opportunity to discover novel genes and pathways that play a causal role in cancer occurrence and in turn may lead to new therapies for the prevention and treatment of cancer. Finding genetic associations with breast and prostate cancer risk that are robust across multiple studies may facilitate the identification of high risk individuals who can be targeted for early screening and preventive interventions.

We report GWAS results for breast cancer and prostate cancer phenotypes in a community-based sample of adults from two generations of the same families. Although none of the SNP associations achieved genome-wide significance in GEE or FBAT models, this resource has the potential to detect novel cancer susceptibility genes and to explore the relevance of promising candidate gene associations to human cancer. Our results can be compared to those from other genome-wide association studies such as the National Cancer Institute's CGEMS [17]. Although the two studies used different genotyping platforms limiting overlap in the SNPs examined, we were able to determine the physical position of the SNPs. Using this strategy, SNPs in the *ERRB4* gene (CGEMS DSSNP_ID rs2371438 and FHS 100K SNP rs10497958) were associated with prostate cancer. ErbB proteins are widely expressed in prostate cells [35] and may play a role in tumor development, growth and progression in human prostate cancer [36,37]. We also examined the 8q24 region previously associated with prostate cancer risk. CGEMS investigators recently reported a second independent risk SNP (rs6983267) within the 8q24 strongly associated with prostate cancer [34]. The Affymetrix 100K

Table 2: Cancer Phenotypes for FHS 100K Project: Results of Association Analyses*

2a. GEE results ranked by p-value

SNP	Chr	Physical Position	GEE p-value	FBAT p-value	Gene region (within 60 kb)
BREAST CANCER					
rs2075555	17	45,629,290	**8.3×10^{-8}**	0.157	*COL1A1*
rs6556756	5	163,821,858	**5.0×10^{-7}**	0.114	
rs1154865	12	72,276,104	**6.6×10^{-7}**	0.052	
rs1978503	18	51,815,280	**9.66×10^{-7}**	0.287	
rs1926657	13	94,672,957	**1.9×10^{-6}**	0.072	*ABCC4*
rs10263639	7	66,503,417	**2.7×10^{-6}**	0.590	
rs10490113	2	59,410,998	**4.7×10^{-6}**	0.077	
rs458685	21	30,099,382	**6.0×10^{-6}**	0.064	*GRIK1*
rs1876206	15	46,687,878	**6.0×10^{-6}**	0.742	*FBN1*
rs9314033	5	163,822,784	**1.2×10^{-5}**	0.058	
rs10486490	7	25,846,789	**1.3×10^{-5}**	0.245	
rs9325024	5	146,185,979	**1.7×10^{-5}**	0.112	*PPP2R2B*
rs1294255	1	229,808,093	**1.9×10^{-5}**	0.470	
rs10513754	3	178,681,862	**2.8×10^{-5}**	0.189	
rs10501093	11	27,946,695	**3.0×10^{-5}**	0.066	*KIF18A*
PROSTATE CANCER					
rs9311171	3	37,971,481	**1.8×10^{-6}**	0.392	*CTDSPL*
rs1529276	13	102,726,008	**1.8×10^{-6}**	0.071	
rs10498792	6	51,774,590	**3.0×10^{-6}**	0.285	*PKHD1*
rs4466137	5	83,021,495	**3.1×10^{-6}**	0.305	*HAPLN1*
rs345013	3	146,656,486	**5.1×10^{-6}**	0.002	
rs10505137	8	111,210,785	**1.3×10^{-5}**	0.002	
rs1873038	3	175,020,343	**1.8×10^{-5}**	0.317	*NLGN1*
rs8019932	14	29,676,177	**1.9×10^{-5}**	0.019	
rs344985	3	146,619,475	**2.1×10^{-5}**	0.002	
rs1920676	11	15,789,323	**2.2×10^{-5}**	0.196	
rs10511206	10	127,209,365	**2.4×10^{-5}**	0.028	
rs1037569	5	119,954,204	**2.8×10^{-5}**	0.165	
rs2109312	2	128,292,061	**3.0×10^{-5}**	0.786	*WDR33*
rs1352416	3	174,963,409	**3.4×10^{-5}**	0.201	*NLGN1*
rs1440714	11	15,,790,399	**3.7×10^{-5}**	0.223	

2b. FBAT results ranked by p-value

SNP	Chr	Physical Position	GEE p-value	FBAT p-value	Gene Region (within 60 kb)
BREAST CANCER					
rs7711990	5	180,307,504	0.010	**8.4×10^{-5}**	
rs1451125	4	137,449,995	0.008	**9.5×10^{-5}**	
rs4266352	4	109,823,951	0.001	**1.2×10^{-4}**	
rs6720918	2	122,873,276	0.027	**1.9×10^{-4}**	
rs6101183	20	59,025,355	0.132	**2.0×10^{-4}**	
rs9307064	4	90,503,827	0.159	**2.2×10^{-4}**	
rs10512287	9	101,538,523	0.807	**2.3×10^{-4}**	*GRIN3A*
rs7190881	16	7,368,424	0.475	**2.9×10^{-4}**	
rs2224402	1	229,359,076	0.418	**3.1×10^{-4}**	*C1orf57*
rs10516690	4	84,822,792	0.954	**3.7×10^{-4}**	
rs10487920	7	145,895,727	0.085	**3.9×10^{-4}**	*CNTNAP2*
rs6479347	9	91,348,121	0.017	**4.5×10^{-4}**	

Table 2: Cancer Phenotypes for FHS 100K Project: Results of Association Analyses* *(Continued)*

rs1924587	13	96,195,380	0.020	**4.6×10^{-4}**	*HS6ST3*
rs2059273	16	5,991,144	0.005	**4.9×10^{-4}**	*A2BP1*
rs2822669	21	14,752,449	0.029	**5.7×10^{-4}**	*SAMSN1*
PROSTATE CANCER					
rs657057	6	56,402,241	0.109	**3.4×10^{-5}**	*DST*
rs6691482	1	170,274,521	0.007	**4.0×10^{-5}**	*SLC9A11*
rs4378061	9	10,930,127	0.633	**6.6×10^{-5}**	
rs2056387	1	234,250,153	0.159	**1.1×10^{-4}**	*RYR2*
rs2274626	1	176,253,589	0.005	**1.4×10^{-4}**	*NPHS2*
rs181247	17	53,562,730	0.552	**3.2×10^{-4}**	*OR4D1|OR4D2*
rs7827348	8	3,020,222	0.054	**3.6×10^{-4}**	
rs255561	5	82,386,566	0.040	**3.9×10^{-4}**	*XRCC4*
rs1504858	4	137,063,555	0.010	**3.9×10^{-4}**	
rs954709	4	167,317,424	0.049	**4.2×10^{-4}**	
rs2179443	20	37,450,516	0.057	**4.4×10^{-4}**	
rs35239	5	68,064,834	0.657	**4.6×10^{-4}**	
rs477516	8	16,758,362	0.057	**5.1×10^{-4}**	
rs595725	1	28,993,331	0.239	**5.3×10^{-4}**	*OPRD1*
rs10515322	5	101,539,543	0.051	**5.4×10^{-4}**	*SLCO4C1*

*Autosomal SNPs with genotypic call rate ≥80%, minor allele frequency ≥10%, Hardy-Weinberg test p ≥ 0.001, and ≥10 informative families for FBAT. The physical location of the SNPs was based on Build 35 of Genome; however, the 100K web browser was based on Build 36

GeneChip did not include either of the previously reported risk SNPs; however, we did identify five other SNPs in this region associated with prostate cancer. The underlying biologic mechanism mediating prostate cancer risk associated with the SNPs and chromosomal region remains unknown. A two-stage approach, genome-wide association followed by selective genotyping of SNPs with suggestive evidence of association, may provide an efficient strategy for pursuing initial genome-wide results [2,38,39].

Several important limitations merit comment. First, this study used cancer cases identified through surveillance of a multigenerational community-based sample. The enrollment and examination of Original Cohort and Offspring Cohort participants began years before DNA collection occurred. Thus, a survival bias may have been introduced. Our cases may be comprised of early-staged and less lethal cancers. To address this potential bias, we adjusted for covariates using the full Framingham sample, and used the residual traits for the subset of individuals genotyped using the 100K Affymetrix GeneChip to test for association with the SNPs in linear regression models. Residual traits from Cox models typically are not ideally distributed for linear regression models, but our adjustment method using the full Framingham sample precludes the testing of SNP associations with cancer traits using Cox models. Second, we had a small number of cancer events (250 all cancer cases, 58 breast cancer cases and 59 prostate cancer cases) limiting our ability to detect SNP associations. In a recent small GWAS of age-related macular degeneration that included 96 cases and 50 controls, an association with the *CFH* gene was identified [40] and confirmed in larger studies [41-43]. However, in that report, individuals homozygous for the *CFH* risk allele had a sevenfold increased likelihood of age-related macular degeneration [40]. It is very unlikely that common genetic variants for cancer phenotypes will confer a risk for cancer susceptibility of that magnitude. For example, the odds ratio associated with the risk marker identified

Table 3: All SNP Associations within Selected Breast and Prostate Candidate Genes (up to 60 kb)

Associations within Candidate Genes Selected by Cancer Type: FBAT or GEE p-value < 0.01						
Cancer Trait	**Gene**	**SNP**	**Chr**	**Physical Position**	**GEE p-value**	**FBAT p-value**
Breast cancer	*ERBB4*	rs905883	2	213,013,976	2.0×10^{-4}	0.8517
Breast cancer	*ERBB4*	rs7564590	2	213,213,406	0.0032	0.2567
Breast cancer	*ERBB4*	rs7558615	2	212,771,655	0.008	0.8457
Prostate cancer	*MSR1*	rs9325782	8	16,134,844	8.2×10^{-4}	0.3448
Prostate cancer	*MSR1*	rs2410373	8	15,968,877	0.014	0.02051

Table 4: SNPs in the Chromosome 8q24 region Associated with Prostate Cancer:GEE or FBAT p-value < 0.01

SNP	Physical Position	Distance from reported marker	MAF	GEE p-value	FBAT p-value
rs7001069	128179828	-253268	0.0188	3.0×10^{-5}	0.39322
rs10505483	128194377	-238719	0.0223	4.0×10^{-5}	0.32697
rs1562871	128470954	37858	0.1696	0.16618	0.00328
rs10505474	128486686	53590	0.406	0.00995	0.0121
rs10505506	129114473	681377	0.3275	0.00829	0.92941

for prostate cancer in region 8q24 was 1.72 in the combined Icelandic sample [5]. Furthermore, the associations between prostate cancer and the SNPs with low minor allele frequency (Table 5) are likely to be false positive associations given the small number of prostate cancer cases in our sample. Third, the 100K Affymetrix GeneChip provides limited coverage of the genome; many of our a priori candidate genes did not have any SNP coverage on the chip and coverage of some candidate genes that were present on the chip was suboptimal. Importantly, the replicated risk SNP, rs1447295, for prostate cancer [5] was not included on the chip. NHLBI has committed funds for a 550 K genome-wide scan on all FHS participants. This will enable us to confirm our initial 100K SNP associations in a larger sample with a greater number of cancer cases and with denser coverage of the genome. We did not examine epistasis or gene-environment interactions which may modify the associations noted in this study. Lastly, most of our associations are likely to be due to chance. Replication studies are needed to determine if any of the results we report are indicative of true associations. It is important that our data be used in conjunction with data from other samples given the high probability of false positive associations.

Conclusion

In summary, the untargeted genome-wide approach to detect genetic associations for cancer traits provides an opportunity to identify novel biologic pathways related to cancer occurrence and to direct future study of candidate genes that hold the most promise for relevance to cancer risk in humans. Enhancing our understanding of the

Table 5: Prostate Cancer SNP Associations Common to Both CGEMS Top 500 Ranked SNPs and FHS 100K SNPs

		CGEMS Phase 1a				FHS 100K				
DBSNP_ID	Chr	Physical Position	CGEMS p-value	CGEMS Rank	100K SNP	Physical Position	GEE p-value	FBAT p-value	MAF	Nearest Gene
Minor Allele Frequency ≥10%										
rs3017183	18	51,319,241	0.00024100	85	rs3794889	51,212,935	0.017	0.799	0.224	*TCF4*
rs3017183	18	51,319,241	0.00024100	85	rs4801149	51,214,227	0.044	0.983	0.221	*TCF4*
rs2253319	21	35,109,916	0.00041100	137	rs2834645	35,109,556	0.009	0.293	0.206	*RUNX1*
rs392715	16	24,023,323	0.00041800	141	rs10492797	24,122,453	0.036	0.407	0.175	*PRKCB1*
rs10256504	7	3,516,796	0.00092200	295	rs4418248	3,194,181	1.8×10^{-4}	0.056	0.109	*SDK1*
rs6102912	20	40,636,349	0.00111000	350	rs986831	40,594,892	0.009	0.094	0.407	*PTPRT*
rs4782742	16	81,573,151	0.00127400	395	rs254315	82,313,913	7.1×10^{-4}	0.011	0.141	*CDH13*
rs2371438	2	212,791,037	0.00128500	401	rs10497958	212,771,465	0.055	0.047	0.164	*ERBB4*
rs6577648	8	135,555,688	0.00144900	445	rs10505624	135,607,862	4.9×10^{-4}	0.007	0.147	*ZFAT1*
rs9327886	5	102,931,646	0.00154600	468	rs10515347	102,950,833	0.006	0.062	0.354	*NUDT12*
rs6555491	5	7,799,792	0.00158700	478	rs10512920	7,455,561	0.009	0.647	0.244	*ADCY2*
Minor Allele Frequency <10%										
rs38276	7	14,078,240	0.00055700	181	rs10486031	14,258,703	1.7×10^{-5}	0.759	0.032	*DGKB*
rs216666	2	80,745,015	0.00057000	184	rs10520247	79,693,452	2.8×10^{-5}	0.240	0.065	*CTNNA2*
rs216666	2	80,745,015	0.00057000	184	rs10520246	79,693,324	1.5×10^{-4}	0.198	0.050	*CTNNA2*
rs1261256	5	22,490,853	0.00060200	193	rs10520880	22,634,003	8.5×10^{-5}	0.356	0.065	*CDH12*
rs7610584	3	171,635,885	0.00061900	196	rs10513681	171,624,010	7.8×10^{-7}	0.402	0.025	*CLDN11*
rs7329659	13	97,872,564	0.00064700	209	rs7989050	97,680,299	6.0×10^{-6}	0.258	0.014	*FARP1*
rs208354	7	2,554,104	0.00065000	211	rs10487577	2,583,719	2.0×10^{-7}	N/A	0.012	*GNA12*
rs3751832	16	77,771,773	0.00079900	257	rs10514443	77,353,386	9.3×10^{-4}	N/A	0.008	*WWOX*

CGEMS = Cancer Genetics Markers of Susceptibility; MAF = minor allele frequency; N/A= not available, less than 10 informative families for FBAT

mechanisms responsible for cancer susceptibility may in turn identify novel strategies for early detection, prevention, and treatment of breast and prostate cancers. These data serve as a resource for replication in other population-based samples.

Abbreviations

CGEMS = cancer genetics markers of susceptibility; FBAT = family-based association test; FHS = Framingham Heart Study; GEE = generalized estimating equations; GWAS = genome-wide association study; ICD-O = international classification of diseases for oncology; MAF = minor allele frequency; SNP = single nucleotide polymorphism.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors have made substantial contributions to conception and design or acquisition of phenotypic data. JMM, SJW, DF, DL, CLR contributed to the analysis and interpretation of data. JMM, CLR, DL, DF, and SJW have been involved in drafting the manuscript or revising it critically for important intellectual content and all authors have read and approved the final manuscript.

Additional material

Additional file 1

List of selected candidate genes for breast and prostate cancer.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2350-8-S1-S6-S1.doc]

Acknowledgements

We thank Drs Emelia J. Benjamin and Martin G. Larson for their participation in the Framingham Heart Study Cancer Phenotype Working Group and for their contribution to the conception, design and interpretation of the data and comments on the drafting of the manuscript.

We thank the FHS participants for their ongoing participation and dedication to the study making this work possible.

NHLBI's Framingham Heart Study is supported by contract number N01-HC-25195. FHS 100K analyses were conducted using the Boston University. Linux Cluster for Genetic Analysis (LinGA) funded by the NIH NCRR (National Center for Research Resources) Shared Instrumentation grant 1S10RR163736-01A1 http://www.bu.edu/dbin/sph/departments/biostatistics/linga_publications.php.

This article has been published as part of *BMC Medical Genetics* Volume 8 Supplement 1, 2007: The Framingham Heart Study 100,000 single nucleotide polymorphisms resource. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2350/8?issue=S1.

References

1. American Cancer Society: **Cancer Facts and Figures 2006.** *Atlanta: American Cancer Society*; 2006.
2. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Luccarini C, Conroy D, Shah M, Munday H, Jordan C, Perkins B, West J, Redman K, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo KY, Noh DY, Ahn SH, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low YL, Bogdanova N, Schurmann P, Dork T, Tollenaar RA, Jacobi CE, Devilee P, Klijn JG, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, Macpherson G, Reed MW, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den OA, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko YD, Spurdle AB, Beesley J, Chen X, Aghmesheh M, Amor D, Andrews L, Antill Y, Armes J, Armitage S, Arnold L, Balleine R, Begley G, Beilby J, Bennett I, Bennett B, Berry G, Blackburn A, Brennan M, Brown M, Buckley M, Burke J, Butow P, Byron K, Callen D, Campbell I, Chenevix-Trench G, Clarke C, Colley A, Cotton D, Cui J, Culling B, Cummings M, Dawson SJ, Dixon J, Dobrovic A, Dudding T, Edkins T, Eisenbruch M, Farshid G, Fawcett S, Field M, Firgaira F, Fleming J, Forbes J, Friedlander M, Gaff C, Gardner M, Gattas M, George P, Giles G, Gill G, Goldblatt J, Greening S, Grist S, Haan E, Harris M, Hart S, Hayward N, Hopper J, Humphrey E, Jenkins M, Jones A, Kefford R, Kirk J, Kollias J, Kovalenko S, Lakhani S, Leary J, Lim J, Lindeman G, Lipton L, Lobb L, Maclurcan M, Mann G, Marsh D, McCredie M, McKay M, Anne MS, Meiser B, Milne R, Mitchell G, Newman B, O'loughlin I, Osborne R, Peters L, Phillips K, Price M, Reeve J, Reeve T, Richards R, Rinehart G, Robinson B, Rudzki B, Salisbury E, Sambrook J, Saunders C, Scott C, Scott E, Scott R, Seshadri R, Shelling A, Southey M, Spurdle A, Suthers G, Taylor D, Tennant C, Thorne H, Townshend S, Tucker K, Tyler J, Venter D, Visvader J, Walpole I, Ward R, Waring P, Warner B, Warren G, Watson E, Williams R, Wilson J, Winship I, Young MA, Bowtell D, Green A, Defazio A, Chenevix-Trench G, Gertig D, Webb P, Mannermaa A, Kosma VM, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BA: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007 in press.
3. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover RN, Thomas G, Chanock SJ: **A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer.** *Nat Genet* 2007 in press.
4. Cupples LA, Arruda H, Benjamin EJ, D'Agostino RB Sr, Demissie S, DeStefano AL, Dupuis J, Falls K, Fox CS, Gottlieb D, Govindaraju DR, Guo CY, Heard-Costa N, Hwang SJ, Katherisan S, Kiel D, Laramie JM, Larson MG, Levy D, Liu CY, Lunetta KL, Mailman M, Manning AK, Meigs JB, Murabito JM, Newton-Cheh C, O'Connor GT, O'Donnell CJ, Pandey MA, Seshadri S, Vasan RS, Wang ZY, Wilk JB, Wolf PA, Yang Q, Atwood LD: **The Framingham Heart Study 100K SNP genome-wide association study resource: Overview of 17 phenotype working group reports.** *BMC Med Genet* 2007, **8(Suppl 1):**S1.
5. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birgisdottir B, Le Roux L, Olafsdottir A, Blondal T, Andresdottir M, Gretarsdottir OS, Bergthorsson JT, Gudbjartsson D, Gylfason A, Thorleifsson G, Manolescu A, Kristjansson K, Geirsson G, Isaksson H, Douglas J, Johansson JE, Balter K, Wiklund F, Montie JE, Yu X, Suarez BK, Ober C, Cooney KA, Gronberg H, Catalona WJ, Einarsson GV, Barkardottir RB, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K: **A common variant associated with prostate cancer in European and African populations.** *Nat Genet* 2006, **38:**652-658.
6. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM, Oakley-Girvan I, Whittemore AS, Cooney KA, Ingles SA, Altshuler D, Hender-

son BE, Reich D: **Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men.** *Proc Natl Acad Sci USA* 2006, **103:**14068-14073.
7. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ: **The UCSC Genome Browser Database.** *Nucleic Acids Res* 2003, **31:**51-54.
8. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12:**996-1006.
9. 2007 [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000007].
10. Kashuba VI, Li J, Wang F, Senchenko VN, Protopopov A, Malyukova A, Kutsenko AS, Kadyrova E, Zabarovska VI, Muravenko OV, Zelenin AV, Kisselev LL, Kuzmin I, Minna JD, Winberg G, Ernberg I, Braga E, Lerman MI, Klein G, Zabarovsky ER: **RBSP3 (HYA22) is a tumor suppressor gene implicated in major epithelial malignancies.** *Proc Natl Acad Sci USA* 2004, **101:**4906-4911.
11. Khanim FL, Gommersall LM, Wood VH, Smith KL, Montalvo L, O'Neill LP, Xu Y, Peehl DM, Stewart PM, Turner BM, Campbell MJ: **Altered SMRT levels disrupt vitamin D3 receptor signalling in prostate cancer cells.** *Oncogene* 2004, **23:**6712-6725.
12. Abedin SA, Banwell CM, Colston KW, Carlberg C, Campbell MJ: **Epigenetic corruption of VDR signalling in malignancy.** *Anticancer Res* 2006, **26:**2557-2566.
13. 2006 [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd].
14. Theodorou V, Boer M, Weigelt B, Jonkers J, van d V, Hilkens J: **Fgf10 is an oncogene activated by MMTV insertional mutagenesis in mouse mammary tumors and overexpressed in a subset of human breast carcinomas.** *Oncogene* 2004, **23:**6047-6055.
15. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM: **Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer.** *Science* 2005, **310:**644-648.
16. Liu W, Chang B, Sauvageot J, Dimitrov L, Gielzak M, Li T, Yan G, Sun J, Sun J, Adams TS, Turner AR, Kim JW, Meyers DA, Zheng SL, Isaacs WB, Xu J: **Comprehensive assessment of DNA copy number alterations in human prostate cancers using Affymetrix 100K SNP mapping array.** *Genes Chromosomes Cancer* 2006, **45:**1018-1032.
17. 2006 [http://cgems.cancer.gov/about/executive_summary.asp].
18. 2006 [https://caintegrator.nci.nih.gov/cgems/].
19. Qin HR, Iliopoulos D, Semba S, Fabbri M, Druck T, Volinia S, Croce CM, Morrison CD, Klein RD, Huebner K: **A role for the WWOX gene in prostate cancer.** *Cancer Res* 2006, **66:**6477-6481.
20. Xu J, Dimitrov L, Chang BL, Adams TS, Turner AR, Meyers DA, Eeles RA, Easton DF, Foulkes WD, Simard J, Giles GG, Hopper JL, Mahle L, Moller P, Bishop T, Evans C, Edwards S, Meitz J, Bullock S, Hope Q, Hsieh CL, Halpern J, Balise RN, Oakley-Girvan I, Whittemore AS, Ewing CM, Gielzak M, Isaacs SD, Walsh PC, Wiley KE, Isaacs WB, Thibodeau SN, McDonnell SK, Cunningham JM, Zarfas KE, Hebbring S, Schaid DJ, Friedrichsen DM, Deutsch K, Kolb S, Badzioch M, Jarvik GP, Janer M, Hood L, Ostrander EA, Stanford JL, Lange EM, Beebe-Dimmer JL, Mohai CE, Cooney KA, Ikonen T, Baffoe-Bonnie A, Fredriksson H, Matikainen MP, Tammela TL, Bailey-Wilson J, Schleutker J, Maier C, Herkommer K, Hoegel JJ, Vogel W, Paiss T, Wiklund F, Emanuelsson M, Stenman E, Jonsson BA, Gronberg H, Camp NJ, Farnham J, Cannon-Albright LA, Seminara D: **A combined genome-wide linkage scan of 1,233 families for prostate cancer-susceptibility genes conducted by the international consortium for prostate cancer genetics.** *Am J Hum Genet* 2005, **77:**219-229.
21. Kammerer S, Roth RB, Reneland R, Marnellos G, Hoyal CR, Markward NJ, Ebner F, Kiechle M, Schwarz-Boeger U, Griffiths LR, Ulbrich C, Chrobok K, Forster G, Praetorius GM, Meyer P, Rehbock J, Cantor CR, Nelson MR, Braun A: **Large-scale association study identifies ICAM gene region as breast and prostate cancer susceptibility locus.** *Cancer Res* 2004, **64:**8906-8910.
22. Xu J, Zheng SL, Hawkins GA, Faith DA, Kelly B, Isaacs SD, Wiley KE, Chang B, Ewing CM, Bujnovszky P, Carpten JD, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB: **Linkage and association studies of prostate cancer susceptibility: evidence for linkage at 8p22–23.** *Am J Hum Genet* 2001, **69:**341-350.
23. Kraft P, Pharoah P, Chanock SJ, Albanes D, Kolonel LN, Hayes RB, Altshuler D, Andriole G, Berg C, Boeing H, Burtt NP, Bueno-de-Mesquita B, Calle EE, Cann H, Canzian F, Chen YC, Crawford DE, Dunning AM, Feigelson HS, Freedman ML, Gaziano JM, Giovannucci E, Gonzalez CA, Haiman CA, Hallmans G, Henderson BE, Hirschhorn JN, Hunter DJ, Kaaks R, Key T, Le Marchand L, Ma J, Overvad K, Palli D, Pike MC, Riboli E, Rodriguez C, Setiawan WV, Stampfer MJ, Stram DO, Thomas G, Thun MJ, Travis R, Trichopoulou A, Virtamo J, Wacholder S: **Genetic variation in the HSD17B1 gene and risk of prostate cancer.** *PLoS Genet* 2005, **1:**e68.
24. Breast Cancer Association Consortium: **Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium.** *J Natl Cancer Inst* 2006, **98:**1382-1396.
25. Antoniou AC, Easton DF: **Models of genetic susceptibility to breast cancer.** *Oncogene* 2006, **25:**5898-5905.
26. Smith P, McGuffog L, Easton DF, Mann GJ, Pupo GM, Newman B, Chenevix-Trench G, Szabo C, Southey M, Renard H, Odefrey F, Lynch H, Stoppa-Lyonnet D, Couch F, Hopper JL, Giles GG, McCredie MR, Buys S, Andrulis I, Senie R, Goldgar DE, Oldenburg R, Kroeze-Jansema K, Kraan J, Meijers-Heijboer H, Klijn JG, van Asperen C, van L I, Vasen HF, Cornelisse CJ, Devilee P, Baskcomb L, Seal S, Barfoot R, Mangion J, Hall A, Edkins S, Rapley E, Wooster R, Chang-Claude J, Eccles D, Evans DG, Futreal PA, Nathanson KL, Weber BL, Rahman N, Stratton MR: **A genome wide linkage search for breast cancer susceptibility genes.** *Genes Chromosomes Cancer* 2006, **45:**646-655.
27. Walsh T, Casadei S, Coats KH, Swisher E, Stray SM, Higgins J, Roach KC, Mandell J, Lee MK, Ciernikova S, Foretova L, Soucek P, King MC: **Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer.** *JAMA* 2006, **295:**1379-1388.
28. Berndt SI, Chatterjee N, Huang WY, Chanock SJ, Welch R, Crawford ED, Hayes RB: **Variant in sex hormone-binding globulin gene and the risk of prostate cancer.** *Cancer Epidemiol Biomarkers Prev* 2007, **16:**165-168.
29. Bergman A, Karlsson P, Berggren J, Martinsson T, Bjorck K, Nilsson S, Wahlstrom J, Wallgren A, Nordling M: **Genome-wide linkage scan for breast cancer susceptibility loci in Swedish hereditary non-BRCA1/2 families: suggestive linkage to 10q23.32–q25.3.** *Genes Chromosomes Cancer* 2007, **46:**302-309.
30. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, Jayatilake H, McGuffog L, Hanks S, Evans DG, Eccles D, Easton DF, Stratton MR: **PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene.** *Nat Genet* 2007, **39:**165-167.
31. Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, Scollen S, Baynes C, Ponder BA, Chanock S, Lissowska J, Brinton L, Peplonska B, Southey MC, Hopper JL, McCredie MR, Giles GG, Fletcher O, Johnson N, Dos SSI, Gibson L, Bojesen SE, Nordestgaard BG, Axelsson CK, Torres D, Hamann U, Justenhoven C, Brauch H, Chang-Claude J, Kropp S, Risch A, Wang-Gohrke S, Schurmann P, Bogdanova N, Dork T, Fagerholm R, Aaltonen K, Blomqvist C, Nevanlinna H, Seal S, Renwick A, Stratton MR, Rahman N, Sangrajrang S, Hughes D, Odefrey F, Brennan P, Spurdle AB, Chenevix-Trench G, Beesley J, Mannermaa A, Hartikainen J, Kataja V, Kosma VM, Couch FJ, Olson JE, Goode EL, Broeks A, Schmidt MK, Hogervorst FB, Veer LJ, Kang D, Yoo KY, Noh DY, Ahn SH, Wedren S, Hall P, Low YL, Liu J, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Sigurdson AJ, Stredrick DL, Alexander BH, Struewing JP, Pharoah PD, Easton DF: **A common coding variant in CASP8 is associated with breast cancer risk.** *Nat Genet* 2007, **39:**352-358.
32. Camp NJ, Farnham JM, Cannon-Albright LA: **Localization of a prostate cancer predisposition gene to an 880-kb region on chromosome 22q12.3 in Utah high-risk pedigrees.** *Cancer Res* 2006, **66:**10205-10212.
33. Lange EM, Ho LA, Beebe-Dimmer JL, Wang Y, Gillanders EM, Trent JM, Lange LA, Wood DP, Cooney KA: **Genome-wide linkage scan for prostate cancer susceptibility genes in men with aggressive disease: significant evidence for linkage at chromosome 15q12.** *Hum Genet* 2006, **119:**400-407.
34. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover R, Hunter DJ, Chanock SJ, Thomas G: **Genome-wide association study of**

prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 2007, **39:**645-649.
35. Nasu K, Tanji N, Nishioka R, Wang J, Yanagihara Y, Ozawa A, Yokoyama M, Sakayama K: **EXpression of ErbB proteins in human prostate.** *Arch Androl* 2006, **52:**185-190.
36. Soung YH, Lee JW, Kim SY, Wang YP, Jo KH, Moon SW, Park WS, Nam SW, Lee JY, Yoo NJ, Lee SH: **Somatic mutations of the ERBB4 kinase domain in human cancers.** *Int J Cancer* 2006, **118:**1426-1429.
37. Gallo RM, Bryant I, Fry R, Williams EE, Riese DJ: **Phosphorylation of ErbB4 on Tyr1056 is critical for inhibition of colony formation by prostate tumor cell lines.** *Biochem Biophys Res Commun* 2006, **349:**372-382.
38. Skol AD, Scott LJ, Abecasis GR, Boehnke M: **Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies.** *Nat Genet* 2006, **38:**209-213.
39. Wang H, Thomas DC, Pe'er I, Stram DO: **Optimal two-stage genotyping designs for genome-wide association scans.** *Genet Epidemiol* 2006, **30:**356-368.
40. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnss C, Hoh J: **Complement factor H polymorphism in age-related macular degeneration.** *Science* 2005, **308:**385-389.
41. Despriet DD, Klaver CC, Witteman JC, Bergen AA, Kardys I, de Maat MP, Boekhoorn SS, Vingerling JR, Hofman A, Oostra BA, Uitterlinden AG, Stijnen T, van Duijn CM, de Jong PT: **Complement factor H polymorphism, complement activators, and risk of age-related macular degeneration.** *JAMA* 2006, **296:**301-309.
42. Edwards AO, Ritter R III, Abel KJ, Manning A, Panhuysen C, Farrer LA: **Complement factor H polymorphism and age-related macular degeneration.** *Science* 2005, **308:**421-424.
43. Schaumberg DA, Christen WG, Kozlowski P, Miller DT, Ridker PM, Zee RY: **A prospective assessment of the Y402H variant in complement factor H, genetic variants in C-reactive protein, and risk of age-related macular degeneration.** *Invest Ophthalmol Vis Sci* 2006, **47:**2336-2340.

Health Literacy

An Overlooked Factor in Understanding HIV Health Disparities

Chandra Y. Osborn, PhD, Michael K. Paasche-Orlow, MD, MA, MPH, Terry C. Davis, PhD, Michael S. Wolf, PhD, MPH

Background: Limited health literacy may be a contributing factor to racial disparities in health care. This study examined the mediating effect of limited health literacy on the relationship between race and HIV-medication adherence.

Methods: A total of 204 patients infected with HIV were recruited from two clinics in 2001. Structured in-person interviews were conducted to obtain information on patient demographics, medication adherence, and health literacy. Multivariate regression models were run in 2006 to examine the associations among race, literacy, and HIV-medication adherence after adjusting for relevant covariates.

Results: In an adjusted analysis that excluded literacy, African Americans were 2.40 times more likely to be nonadherent to their HIV-medication regimen than whites (95% confidence interval [CI]=1.14–5.08). When literacy was included in the final model, the effect estimates of race diminished 25% to nonsignificance. Literacy remained a significant independent predictor of nonadherence (adjusted odds ratio [AOR]=2.12, 95% CI=1.93–2.32).

Conclusions: In this study, limited health literacy mediated the relationship between race and HIV-medication adherence. Investigators need to consider the potential utility of responding to literacy and communication barriers in health care as part of interventions to reduce racial disparities.

(Am J Prev Med 2007;33(5):374–378)

Introduction

Recent studies have suggested that limited literacy in adults may contribute to racial disparities in health.[1–3] This is important, as reducing disparities in health care is one of the two primary federal goals for public health in the United States.[4,5] According to the Institute of Medicine (IOM), 90 million people in the U.S. lack the literacy proficiency needed to properly understand and act on health information.[6] This has often been referred to as health literacy, a reflection of both a patient's ability and the literacy prerequisites of the healthcare system.[7] Although limited health literacy has been associated with poorer health outcomes, and also has been shown to be more prevalent among African Americans than whites,[1–3,8,9] it is unclear whether racial differences in health literacy explain the presence of health disparities.

In the context of HIV, individuals with low literacy skills have been more likely to possess a poor working knowledge of their disease and its treatment.[10–15] Kalichman et al. [10,11] found that infected patients with limited literacy had less general knowledge of the disease and their own treatment compared to patients with adequate literacy, and that they were less likely to have an undetectable viral load. While Paasche-Orlow et al.[16] found no association between health literacy and HIV-medication adherence or viral load suppression, Wolf et al.[15] found that patients with low literacy were more than three times as likely to be nonadherent to their anti-retroviral regimens than those with adequate literacy.

The HIV literature also documents substantial racial disparities in health outcomes. Most noteworthy, African Americans with HIV infection are at greater risk for a faster progression to AIDS and shorter survival in comparison to whites.[17,18] Racial differences in medication usage contribute to disparities in health outcomes.[19–21] Possible explanatory factors have been of-

From the Health Literacy and Learning Program, Institute for Healthcare Studies, Feinberg School of Medicine, Northwestern University (Osborn, Wolf), Chicago, Illinois; Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine (Paasche-Orlow), Boston, Massachusetts; and Department of Medicine–Pediatrics, Louisiana State University Health Sciences Center at Shreveport (Davis), Shreveport, Louisiana

Address correspondence and reprint requests to: Chandra Y. Osborn, PhD, Institute for Healthcare Studies, Feinberg School of Medicine, Northwestern University, 676 N. St. Clair St., Suite 200, Chicago IL 60611. E-mail: c-osborn@northwestern.edu.

The full text of this article is available via AJPM Online at www.ajpm-online.net; 1 unit of Category-1 CME credit is also available, with details on the website.

0749-3797/07/$–see front matter
doi:10.1016/j.amepre.2007.07.022

fered, including variability in access to medication[22,23] and adherence practices.[6,24] Although health literacy and race are independently associated with medication adherence,[25] it is unclear to what extent health literacy might be an underlying mechanism that promulgates racial differences in HIV-medication adherence.

Recent studies have suggested that health literacy is a more powerful predictor of health status than race.[1–3] What has yet to be presented are mediating analyses exploring health literacy as a determinant of racial differences in the performance of health-promoting behaviors, such as medication adherence. As such, the objective of this study was to examine the mediating effect of health literacy on the relationship between race and HIV medication adherence.

Methods

Sample

The study sample and methods have been described previously in detail.[13] From June to September 2001, a total of 204 consecutive HIV-infected patients on one or more antiretroviral medications were recruited from outpatient infectious disease clinics at the Northwestern Memorial Hospital (Chicago site) and the Louisiana State University Health Sciences Center at Shreveport (LSUHSC). Patients who had been on their current regimen for less than 2 weeks were excluded from participation, as were those with one or more of the following conditions, as noted in the medical record: (1) dementia, (2) blindness or severely impaired vision not correctable with eyeglasses, (3) deafness or hearing problems uncorrectable with a hearing aid, and/or (4) too ill to participate in the survey. Approval for human subjects research was obtained from institutional review boards at both study sites prior to consenting patients to the study.

Data and Procedure

Trained research assistants received referrals of interested and eligible patients from clinic health providers, then engaged in an informed consent process and conducted a structured interview with recruited patients. All interviews were conducted in a private room at each respective clinic immediately prior to the patients' scheduled physician visits. Information gathered pertained to patient demographic information, medication adherence, and health literacy. Demographic questions specifically included patient age, gender, race, level of educational attainment, employment status, monthly income, and health insurance coverage.

Medication Adherence

Patient anti-retroviral agents, as well as comorbidities and non-HIV prescriptions, were obtained through medical chart reviews. Patients reported any recent missed doses using pages that contained the names and color photographs of common HIV medications included in a revised version of the Patient Medication Adherence Questionnaire (PMAQ).[26,27]

The PMAQ requires patients to identify their medication and then report on a missed dose in the past 4 days for each anti-retroviral agent. Specifically, four questions were asked regarding whether the patient had missed taking a dose yesterday, the day before yesterday, 3 days ago, and over the past weekend. Patients were rated as having proper adherence if they reported no missed doses in this time period, while those acknowledging one or more missed doses within this short timeframe were considered nonadherent.

Health Literacy

Patients were asked to read aloud as many words as they could from a word-recognition list composed of 66 health-related words, Rapid Estimate of Adult Literacy in Medicine (REALM), that were arranged in order of increasing difficulty.[28–30] Scores were based on the total number of words pronounced correctly, with standard English pronunciation being the scoring standard. Correct pronunciation of 0–18 words corresponded to a reading level of third grade or less, and correct pronunciation of 19–44 words corresponded to a fourth- to sixth-grade reading level. Patients who scored from 0 to 44 were considered to have low literacy. Correct pronunciation of 45–60 words indicated a seventh- or eighth-grade reading level, and patients scoring in this level were considered to have marginal literacy. Correct pronunciation of 61–66 words indicated a reading level at the ninth grade or above, and individuals with scores in this range were considered to have adequate literacy.

Data Analysis

All analyses were performed in 2006 (Stata version 9). Chi-square and Student's t-tests were used to evaluate the associations among race, literacy, self-reported missed medication doses, and other sociodemographic characteristics. Patient literacy was classified as adequate (9th grade and higher), marginal (7th–8th grade), and low (6th grade and below). Race was entered as African-American or other. Multivariate regression models were used to analyze the mediational effect of literacy on racial disparities in HIV-medication adherence.[31] Mediating variables were those thought to lie in a causal pathway between the main predictor variable and the outcome. First, the independent relationship between race and medication adherence was established after adjusting for exogenous covariates (gender, age, income, number of medications in regimen, and non-HIV comorbid condition) and potential interaction effects (Model 1). Education, insurance, and employment were not included in the multivariate model due to colinearity with either literacy or other socioeconomic variables in the model. Due to the low prevalence of drug and alcohol abuse (self-report of treatment in past 6 months), this variable also was not included. Second, the relationship between literacy and HIV adherence was to be tested, which had already been demonstrated and confirmed in a prior study using this same cohort.[15] Finally, literacy was added to Model 1 as a mediator, and changes in odds ratios for race were analyzed (Model 2). This approach has been used by others to study the degree to which health behaviors mediate the effect of socioeconomic status on health.[32] Model calibration and discrimination were estimated using the Hosmer–Lemeshow goodness-of-fit chi-square test and the C statistic from receiver operating characteristic (ROC) curves.

Table 1. Sociodemographic characteristics of patients (N=204)

Variable	Percentage
Race	
Black	45.1
Gender	
Male	79.9
Age group	
<40	58.8
40–49	28.4
≥50	12.6
Education	
<High school	12.3
High school graduate	26.0
>High school	61.8
Literacy level	
≤6th grade (Low)	11.3
7th–8th grade (Marginal)	20.1
≥9th grade (Adequate)	68.6
Annual income	
<$10,000	39.7
$10,000–$11,999	23.0
$12,000–$17,999	9.8
≥$18,000	27.5
Insurance	
Private	27.5
Medicare	19.6
Medicaid or free care	52.9
Employment	
Unemployed	55.9
Employed, part-time	15.2
Employed, full-time	28.9
Number of HIV medications in regimen	
1–2 medicines	29.9
≥3 medicines	70.1
Non-HIV comorbid conditions	
≥1 non-HIV comorbid conditions	52.5
Mental illness	
Treated in past 6 months	29.9
Drug or alcohol abuse	
Treated in past 6 months	9.3

Results

Respondents had a mean age of 40.1 years (SD=9.2 years), 79.9% were male and 45.1% were African American. More than half (55.9%) of respondents were unemployed, 39.7% had an annual household income less than $10,000, and 52.9% were either covered by Medicaid or did not carry any health insurance. Over 60% reported at least some college education. Approximately one third had either low or marginal literacy skills; 11.3% were reading at or below a 6th-grade level (low literacy), and 20.1% were reading at a 7th- to 8th-grade level (marginal literacy). More than half (52.5%) of all patients were also receiving treatment for a non–HIV-related chronic illness. Nearly one third reported receiving mental health services, and 9.3% had received treatment for alcohol or illicit drug use in the past 6 months. Over 70% were taking three or more HIV medications. Demographic and clinical characteristics are shown in Table 1.

African-American patients were more likely to possess marginal or low literacy skills compared to non–African-American patients (52.1% vs 14.3%, $p<0.001$), and were significantly less likely to self-report adherence to their medication regimens in the past 4 days (60.1% vs 76.8%, $p=0.014$). Patients with low literacy were more likely to be nonadherent (52.2%) than patients with adequate literacy (30.0%, $p=0.01$).

To confirm the association between race and the potential mediating variable (literacy), a preliminary regression model was analyzed with African-American race as the independent variable and low literacy skills as the dependent variable, controlling for age and study site. A significant association between African-American race and low literacy was confirmed (adjusted odds ratio [AOR]=7.4, 95% confidence interval [CI]=1.49–10.9). For Model 1 examining race and medication adherence, African Americans were 2.40 (95% CI=1.14–5.08; C statistic=0.67) times more likely to be nonadherent to their HIV-medication regimen than non–African Americans (Table 2). As illustrated in Table 2, when literacy was included in the second model, the effect estimates of African-American race diminished by 25% to a point of nonsignificance (AOR=1.80, 95% CI=0.51–5.85; C statistic=0.72). Low literacy remained a significant independent predictor of nonadherence (AOR=2.12, 95% CI=1.93–2.32). Older age and a greater number of medicines in the patient's regimen were also significantly associated with a greater likelihood of missed doses. Patients with an annual income below $10,000 and those also receiving treatment for a non-HIV comorbid condition were less likely to self-report missed doses in the past 4 days.

Discussion

Understanding the reason for pervasive health disparities across race and ethnicity is a major research, practice, and policy goal in the U.S.[33,34] To provide new insights into the pathways that lead to health disparities, patients were recruited from two U.S. regions to examine whether health literacy mediates race disparities in HIV-medication adherence. It was found that African-American patients had a twofold greater likelihood of being nonadherent to their anti-retroviral regimens. Yet consistent with predictions, the inclusion of health literacy reduced the explanatory power of race. In the final model, health literacy, not race, significantly predicted nonadherence. It is believed that this study is the first to examine health literacy as a mediator of race for HIV-medication nonadherence.

While the association between race and medication adherence was reduced by 25% once literacy was entered in the model, clearly other factors contributed to racial disparities. In this study alone it was found that treatment-related aspects of regimen complexity and comorbidity independently predicted adherence, as

Table 2. Multivariate regression analysis for nonadherence to HIV-medication regimens, with and without literacy level

	Model 1		Model 2	
	AOR	**95% CI**	**AOR**	**95% CI**
Race				
White	1.00		1.00	
Black	2.40	1.14–5.08	1.80	0.51–5.85
Gender				
Female	1.00		1.00	
Male	0.94	0.84–2.01	0.97	0.80–1.18
Age group				
<40	1.00		1.00	
40–49	1.29	0.64–2.02	1.29	0.61–2.79
≥50	1.48	1.09–5.99	1.52	1.33–1.72
Annual income				
≥$18,000	1.00		1.00	
$12,000–$17,999	2.26	1.20–1.53	2.19	0.80–6.05
$10,000–$11,999	1.36	0.94–5.47	1.10	0.64–1.90
<$10,000	0.42	0.19–0.93	0.45	0.45–0.78
Number of HIV medications in regimen				
1–2 medicines	1.00		1.00	
≥3 medicines	1.24	1.17–1.32	1.26	1.12–1.32
Non-HIV comorbid condition				
No	1.00		1.00	
Yes	0.74	0.66–0.82	0.70	0.63–0.78
Mental illness				
No prior treatment	1.00		1.00	
Treatment in past 6 months	1.31	0.68–2.47	1.11	0.66–2.59
Literacy level				
≥9th grade (adequate)	–		1.00	
7th–8th grade (marginal)			1.55	0.93–2.45
≤6th grade (low)			2.12	1.93–2.32
Model fit (C statistic)	0.68		0.74	

AOR, adjusted odds ratio; CI, confidence interval.

did age and income. Interestingly, patients with the lowest annual household income or those who were contending with a non-HIV comorbid condition were more likely to be adherent than patients with higher household incomes or those without a non-HIV comorbidity. While these relationships were not entirely clear, it is possible that individuals with the least amount of financial resources were being identified by the healthcare system as being at high risk and subsequently intervined upon. This also could be true for those with greater comorbidity, although it is important to note that only medication-taking behaviors were assessed and not health status or actual treatment outcomes. Future studies should seek to investigate in more detail the role of other patient psychosocial characteristics, such as culture and social support, as possible mediators of the now well-established relationships among literacy, health behaviors, and outcomes.[6,35]

Certain limitations to the study should be acknowledged. First, adherence could not be assessed using one or a combination of other more objective measures, such as random pill counts, Medical Equipment Management Systems (MEMS) caps, or pharmacokinetic laboratory assessments. Although an existing, validated assessment tool was used to measure HIV-medication adherence,[26,27] through questionnaires patients might have underreported missed doses. However, several recent studies have concluded that self-reported measures, such as the PMAQ, are viable and accurate means to measure adherence.[20,36] Second, these data were derived from a cohort of HIV-infected patients interviewed 5 years ago, and might not reflect directly the experience of those on more current HIV-medication regimens. While more recent advances offer the potential for simplified and less restrictive dosing schedules, adherence still remains a significant challenge for patients with the disease.[37,38] Therefore, it is believed that these findings are relevant to the present day. Despite these limitations, this study is the first to assess the impact of limited health literacy in explaining racial/ethnic differences in medication adherence among a sample of patients from both urban and rural settings.

Limited health literacy presents a wide-reaching barrier to disease prevention that, unlike race/ethnicity, is potentially modifiable, which is why this study has important policy implications, particularly with respect to the development of communication strategies to promote HIV-medication adherence that are proven to be effective with patients of all literacy levels. Most health education materials describing medication management and adherence have been written at the high school or college level and may be difficult to understand by individuals with low literacy skills.[39] The development of educational strategies that are both appropriate for lower-literacy audiences and culturally sensitive may benefit the large number of patients who are at risk of nonadherence.[15,40–43] The design of better medication instructions for patients with more-limited literacy, such as improved packaging, labeling, and dispensing practices, is sorely needed.[44] Instructional and educational improvements such as these may help decrease the racial disparities in medication adherence; and, ultimately, contribute to reducing disparities in the early development of AIDS and AIDS-related mortality.

Dr. Osborn conducted this research as a National Research Service Award postdoctoral fellow at the Institute for Healthcare Studies, Feinberg School of Medicine at Northwestern University under an institutional award from the Agency for Healthcare Research and Quality. Dr. Wolf is supported by a Career Development Award through the Centers for Disease Control and Prevention (K01 EH000067–01).

No financial disclosures were reported by the authors of this paper.

References

1. 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.
2. Sentell TL, Halpin HA. Importance of adult literacy in understanding health disparities. J Gen Intern Med 2006;21:862–6.
3. Wolf MS, Knight SJ, Lyons EA, et al. Literacy, race, and PSA level among low-income men newly diagnosed with prostate cancer. Urology 2006; 68:89–93.
4. Bowyer NK, Kleinstein RN. Healthy People 2010—vision objectives for the nation. Optometry 2000;71:569–78.
5. Hixon AL, Chapman RW. Healthy People 2010: The role of family physicians in addressing health disparities. Am Fam Physician 2000;62: 1971–2, 1975, 1978.
6. Institute of Medicine (IOM). Health literacy: a prescription to end confusion. Washington DC: National Academies Press, 2004.
7. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy with health outcomes. Am J Health Behav 2007 in press.
8. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005;20:175–84.
9. Sudore RL, Mehta KM, Simonsick EM, et al. Limited literacy in older people and disparities in health and healthcare access. J Am Geriatr Soc 2006;54:770–6.
10. Kalichman SC, Benotsch E, Suarez T, Catz S, Miller J, Rompa D. Health literacy and health-related knowledge among persons living with HIV/AIDS. Am J Prev Med 2000;18:325–31.
11. Kalichman SC, Ramachandran B, Catz S. Adherence to combination antiretroviral therapies in HIV patients of low health literacy. J Gen Intern Med 1999;14:267–73.
12. Miller LG, Liu H, Hays RD, et al. Knowledge of antiretroviral regimen dosing and adherence: a longitudinal study. Clin Infect Dis 2003;36:514–8.
13. Wolf MS, Davis TC, Arozullah A, et al. Relation between literacy and HIV treatment knowledge among patients on HAART regimens. AIDS Care 2005;17:863–73.
14. Wolf MS, Davis TC, Cross JT, Marin E, Green K, Bennett CL. Health literacy and patient knowledge in a Southern U.S. HIV clinic. Int J STD AIDS 2004;15:747–52.
15. 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.
16. 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.
17. McGinnis KA, Fine MJ, Sharma RK, et al. Understanding racial disparities in HIV using data from the veterans aging cohort 3-site study and VA administrative data. Am J Public Health 2003;93:1728–33.
18. Muir KW, Santiago-Turla C, Stinnett SS, et al. Health literacy and adherence to glaucoma therapy. Am J Ophthalmol 2006;142:223–6.
19. Demmer C. Attitudes toward HIV protease inhibitors and medication adherence in an inner city HIV population. AIDS Patient Care STDs 2003;17:575–80.
20. Fogarty L, Roter D, Larson S, Burke J, Gillespie J, Levy R. Patient adherence to HIV medication regimens: a review of published and abstract reports. Patient Educ Couns 2002;46:93–108.
21. Palacio H, Kahn JG, Richards TA, Morin SF. Effect of race and/or ethnicity in use of antiretrovirals and prophylaxis for opportunistic infection: a review of the literature. Public Health Rep 2002;117:2:252–62.
22. Kahn JG, Zhang X, Cross LT, Palacio H, Birkhead GS, Morin SF. Access to and use of HIV antiretroviral therapy: variation by race/ethnicity in two public insurance programs in the U.S. Public Health Rep 2002;117:263–72.
23. Morin SF, Sengupta S, Cozen M, et al. Responding to racial and ethnic disparities in use of HIV drugs: analysis of state policies. Public Health Rep 2002;117:263–72.
24. Siegel K, Karus D, Schrimshaw EW. Racial differences in attitudes toward protease inhibitors among older HIV-infected men. AIDS Care 2000;12: 423–34.
25. 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.
26. DeMasi RA, Graham NM, Tolson JM, et al. Correlation between self-reported adherence to highly active antiretroviral therapy (HAART) and virologic outcome. Adv Ther 2001;18:163–73.
27. DeMasi RA, Tolson JM, Pham SV, Capuano GA, Fisher RL. Self-reported adherence to HAART and correlation with HIV RNA: Initial results with the patient medication adherence questionnaire. Paper presented at: 6th Conference on Retroviruses and Opportunistic Infections, 1999; Chicago, Illinois.
28. Davis TC, Kennen EM, Gazmararian JA, Williams MV. Literacy testing in health care research. In: Schwartzberg JG, VanGeest JB, Wang CC, eds. Understanding health literacy: Implications for medicine and public health. Chicago IL: AMA Press; 2004:157–179.
29. Davis TC, Long SW, Jackson RH, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993;25:391–5.
30. Davis TC, Michielutte R, Askov EN, Williams MV, Weiss BD. Practical assessment of adult literacy in health care. Health Educ Behav 1998;25: 613–24.
31. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986;51:1173–82.
32. Lantz PM, Lynch JW, House JS, et al. Socioeconomic disparities in health change in a longitudinal study of U.S. adults: the role of health-risk behaviors. Soc Sci Med 2001;53:29–40.
33. Williams DR, Jackson PB. Social sources of racial disparities in health. Health Aff 2005;24:325–34.
34. Lurie N, Jung M, Lavizzo-Mourey R. Disparities and quality improvement: federal policy levers. Health Aff 2005;24:354–64.
35. Lee SY, Arozullah AM, Cho YI. Health literacy, social support, and health: a research agenda. Soc Sci Med 2004;58:1309–21.
36. Simoni JM, Kurth AE, Pearson CR, Pantalone DW, Merrill JO, Frick PA. Self-report measures of antiretroviral therapy adherence: A review with recommendations for HIV research and clinical management. AIDS Behav 2006;10:227–45.
37. 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.
38. Glass TR, De Geest S, Weber R, et al. Correlates of self-reported nonadherence to antiretroviral therapy in HIV-infected patients: the Swiss HIV Cohort Study. J Acquir Immune Defic Syndr 2006;41:385–92.
39. Wilson FL, Racine E, Tekieli V, Williams B. Literacy, readability and cultural barriers: critical factors to consider when educating older African Americans about anticoagulation therapy. J Clin Nurs 2003;12:275–82.
40. Mullins CD, Blatt L, Gbarayor CM, Yang HW, Baquet C. Health disparities: a barrier to high-quality care. Am J Health Syst Pharm 2005;62:1873–82.
41. Hussey LC. Minimizing effects of low literacy on medication knowledge and compliance among the elderly. Clin Nurs Res 1994;3:132–45.
42. Kalichman SC, Cherry J, Cain D. Nurse-delivered antiretroviral treatment adherence intervention for people with low literacy skills and living with HIV/AIDS. J Assoc Nurses AIDS Care 2005;16:3–15.
43. Paasche-Orlow MK, Riekert KA, Bilderback A, et al. Tailored education may reduce health literacy disparities in asthma self-management. Am J Respir Crit Care Med 2005;172:980–6.
44. Wolf MS, Davis TC, Tilson HH, Bass PF 3rd, Parker RM. Misunderstanding of prescription drug warning labels among patients with low literacy. Am J Health Syst Pharm 2006;63:1048–55.

The Causal Pathways Linking Health Literacy to Health Outcomes

Michael K. Paasche-Orlow, MD, MA, MPH; Michael S. Wolf, PhD, MPH

***Objective*: To provide an evidence-based review of plausible causal pathways that could best explain well-established associations between limited health literacy and health outcomes. *Methods*: Through analysis of current findings in medical and public health literature on health literacy we derived a conceptual causal model. *Results*: Health literacy should be viewed as both a patient and a system phenomenon. Three distinct points along a continuum of health care are suggested to be influenced by health literacy: (1) access and utilization of health care, (2) patient-provider relationship, and (3) self-care. *Conclusions*: The conceptual model organizes what has been learned to date and underscores promising areas of future inquiry and intervention.**

***Key words*: causal pathways, health literacy, literacy, conceptual model**

***Am J Health Behav.* 2007;31(Suppl 1):S19-S26**

Though limited health literacy has been shown to be associated with worse health outcomes,[1] the causal pathways are not entirely known,[2,3] and several projects have focused on explaining potential mechanisms.[3-8] This paper provides a review of the current evidence and proposes a conceptual model describing the systemic, interactional, and self-care mechanisms by which limited health literacy is most likely to lead to worse health outcomes. By doing so, we begin to highlight some of the most promising areas for intervention research as well as important gaps in our current understanding of the pathways linking literacy and health. This is one step in what will need to be an iterative process of model specification and clarification. It is our hope that investigators will increasingly pursue research designs and analytic approaches to refine the model so that the most valid and useful explanations of the relationship between health literacy and outcomes can inform professional responses to the problem in the many diverse contexts of health care.[9]

It should first be recognized that limited health literacy is strongly associated with other socioeconomic indicators, including educational attainment, race/ethnicity, and age.[10] Such associations make it difficult to discern the independent effect of health literacy from the complex relationships known between these latent and evolving traits that are also interrelated themselves. Yet many of the determinants of disease, in general,

Michael K. Paasche-Orlow, Assistant Professor of Medicine, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts and Boston University Medical Center, Boston, MA. Michael S. Wolf, Assistant Professor and Director, Health Literacy and Learning Program, Institute for Healthcare Studies, Feinberg School of Medicine, Northwestern University, Chicago, IL.

Address correspondence to Dr Paasche-Orlow, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts and Boston University Medical Center, 91 East Concord Street, Suite 200, Boston, MA 02118. E-mail: mpo@bu.edu

are interconnected in complex ways.[11] In this sense, the current model is focused on the direct pathways between literacy and health outcomes, while recognizing there are likely to be many indirect associations that will eventually need to be explored by researchers in greater detail.

Definition of Health Literacy and Implications for a Causal Framework

The National Institutes of Health (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."[12,13] We agree with this statement. However, specific clarification is required. According to this definition, health literacy relates to an individual's possession of requisite skills for making health-related decisions. This means that health literacy must always be examined in the context of the specific tasks that need to be accomplished. When we refer to health literacy, we include not only a person's ability, but also the complexity of the tasks at hand. Much of the literature on health literacy focuses on the capacity of individuals in the context of health activities. However, we feel that it is important to underscore the importance of a contextual appreciation of health literacy. As such, causal mechanisms of the health literacy-health outcomes relationship are due not only to patient-level characteristics, but also to those attributes of the health care system. Therefore, the proposed conceptual causal model diagrammed in Figure 1 recognizes both individual and system-level factors that affect access to health care, medical encounters, and self-care activities.

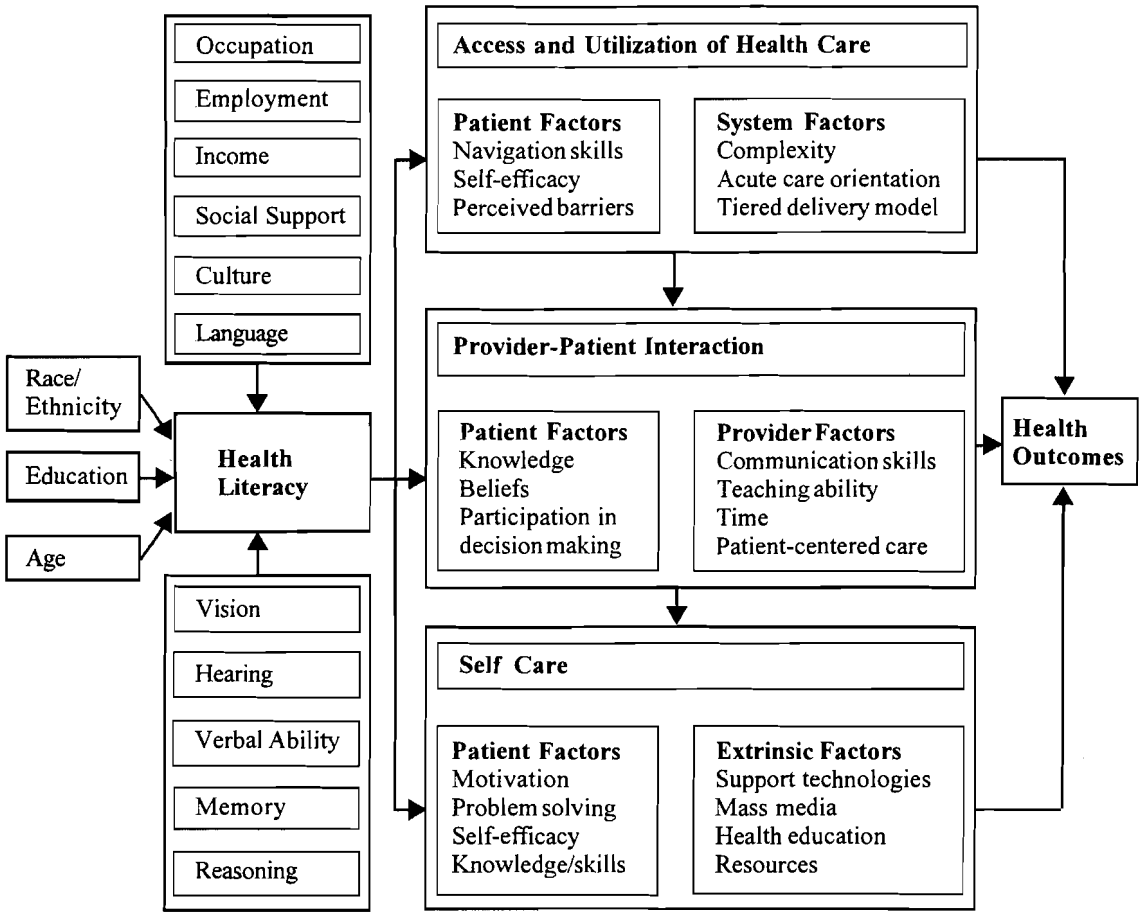
Access and Utilization of Health Care

Patient factors. People typically initiate medical care when they have a medical need. However, people with limited health literacy may delay seeking care due to not understanding prevention and/or knowing the signs and symptoms of disease that should motivate people to action. Additionally, adults with limited health literacy may be less comfortable with the social interactions that occur in medical settings, worry that their limited literacy will be exposed, and often have feelings of shame.[14] Similarly, limited health literacy has been associated with more distrust of providers, pessimism about treatment, lower satisfaction, and a worse assessment of the quality of care.[15,16] As much of routine prevention is dependent on patients coming into primary care, it is not surprising that people with limited health literacy receive less primary prevention.[17-19]

A concept that has been advanced to explain how people with limited health literacy may have difficulty with using health services is the notion of navigation. The concept is meant to include all of the skills needed to go from one place to another in the pursuit of medical care. For example, people with limited health literacy have reported difficulty finding their way in health care facilities, likely the result of poor communication (verbal directions, signage and placards, complexity of instructions). To date, measures of navigation in this sense have not been presented, and no data have been published that directly reveal how patients maneuver their way through the trajectory of large and complicated health systems.

Utilization of care, however, can be viewed as a global proxy measure of functional navigation. Multiple studies have described higher rates of emergency room utilization, hospitalization, and rehospitalization.[15,20-23] However, self-reported access to physician visits was not significantly worse for patients with limited health literacy,[24] and among new Medicare managed care enrollees, there were no significant differences in the mean number of primary care provider visits or the time to a first visit.[25] Another study by Lincoln et al examined utilization of various health services without finding differences by level of literacy.[26] Unfortunately, as almost all of the data presented in the medical literature on health literacy relates to patients evaluated in medical contexts, few studies have captured the experience of people who remain largely outside of the health care system. Millions of people are eligible but not enrolled in programs such as Medicaid and other means-tested health benefits.[27] Although such factors have not been thoroughly examined, it is possible that shame or lack of understanding of potential benefits act as perceived barriers to enrollment, an important and unexplored pathway leading to worse health among patients with limited literacy.

Figure 1
Causal Pathways between Limited Health Literacy and Health Outcomes



System factors. One consequence of having multiple nonprofit, for-profit, and government payers in the health care marketplace is that the US health care system has become increasingly complicated and confusing. The financial structure of medical insurance is directly organized around the idea that higher premiums command fewer logistical barriers to care. The rules and regulations of how to use health plans vary, and it is easy to misunderstand what services are covered and how to use various types of programs. Similarly, health programs do not comprehensively manage the flow of information that commonly has to be processed separately for medical, pharmacy, and dental benefits – if one has coverage. Furthermore, rule changes (eg, doctors are shifted out of network and medications are dropped from formularies) frequently befuddle clinicians and patients alike. No metric has emerged for how much of the health disparity associated with health literacy might be due to the overall complexity of the health care system.

Patient-Provider Interactions

Patient factors. Knowing less about a disease may have broad ramifications on how patients with limited literacy interact with their providers. Patients may or may not have insight into the fact that they do not know very much about their condition. Beyond the issue of how shame

might lead some people with low health literacy to avoid asking questions of their clinicians, low literacy may be associated more generally with worse patient activation. In addition, low literacy has been linked to worse mental health functioning and worse depression.[26,28,29] These connections may lead patients with low health literacy to have a tendency to be more passive and may complicate the patient-physician interaction and lead to miscommunication.

Provider factors. Providers tend to be unaware of their patients' limited literacy.[30,31] However, it is not clear that even with such knowledge providers would be able to facilitate successful communication. To do so, they would need to learn how to communicate in plain terms to confirm important items in a patient's history and evaluate patient comprehension of important action items.[7,32] Avoiding medical jargon is not always sufficient; though underutilized, pictures, multimedia, and decision aids can often help. Oftentimes, a great amount of information is relayed to patients, but providers seldom evaluate patient comprehension in any meaningful manner.[33] Providers frequently do not employ interview techniques that are useful for clinical encounters with patients with low literacy. It is unclear how much of this problem is due to a lack of interviewing skills, a lack of engagement and commitment to patient education, or a maladaptive response to imposed time pressures. During the fast-paced and potentially anxiety-driven encounter, patients with limited literacy may be less responsive and reluctant to ask questions, admit confusion, use different descriptions of medical problems than clinicians are used to, and provide less accurate information about their medical histories and medications.[14,34,35]

A patient-centered approach of confirming comprehension, known as teachback, has been advanced as a new standard of care for clinicians.[36] Although this approach is sensible and will not put patients at risk, it is appropriate to note that evidence has not yet been presented showing that adopting universal precautions for confirming patient understanding in the medical encounter leads to improved health outcomes for patients with limited literacy. Similarly, although the teach-to-goal educational technique, in which a learner is asked to exhibit comprehension and gets repeated rounds of focused teaching until mastery is exhibited, has been advanced as a key method for improving comprehension, it has been shown to be effective only in a small and short study of adults with asthma.[23] Comprehension of instrumental activities, such inhaler technique, may play a significant causal role in chronic disease management; however, insufficient data have been collected to actually know if this will be a generalizable intervention approach.

Self-care

Patient factors. People with limited health literacy have been found not only to have lower levels of disease specific knowledge, but also to know less of the practical and instrumental knowledge critical to self-management such as which medication to take in an asthma exacerbation. In 2 studies, asthma knowledge mediated the relationship between low literacy and asthma-related skills (inhaler technique) and outcomes (asthma quality of life and treatment in emergency department).[23,29] Although few data have been presented, limited problem-solving skills and low self-efficacy have been suggested as well. Two recent observational studies showed no association between self-efficacy and low literacy;[29,37] however in 2 diabetes management intervention trials subjects with low literacy have been shown to have more improvement than subjects with adequate literacy, which suggests a possible moderating role of self-efficacy.[38,39]

Self-management requires not only the knowledge of what to do but also the will and capacity to carry out the health care plan. A critical element of successful self-management is medication adherence. On this front, the evidence has been mixed. Although patients with limited literacy have more trouble understanding primary and precautionary medication label instructions and are less likely to be able to report the name of their medication, there is no consistent finding of worse medication adherence among patients with limited literacy.[40,41] Indeed, although Wolf et al found patients with low literacy and HIV to have a 3.3 (95% confidence interval 1.3 to 8.7) adjusted odds of nonadherence to antiretroviral therapy, Paasche-Orlow et al found patients with

low literacy and HIV to have a 1.9 (95% confidence interval 0.84 to 4.30) adjusted odds ratio of adherence.[42,43] Little has been presented on the independent relationship between health literacy and specific healthy and unhealthy behaviors, yet 3 papers provide compelling evidence suggesting a lack of association between health literacy and unhealthy alcohol and drug use, among other risk factors.[26,43,44] In light of other conflicting evidence detailing the association between low health literacy and medication adherence behaviors,[45-48] other proposed variables, such as self-efficacy to effectively manage the set of medication management tasks, might be important.[42] However, DeWalt et al did not find a similar mediating relationship of self-efficacy to the literacy-health outcomes relationship for glycemic control among type 2 diabetics.[49] Their study suggests other mediating factors, such as desire to participate in care and health knowledge.

Extrinsic factors. Self-management is now the accepted term used to describe the day-to-day decisions and activities that patients, with the help of loved ones, engage in to live with and control their illnesses.[50-52] Patients need to interpret their symptoms and make decisions. Which over-the-counter medications will they use, and how should they use it? What is a side effect that warrants stopping a medication? When a provider gives a patient with low health literacy a prescription, will the patient know how to take it? Pharmacies typically fulfill their obligation to educate patients about their prescriptions by giving patient education materials that they cannot read. Similarly, patients contend with competing and confusing health messages from the news, the public health sector, and direct-to-consumer advertising.

Patients with chronic disease are commonly in the position of needing to monitor their condition and adjust their therapy. Because patients with low health literacy are less able to understand their disease and symptoms or to know how to monitor their disease (eg, peak flow meters), they also face greater difficulty using monitoring devices correctly (eg, glucometer) and may not be able to interpret results to know what to do (eg, increase Lasix dose). A critical limitation is the lack of supportive technology and 2-way data.[36,53] Although most activities of self-care can be supported by technologies that allow remote synchronous or asynchronous assistance, few self-management support systems have been tested for use with patients with low literacy.[39,54]

DISCUSSION

In presenting a causal model, we add a word of caution that we do not offer a deterministic cause-and-effect schema. The model we have detailed takes the form of a component-cause model as is commonly undertaken in epidemiological research.[55] At one level, this could be merely an exercise in measurement of probability distributions (eg, How much of the relationship between a cause and an effect is ascribed to an intervening phenomenon?) and not the result of a criterion-driven process.[56] The primary reason that the current model is not more heavily driven by empirical evidence is that relevant data are limited. Similarly, much of the literature on literacy and health has been cross-sectional, which limits attempts to determine cause and effect relationships.[57] Nonetheless, there are certain rules that govern what can be rightfully viewed as causal (eg, Does the cause precede the effect? Is the potential variance due to the cause deemed significant? Is the effect plausible?) such that subjective judgments based on observations by investigators in the field can also contribute important details to the theoretical model we present.[58]

Though it is appealing to have a conceptual framework that incorporates the entire array of critical interrelated phenomena like socioeconomic status, ethnicity, environment, social support, and other variables thought to influence how people are affected by limited health literacy, there is also benefit from keeping the framework straightforward and unencumbered by controversies in other domains. For this reason, we have focused on only the most direct paths between literacy and health. Yet the model remains incomplete in its current form, as it does not address several areas of complexity that may later be revealed as having considerable importance for different reasons. For instance, the model treats the concept of literacy as a single entity and ignores key problems in the field that stem from measurement issues. The concept of health literacy includes multiple elements (eg, listening,

verbal fluency, memory span, navigation) that are not entirely captured in current tests. Strictly speaking, even the current literature which is restricted to instruments that deal exclusively with the tasks of reading comprehension, word pronunciation, and calculation is challenging to interpret because studies have used varying inclusion and exclusion criteria.[59] It is likely that the causal interactions are sensitive to what concepts are included in the dependent variable.

Our proposed model also treats literacy as a fixed characteristic, even though it is subject to change over time; and depending on the cohort, the nature of change may not be easily predicted. Literacy improvement can occur in a general sense (adult basic education) or in a specific health literacy fashion, such as a specific self-care skill.[23,60] Literacy is also most likely to decline over time, in relation to or independent of conditions that increase in prevalence as we age, such as dementia. The relationships depicted in our model might also be thought to imply that health literacy is a dichotomous variable. Though most studies have presented data in this way, the form of the relationship between literacy and various elements in the model should be considered an empirical question. However, there is no reason to think that literacy will relate to health outcomes in a simple linear fashion; rather, a step function with a threshold effect may be more likely.

One of the reasons that literacy may not relate to health outcomes in a linear fashion relates to the fact that people generally exist within a web of social relationships. Below a certain level of function, much of the day-to-day detail of chronic disease management often needs to be facilitated by others. Similarly, there are many clinical scenarios (eg, pediatrics, dementia, severe cognitive impairment) in which the obvious focus of health literacy research needs to be directed to the literacy skills of the relevant care providers. Although the health impact of social cohort effects has not been fully elucidated in the context of health literacy, the interaction between health literacy and social support is likely to have complicated and subtle implications, as patients with low educational attainment are more likely to have social support from people who also have lower educational attainment than do patients with higher educational attainment.

Finally, the model presented herein does not specify the nature of causation. Some variables appear to operate primarily as mediators (eg, people with low literacy have worse health outcomes due to the fact that they have worse knowledge), and other variables appear to operate primarily as moderators (ie, people with low literacy have worse health outcomes due to the heightened dependence on social support).[61] There have been no longitudinal empirical evaluations to parse the manner in which many of the variables actually operate. Indeed, such effects can be difficult to elucidate, and a given variable may be both mediating and moderating in different contexts.[62]

Each of these limitations relates to ways the model simplifies a multitude of complex relationships, yet exhibiting the multicausality inherent in the current model was a sufficiently complicated task that requires acknowledgment.[56] Researchers must begin to think more deeply about these relationships and, by doing so, explore the greater implications of the model for the design of appropriate intervention strategies. In turn, thinking conceptually about the different causal pathways between health literacy and specific outcomes of interest may lead to the consideration of different approaches, pending the identified context. Ultimately, intervention efforts are better served by acknowledging the interdependence of the various sociocultural factors that need to be managed in order to succeed in promoting improved health care for those with limited health literacy. ■

REFERENCES

1. DeWalt DA, Berkman ND, Sheridan S, et al. Literacy and health outcomes: a systematic review of the literature. *J Gen Intern Med.* 2004;19(12):1228-1239.
2. DeWalt DA, Pignone MP. Reading is fundamental: the relationship between literacy and hHealth. *Arch Intern Med.* 2005;165:1943-1944.
3. Baker DW. Reading between the lines: deciphering the connections between literacy and health. *J Gen Intern Med.* 1999;14(5):315-317.
4. Pignone M, DeWalt D, Sheridan S, et al. Interventions to improve health outcomes for patients with low literacy: a systematic review. *J Gen Intern Med.* 2005;20(2)185-192.
5. Roter D. Health literacy and the patient-provider relationship. In: Schwartzberg JG, VanGeest JB, Wang CC, editors. Understand-

ing Health Literacy: Implications for Medicine and Public Health. American Medical Association Press 2005:87-100.
6. Rudd RR, Renzulli D, Pereira A, Daltroy L. Literacy demands in health care settings: the patient perspective. In: Schwartzberg JG, VanGeest JB, Wang CC, editors. Understanding Health Literacy: Implications for Medicine and Public Health. American Medical Association Press 2005:69-84.
7. Castro CM, Wilson C, Schillinger D. Babel babble: physicians' use of jargon with diabetes patients. *J Gen Intern Med.* 2004;19(Suppl 1):124.
8. Rootman I, Ronson B. Literacy and health research in Canada: where have we been and where should we go? *Can J Public Health.* 2005;96(Suppl 2):S62-S77.
9. Olsen J. What characterises a useful concept of causation in epidemiology? *J Epidemiol Comm Health.* 2003;57(2):86-88.
10. Brown H, Prisuta R, Jacobs B, Campbell A. U.S. Department of Education. National Center for Education Statistics. Literacy of Older Adults in America: Results from the National Adult Literacy Survey, NCES 97-576. Washington, DC: 1996.
11. Lahelma E, Martikainen P, Laaksonen M, Aittomaki A. Pathways between socioeconomic determinants of health. *J Epidemiol Comm Health.* 2004;58(4):327-332.
12. Institute of Medicine. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press 2004.
13. Ratzan SC, Parker RM. Introduction. Selden CR, Zorn M, Ratzan SC, Parker RM, Editors. In: National Library of Medicine Current Bibliographies in Medicine: Health Literacy. Available at: http://www.nlm.nih.gov/pubs/cbm/hliteracy.html. Accessed January 15, 2004]. Vol. NLM Pub. No. CBM 2000-1. 2000. National Institutes of Health, U.S. Department of Health and Human Services, Bethesda, MD.
14. Baker DW, Parker RM, Williams MV, et al. The health care experience of patients with low literacy. *Arch Fam Med.* 1996;5(6):329-334.
15. Kalichman SC, Rompa D. Functional health literacy is associated with health status and health-related knowledge in people living with HIV-AIDS. *J Acquir Immune Defic Syndr.* 2000;25(4):337-344.
16. Mancuso CA, Rincon M. Asthma patients' assessments of health care and medical decision making: the role of health literacy. *J Asthma.* 2006;43(1):41-44.
17. Scott TL, Gazmararian JA, Williams MV, Baker DW. Health literacy and preventive health care use among medicare enrollees in a managed care organization. *Med Care.* 2002;40(5):395-404.
18. Bennett CL, Ferreira MR, Davis TC, et al. Relation between literacy, race, and stage of presentation among low-income patients with prostate cancer. *J Clin Oncol.* 1998;16(9):3101-3104.
19. Wolf MS, Knight SJ, Lyons EA, et al. Literacy, race, and PSA level among low-income men newly diagnosed with prostate cancer. *Urology.* 2006;68(1):89-93.
20. Baker DW, Parker RM, Williams MV, Clark WS. Health literacy and the risk of hospital admission. *J Gen Intern Med.* 1998;13(12):791-798.
21. Baker DW, Gazmararian JA, Williams MV, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. *Am J Public Health.* 2002;92(8):1278-1283.
22. Gordon MM, Hampson R, Capell HA, Madhok R. Illiteracy in rheumatoid arthritis patients as determined by the Rapid Estimate of Adult Literacy in Medicine (REALM) score. *Rheumatology* (Oxford). 2002;41(7):750-754.
23. Paasche-Orlow MK, Riekert KA, Bilderback A, et al. Tailored education may reduce health literacy disparities in asthma self-management. *Am J Respir Crit Care Med.* 2005;172(8):980-986.
24. Baker DW, Parker RM, Williams MV, et al. The relationship of patient reading ability to self-reported health and use of health services. *Am J Public Health.* 1997;87(6):1027-1030.
25. Baker DW, Gazmararian JA, Williams MV, et al. Health literacy and use of outpatient physician services by Medicare managed care enrollees. *J Gen Intern Med.* 2004;19(3):215-220.
26. Lincoln A, Paasche-Orlow MK, Cheng DM, 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(8):818-822.
27. Piette JD, Wagner TH, Potter MB, Schillinger D. Health insurance status, cost-related medication underuse, and outcomes among diabetes patients in three systems of care. *Med Care.* 2004;42(2):102-109.
28. Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional health status among older adults. *Arch Intern Med.* 2005;165:1946-1952.
29. Mancuso CA, Rincon M. Impact of health literacy on longitudinal asthma outcomes. *J Gen Intern Med.* 2006;21(8):813-817.
30. Bass PF, III, Wilson JF, Griffith CH, Barnett DR. Residents' ability to identify patients with poor literacy skills. *Acad Med.* 2002;77(10):1039-1041.
31. 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(5):938-943.
32. Schillinger D, Bindman A, Wang F, et al. Functional health literacy and the quality of physician-patient communication among diabetes patients. *Patient Educ Couns.*

2004;52(3):315-323.
33. Schillinger D, Piette J, Grumbach K, et al. Closing the loop: physician communication with diabetic patients who have low health literacy. *Arch Intern Med.* 2003;163(1):83-90.
34. Kalichman SC, Rompa D, Cage M. Reliability and validity of self-reported CD4 lymphocyte count and viral load test results in people living with HIV/AIDS. *Int J STD AIDS.* 2000;11(9):579-585.
35. Paasche-Orlow MK, Parker RM. Improving the effectiveness of patient education: a fFocus on limited health literacy. In: Talmadge E.King, Margaret B.Wheeler, editors. Medical Management of Vulnerable and Underserved Patients: Principles, Practice, and Populations. New York: McGraw Hill 2007:101-109.
36. Paasche-Orlow MK, Schillinger D, Greene SM, Wagner EH. How health care systems can begin to address the challenge of limited literacy. *J Gen Intern Med.* 2006;21(8):884-887.
37. Sarkar U, Fisher L, Schillinger D. Is self-efficacy associated with diabetes self-management across race/ethnicity and health literacy? *Diabetes Care.* 2006;29(4):823-829.
38. Kim S, Love F, Quistberg DA, Shea JA. Association of health literacy with self-management behavior in patients with diabetes. *Diabetes Care.* 2004;27(12):2980-2982.
39. Rothman RL, DeWalt DA, Malone R, et al. Influence of patient literacy on the effectiveness of a primary care-based diabetes disease management program. *JAMA.* 2004;292(14):1711-1716.
40. Pignone MP, DeWalt DA. Literacy and health outcomes: is adherence the missing link? *J Gen Intern Med.* 2006;21(8):896-897.
41. Gazmararian JA, Kripalani S, Miller MJ, et al. Factors associated with medication refill adherence in cardiovascular-related diseases: a focus on health literacy. *J Gen Intern Med.* 2006;21(12):1215-1221.
42. Wolf MS, Davis TC, Osborn C, et al. Literacy, self-efficacy, and HIV medication adherence. *Patient Educ Couns.* 2006;65(2):253-260.
43. Paasche-Orlow MK, Cheng DM, Palepu A, et al. Health literacy, antiretroviral adherence, and HIV-RNA suppression: a longitudinal perspective. *J Gen Intern Med.* 2006;21(8):835-840.
44. Wolf MS, Gazmararian JA, Baker DW. Health literacy and health risk behaviors among older adults. *Am J Prevent Med.* 2007;32(1):19-24.
45. 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(10):756-765.
46. Kalichman SC, Ramachandran B, Catz S. Adherence to combination antiretroviral therapies in HIV patients of low health literacy. *J Gen Intern Med.* 1999;14(5):267-273.
47. Miller LG, Liu H, Hays RD, et al. Knowledge of antiretroviral regimen dosing and adherence: a longitudinal study. *Clin Infect Dis.* 2003;36(4):514-518.
48. Wolf MS, Davis TC, Cross JT, et al. Health literacy and patient knowledge in a Southern US HIV clinic. *Int J STD AIDS.* 2004;15(11):747-752.
49. DeWalt DA, Boone RS, Pignone MP. Literacy and its relationship with self-efficacy, trust, and participation in medical decision-making. *Am J Health Behav.* 2007;31(Suppl 1 Sep/Oct):27-35.
50. Marks R, Allegrante JP, Lorig K. A review and synthesis of research evidence for self-efficacy-enhancing interventions for reducing chronic disability: implications for health education practice (part II). *Health Promot Pract.* 2005;6(2):148-156.
51. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. *Ann Behav Med.* 2003;26(1):1-7.
52. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. *JAMA.* 2002;20;288(19):2469-2475.
53. Piette JD, Schillinger D. Applying interactive health technologies for vulnerable populations. In: Talmadge E.King, Margaret B.Wheeler, editors. Medical Management of Vulnerable and Underserved Patients: Principles, Practice, and Populations. New York: McGraw Hill 2007:131-140.
54. DeWalt DA, Pignone M, Malone R, et al. Development and pilot testing of a disease management program for low literacy patients with heart failure. *Patient Educ Couns.* 2004;55(1):78-86.
55. Karhausen LR. Causation: the elusive grail of epidemiology. *Med Health Care Philos.* 2000;3(1):59-67.
56. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health.* 2005;95(S1):S144-150.
57. Schillinger D, Barton LR, Karter AJ, et al. Does literacy mediate the relationship between education and health outcomes? A study of a low-income population with diabetes. *Public Health Rep.* 2006;121(3):245-254.
58. Morabia A. Epidemiological causality. *Hist Philos Life Sci.* 2005;27(3-4):365-379.
59. Paasche-Orlow MK, Parker RM, Gazmararian JA, et al. The prevalence of limited health literacy. *J Gen Intern Med.* 2005;20(2):175-184.
60. Weiss BD, Francis L, Senf JH, et al. Literacy education as treatment for depression in patients with limited literacy and depression: a randomized controlled trial. *J Gen Intern Med.* 2006;21(8):823-828.
61. Arozullah AM, Lee SY, Khan T, et al. The roles of low literacy and social support in predicting the preventability of hospital admission. *J Gen Intern Med.* 2006;21(2):140-145.
62. Thrane C. Explaining educational-related inequalities in health: mediation and moderator models. *Soc Sci Med.* 2006;62(2):467-478.

Depressive Symptoms and Subsequent Alcohol Use and Problems: A Prospective Study of Medical Inpatients With Unhealthy Alcohol Use*

TIBOR P. PALFAI, PH.D.,† DEBBIE M. CHENG, SC.D.,† JEFFREY H. SAMET, M.D., M.P.H.,† KEVIN L. KRAEMER, M.D., M.SC.,† MARK S. ROBERTS, M.D., M.P.P.,† AND RICHARD SAITZ, M.D., M.P.H.†

Department of Psychology, College of Arts and Sciences, Boston University, 648 Beacon Street, Boston, Massachusetts, 02215

ABSTRACT. Objective: We sought to determine whether depressive symptoms assessed during hospitalization predicted alcohol use and alcohol-related problems during the subsequent year among medically ill patients. **Method:** The study sample was a cohort of hospitalized medical patients with unhealthy alcohol use who participated in a randomized controlled trial of an alcohol brief intervention. Depressive symptoms at baseline, assessed with the Center for Epidemiologic Studies-Depression Scale (CES-D), were used to predict alcohol use and problems at 3 and 12 months. **Results:** Of the 341 patients enrolled, 90% (men = 220; women = 88) provided data on at least one follow-up time point during the subsequent year. Gender-stratified longitudinal Poisson regression models were fit for each alcohol outcome, adjusting for baseline values of age, physical symptoms, randomization group, alcohol outcome, cocaine use, and socioeconomic indicators. Depressive symptoms were significantly associated with drinks per day (men: incidence rate ratio [IRR] = 1.17 per 10-unit increase in CES-D, $p < .01$; women: IRR = 1.00 per 10-unit increase in CES-D, $p = .98$) and alcohol-related problems (men: IRR = 1.22 per 10-unit increase in CES-D, $p < .001$; women: IRR = 1.05 per 10-unit increase in CES-D, $p = .39$) for men but not for women. They were not significantly associated with the number of days abstinent in men or women. **Conclusions:** In hospitalized medical patients with unhealthy alcohol use, depressive symptoms predict subsequent drinks per day and alcohol-related problems among men. These findings suggest that symptoms of depression may be important to consider in treatment planning for male medical patients with unhealthy patterns of drinking. (*J. Stud. Alcohol Drugs* **68:** 673-680, 2007)

THERE IS CONSIDERABLE EVIDENCE to suggest that patients in medical settings have a higher prevalence of alcohol-use disorders and exhibit more risky drinking behavior than the general population (Maisto and Saitz, 2003). This "unhealthy drinking," which includes the spectrum of drinking patterns that place one at risk for harm to alcohol dependence (Saitz, 2005), is also observed among those who use medical services more frequently (Chou et al., 1996; Curtis et al., 1986; Moore et al., 1989). The prevalence of unhealthy drinking appears to be particularly high among hospitalized medical patients. These patients are more likely to report problems associated with their drinking (Jarman and Kellett, 1979; Moore et al., 1989) and meet criteria for alcohol-use disorders (Booth et al., 1998) than those in the community.

Medical patients who are unhealthy drinkers are also more likely to exhibit clinical depression and elevated depressive symptoms (Alati et al., 2004; Amodei et al., 1994; Kim et al., 2003; Roeloffs et al., 2002). Amodei et al., for example, found that, of those who met criteria for an alcohol-use disorder in an urban primary care center, 38% experienced significant depressive symptoms within the past 30 days, compared with 13% of people without an alcohol-use disorder. Similarly, Alati et al. found that both male and female harmful drinkers in the emergency room, defined by cutoff scores on the Alcohol Use Disorder Identification Test (AUDIT; Saunders et al., 1993), showed higher levels of depressive symptoms compared with those with lower AUDIT scores. The co-occurrence of unhealthy alcohol use and elevated depressive symptoms may be of

Received: February 9, 2007. Revision: March 25, 2007.

*This research was supported by National Institute on Alcohol Abuse and Alcoholism grant RO1-AA-12617.

†Correspondence may be sent to Tibor P. Palfai at the above address or via email at: palfai@bu.edu. Tibor P. Palfai is also with the Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA. Debbie M. Cheng is with the Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, and the Department of Biostatistics, Boston University School of Public Health, Boston, MA. Jeffery H. Samet is with the CARE Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, and the Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA. Kevin L. Kraemer is with the Division of General Internal Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA. Mark S. Roberts is with the Division of General Internal Medicine and the Section of Decision Sciences and Clinical Systems Modeling, University of Pittsburgh School of Medicine, Pittsburgh, PA. Richard Saitz is with the CARE Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, and the Youth Alcohol Prevention Center and Department of Epidemiology, Boston University School of Public Health, Boston, MA.

particular concern for medical patients given their significant negative effects on physical health outcomes (Aneshensel et al., 1984) and mortality risk (Greenfield et al., 2002).

Although research suggests that medical patients with unhealthy drinking may exhibit more depressive symptoms, there is relatively little known about how these symptoms are associated with subsequent alcohol-use patterns and alcohol-related problems. Given that depression may be a transient consequence of changes in alcohol-use patterns (Raimo and Schuckit, 1998), it is unclear whether depressive symptoms experienced by unhealthy drinking patients in the hospital provide any information about alcohol-use patterns during the subsequent year. Higher levels of depressive symptoms among medical patients with unhealthy alcohol use may be the result of the influences of alcohol-use patterns on depression (e.g., substance-induced mood changes) but also to a variety of shared causal mechanisms (e.g., socioeconomic factors) and influences of mood on alcohol-use motivation or self-regulatory functions (Li et al., 2004; Swendsen and Merikangas, 2000). For example, the use of alcohol to cope with negative affect and life stressors may present a particular risk for unhealthy alcohol use among patients with depressive symptoms (Holahan et al., 2003), an alcohol-use motive that has been associated with an increased incidence of alcohol-related problems (Cooper et al., 2000; Gaher et al., 2006) and the development of alcohol-use disorders (Carpenter and Hasin, 1999).

As the association between major depression and alcohol-use disorders may be stronger for women (Grant and Harford, 1995; Kessler et al., 1997), and depression may be more likely to precede alcohol dependence among women than men (Hesselbrock and Hesslebrock, 1997), it is often thought that drinking to self-medicate depressive symptoms may play a more important role in the unhealthy drinking of women (see Schutte et al., 1997). Although longitudinal studies with community samples appear to suggest that depressive symptoms may be a stronger predictor of long-term (2-10 years) alcohol outcomes for women than men (Hartka et al., 1991; Moscato et al., 1997), results from studies with shorter follow-up intervals (≤1 year) have not clearly supported a stronger link between depressive symptoms and alcohol use among women (e.g., Aneshensel and Huba, 1983; Caldwell et al., 2002; Graham et al., 2007), with some evidence suggesting that depression may be a risk factor for increased transitions to high-risk drinking for men but not for women among alcohol users who have experienced an alcohol-related problem (Crum et al., 2001).

Prospective studies of alcohol-treatment samples have also produced inconsistent findings regarding the effects of depression on alcohol use during the subsequent year, with some studies suggesting that indices of major depression may be associated with relapse following treatment for alcohol dependence (Bobo et al., 1998; Greenfield et al., 1998) or greater risk of clinical deterioration (Ilgen and Moos, 2005), whereas others suggest that depression may be linked with better alcohol-related outcomes (Charney et al., 1998) or better outcomes for women (Rounsaville et al., 1987). Rounsaville et al. (1987) found that men with lifetime major depression had more alcohol-related problems and consumed more alcohol per day at 12 months following inpatient alcohol treatment compared with those without additional psychiatric diagnoses, whereas women with depression had better posttreatment outcomes. Depression has also been associated with better short-term alcohol outcomes in studies of both those referred for evaluation following a driving-while-intoxicated conviction (Windle and Miller, 1990) and alcohol treatment-seeking samples (Schutte et al., 1997), prompting some to suggest that depressive symptoms may function to increase recognition of the need to change or to enter treatment for certain populations (Schutte et al., 1997; Wu et al., 1999).

Although the above community and alcohol-treatment studies have not produced results to clearly predict outcomes among hospitalized unhealthy drinkers, they do suggest that the association between depressive symptoms and alcohol outcome and the magnitude of this association for men and women may vary with the indices of alcohol involvement (Graham et al., 2007; Schutte et al., 1997) and patient populations examined (Crum et al., 2001). Currently, there is relatively little known about whether depressive symptoms predict subsequent alcohol use over time among medical inpatients. A better understanding of whether depressive symptoms have prognostic significance for unhealthy drinking among medical patients may be an important first step in the identification of risk factors and the development of interventions for this population. We sought to examine the association between depressive symptoms and subsequent drinking outcomes among unhealthy drinkers on an inpatient medical service. Given the association between depression and alcohol use among medical patients and findings from prospective studies with samples not specifically involved in alcohol treatment, we hypothesized that those with higher levels of depressive symptoms during hospitalization would have fewer abstinent days, consume more drinks per day, and experience more alcohol-related drinking problems during the subsequent year, controlling for corresponding baseline alcohol outcome variable. In addition, we explored the utility of different screening thresholds that have been used to identify clinically significant depressive symptoms among this population using the Center for Epidemiologic Studies-Depression Scale (CES-D; Schulberg et al., 1985; Zich et al., 1990).

Method

Participant selection and recruitment

Participants (N = 5,813) were screened for unhealthy drinking from the inpatient medicine service at an urban

hospital. Of the 986 that screened positive for unhealthy drinking, 341 patients were enrolled in a clinical intervention trial of brief intervention for alcohol use and signed an informed consent form before study participation. Inclusion criteria were an age of 18 or older and either consumption of risky amounts in the past 30 days (>4 drinks on a single occasion or >14 drinks per week for men and >3 drinks on a single occasion or >11 drinks per week for women) (Dawson et al., 2005; Fleming et al., 1997) or AUDIT scores of ≥8 (Saunders et al., 1993). See Saitz et al. (2007) for inclusion and enrollment procedures. The AUDIT is a 10-item self-report measure that provides information about alcohol use, symptoms of alcohol dependence, and alcohol-related problems. It has been extensively validated in medical settings as a screening tool for hazardous and harmful drinking (Conigrave et al., 1995; Saunders et al., 1993). We excluded those who scored below 21 on the Mini Mental State Examination (Folstein et al., 1975), had a specific intention to leave the metropolitan area in the subsequent year, could not provide two contact names, or were not fluent in either Spanish or English. The present study reports results from the 308 patients who provided follow-up data on alcohol use and problems at 3 or 12 months.

Measures

Primary outcomes. The three primary outcomes were drinks per day in the past 30 days, days abstinent in the past 30 days, and alcohol-related problems. The 30-day Timeline Follow Back (TLFB; Sobell and Sobell, 1992) was used to assess drinks per day and days abstinent. This semi-structured interview is used to assess the quantity, frequency, and types of standardized drinks consumed by the participant in the previous 30 days. The TLFB has been extensively validated as a measure of alcohol consumption and has been used to assess drinking patterns in medical patients. Alcohol-related problems were assessed by the Short Inventory of Problems (SIP-2R; Miller et al., 1995). This measure was developed as a short form of the Drinkers Inventory of Consequences (Miller et al., 1995) and provides overall alcohol-related problem and subscale scores. It has been shown to have adequate internal reliability, stability, and validity among medical patients with unhealthy alcohol use (Kenna et al., 2005).

Primary independent variable. The main independent variable was depressive symptoms assessed by the CES-D (Radloff, 1977). This is a 20-item self-report measure that has been extensively validated in community (e.g., Moscato et al., 1997), medical (e.g., Zich et al., 1990), and alcohol/drug-using samples (e.g., Bobo et al., 1998) to assess depressive symptoms and screen for clinical depression. Participants use a 0-3 rating scale to identify the frequency with which they have experienced various symptoms during the previous week. The scale includes four items (e.g., "I was happy") that are reverse scored. CES-D scores of more than 15 are typically used as a screening cutoff for major depression (Radloff, 1977). In the current study, the primary analyses modeled baseline CES-D scores as a continuous variable. Secondary analyses were also performed to assess two potential cutoffs used in previous work with medical patients (Schulberg et al., 1985; Zich et al., 1990) to identify clinically significant elevated depressive symptoms (>15 and >26).

Other assessments. A series of additional instruments were administered to patients at baseline that included measures of demographics, physical symptoms, substance use, and alcohol-use disorders. Age, gender, employment status (i.e., typical employment pattern in the past 3 months), and homelessness (i.e., any nights in past 3 months spent on the streets or overnight shelter) were identified through interview questions (Kertesz et al., 2003). To address physical symptoms, the Short Form 12 Health Survey (SF-12; Ware et al., 1998) was used. In the current study, the SF-12 Physical Component Summary (PCS) scale was used to assess impairment in physical health-related quality of life. Cocaine and other substance use was assessed using the Addiction Severity Index (McLellan et al., 1992). Current alcohol-use disorders were identified with the Composite International Diagnostic Interview Alcohol Module (Robins et al., 1989; World Health Organization, 1996). This interview-administered instrument was used to determine current (past-year) diagnosis of alcohol abuse and dependence according to criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (American Psychiatric Association, 1994; Andrews and Peters, 1998; Cottler et al., 1997).

Procedures

Participants who enrolled in the study completed assessments at baseline, 3 months, and 12 months (Saitz et al., 2007). The full assessment battery took approximately 2 hours to complete and was conducted in the patient's hospital room. At baseline, participants were randomly assigned either to receive a brief (20-30 minute) motivational intervention following the assessment battery or to complete the baseline assessment battery only. Participants were contacted by phone before each follow-up time point and were scheduled for a face-to-face interview.

Statistical analyses

We used descriptive statistics to characterize the study sample at enrollment. Preliminary models were fit for 3- and 12-month outcomes separately, and associations were generally consistent over time. Thus, the primary analyses incorporated both the 3- and 12-month outcomes into a single model using longitudinal generalized estimating

equations (GEE) regression methods. Separate multivariable GEE Poisson regression models were fit to assess the associations between depressive symptoms and each of the three alcohol-related outcomes (i.e., number of drinks per day, days abstinent, and alcohol problems). The longitudinal regression models account for the correlation from using repeated observations for the same subject. The GEE Poisson models were fit using an unstructured working correlation matrix and a log link. The empirical standard errors from the GEE approach were used for all analyses. The three outcomes were modeled as count data and the Pearson chi-square correction was used to account for overdispersion in the data. Poisson regression models were used for the analyses because the outcomes either were count data or nonnegative variables that did not follow a normal distribution (McCullagh and Nelder, 1989). Incidence rate ratios (IRRs) presented from the Poisson regression models represent the relative increase in outcomes associated with changes in each independent variable. For the main independent variable CES-D, the IRRs presented in Table 3 correspond to the relative increase in the alcohol-related outcome associated with a 10-unit increase in CES-D. All models were stratified by gender, as the effects of depressive symptoms on drinking outcomes have been shown to differ by gender in prior research (Aneshensel and Huba, 1983; Crum et al., 2001; Repetto et al., 2004). Depressive symptoms were assessed at study entry, and these CES-D scores were used to predict alcohol outcomes at the 3- and 12-month follow-up visits. Potential confounding factors that have been previously shown to be related to depressive symptoms and alcohol use (e.g., age and physical health-related quality of life) were included as covariates in each of the regression models. Additional covariates included the baseline value of the alcohol-related outcome, cocaine use, randomization group, and postbaseline assessment time point. As a number of investigators have suggested that low socioeconomic status may account for the relation between depressive symptoms and alcohol use (Jones-Webb et al., 1996; Kim et al., 2003), we also included baseline employment status and homelessness (Kertesz et al., 2003) as covariates in regression models. Secondary analyses were conducted to assess two clinical cutoff scores for substantial depressive symptoms at study entry (CES-D > 15 vs ≤ 15 and CES-D > 26 vs ≤ 26). All analyses were conducted with two-sided tests and a significance level of .05. Analyses were performed using SAS software (Version 9.1; SAS Institute, Cary, NC).

Results

Descriptive statistics

Of the original 341 patients enrolled in the trial, 308 (90%) provided follow-up data on alcohol use and problems at the 3- or 12-month outcome points. This final sample of patients was predominantly male (71%) and had a mean (SD) age of 44.5 (10.5). As reported elsewhere (Saitz et al., 2007), the most common principal medical diagnoses were as follows: rule out myocardial infarction, asthma/bronchitis/chronic obstructive pulmonary disease, pancreatitis, and cellulitis. Table 1 presents baseline characteristics of patients by gender. Participants in the study exhibited high levels of alcohol involvement as demonstrated by alcohol use, alcohol-related problems (SIP-2R scores), and the proportion who were alcohol dependent. Among all patients, the median number of drinks per day was 3.6, and the median of the maximum number of drinks on a single drinking occasion in the past month was 14 drinks.

In Table 2, we present the frequency and proportion of patients who met standard (>15) clinical cutoff scores for elevated depressive symptoms (Radloff, 1977) and a higher clinical cutoff score (>26) by gender and alcohol dependence. A large proportion of patients in the present sample met screening thresholds of >15 and >26. Indeed, among alcohol-dependent women, 90% of the patients had CES-D scores of more than 15. For both men and women, clinically significant elevated depressive symptoms were associated with alcohol dependence.

Alcohol use and problems

IRRs for the association between depressive symptoms and each of the three alcohol outcomes are presented by

TABLE 1. Baseline characteristics of medical inpatient sample with unhealthy alcohol use

Variable	Men (n = 220)	Women (n = 88)
Age, mean (SD)	45.5 (10.5)	41.8 (10)
Hispanic, no. (%)	18 (8.2%)	8 (9.1%)
Race, no. (%)		
Black	93 (42.3%)	49 (55.7%)
White	93 (42.3%)	25 (28.4%)
Unemployed, past 3 months, no. (%)	134 (60.9%)	68 (77.3%)
Homeless, ≥1 night(s) past 3 months, no. (%)	57 (25.9%)	24 (27.3%)
Cocaine use, past 30 days, no. (%)	38 (17.3%)	26 (29.5%)
Physical health-related quality of life, mean (SD)[a]	38.3 (9.5)	37.1 (8.2)
Depressive symptoms, CES-D, mean (SD)	23.7 (13.8)	28.9 (12.3)
Alcohol involvement		
Drinks per day, median (IQR)[b]	3.9 (1.5-9.3)	2.8 (0.9-6.6)
Drinks per day, mean (SD)	7.2 (9.3)	6.3 (9.3)
Days abstinent, median (IQR)[b]	16 (3-24)	17 (6-26)
Days abstinent, mean (SD)	14 (10.6)	15.7 (10)
Alcohol problems, SIP-2R, median (IQR)[c]	13 (4-26)	15.5 (6-28)
Alcohol problems, SIP-2R, mean (SD)	15.6 (13.1)	17.7 (13.5)
Current alcohol dependence, no. (%)[d]	170 (77.3%)	69 (78.4%)

Notes: CES-D = Center for Epidemiologic Studies-Depression Scale; IQR = interquartile range (25th and 75th percentiles); SIP-2R = Short Inventory of Problems. [a]Physical component summary scale score from the Short Form-12 Health Survey; [b]drinks per day and days abstinent from the 30-day Timeline Followback; [c]alcohol problems = SIP-2R scores, past 3 months; [d]current alcohol dependence (12 months) as determined by the Composite International Diagnostic Interview-Current Alcohol Module.

TABLE 2. Patients meeting CES-D (depressive symptom) screening criteria of >15 and >26

	CES-D > 15					CES-D > 26				
	Dependent		Nondep.			Dependent		Nondep.		
Gender	No./n^a	%	No./n	%	p	No./n	%	No./n	%	p
Women	61/69	90	11/19	58	.003[b]	47/69	69	6/19	32	.003
Men	132/170	78	16/50	32	<.001	91/170	54	6/50	12	<.001

Notes: Chi-square/Fischer's Exact Tests show that alcohol dependent men were more likely to score above both clinical thresholds than nondependent men. These effects were also observed for women. CES-D = Center for Epidemiologic Studies-Depression scale; nondep. = nondependent. [a]No. = number screening positive, n = total cell size; [b]Fischer's Exact Test used.

gender in Table 3. The IRRs represent the rate of increase in a given drinking variable corresponding to a 10-unit increase in baseline CES-D scores. Among men, depressive symptoms were associated with intensity of alcohol use and alcohol problems. A 10-unit increase in CES-D score was associated with a 1.17-fold increase (95% confidence interval [CI]: 1.06-1.30) in drinks per day at follow-up and a 1.22-fold increase (95% CI: 1.10-1.35) in alcohol problems. Depressive symptoms were not associated with number of abstinent days among men. No significant associations were detected among women.

Secondary regression analyses were conducted using CES-D scores (>15 and >26) that have been used in previous work to screen for clinical depression in medical settings. Results of these analyses were consistent with those observed with the continuous measures. IRRs for each of these screening score thresholds are presented in Table 4. Regardless of the index chosen, depressive symptoms in the hospital remained a predictor of the number of drinks per day and alcohol-related problems over time for male patients.

TABLE 3. The association between baseline depressive symptoms (CES-D score) and alcohol outcomes following hospitalization

Variable	IRR[a] (95% CI)	p
Drinks per day in past month		
Men	1.17 (1.06-1.30)	<.01
Women	1.00 (0.80-1.24)	.98
Abstinent days in past month		
Men	1.02 (0.97-1.07)	.45
Women	1.02 (0.95-1.10)	.62
Alcohol problems in past 3 months[b]		
Men	1.22 (1.10-1.35)	<.001
Women	1.05 (0.94-1.19)	.39

Notes: Longitudinal Poisson regression models adjusting for baseline alcohol involvement, age, physical symptoms, cocaine use, randomization group, homelessness, employment, and time. Models reflect associations over time between depressive symptoms and alcohol outcomes at 3 and 12 months. CES-D = Center for Epidemiologic Studies-Depression scale; IRR = incidence rate ratio; CI = confidence interval. [a]IRR corresponding to a 10-unit increase in baseline CES-D score; [b]alcohol problem scores from the Short Inventory of Problems.

TABLE 4. The association between alternative baseline CES-D cutoff scores and alcohol outcomes following hospitalization

Variable	CES-D (>15) IRR[a] (95% CI)	CES-D (>26) IRR[a] (95% CI)
Drinks per day in past month		
Men	1.70 (1.19-2.44)	1.71 (1.26-2.33)
Women	0.89 (0.41-1.93)	1.34 (0.66-2.73)
Abstinent days in past month		
Men	1.09 (0.96-1.24)	0.97 (0.84-1.13)
Women	0.99 (0.83-1.20)	1.05 (0.85-1.29)
Alcohol problems in past 3 months[b]		
Men	2.15 (1.61-2.87)	1.69 (1.26-2.27)
Women	1.15 (0.77-1.72)	0.93 (0.71-1.20)

Notes: Longitudinal Poisson regression models adjusting for baseline alcohol involvement, age, physical symptoms, cocaine use, randomization group, homelessness, employment, and time. Models reflect associations over time between depressive symptoms and alcohol outcomes at 3 and 12 months. CES-D = Center for Epidemiologic Studies-Depression scale; IRR = incidence rate ratio; CI = confidence interval. [a]IRR corresponding to the presence of elevated depressive symptoms as defined by specified CES-D cutoff score; [b]alcohol problem scores from the Short Inventory of Problems.

Discussion

Despite evidence for the co-occurrence of depressive symptoms and unhealthy alcohol use among medical patients, little previous research has explored the association between depression and subsequent alcohol outcomes among this population. The current study suggests that higher levels of depressive symptoms may predict worse alcohol outcomes among male medical inpatients with unhealthy alcohol use. Depressive symptoms experienced by these patients in the hospital were associated with more drinks per day and alcohol-related problems over time even when controlling for corresponding baseline alcohol variables and potential confounders such as age, physical symptoms, cocaine use, employment, and homelessness. Depressive symptoms did not appear to be prospectively associated with number of days abstinent, however, for either men or women.

The predictive value of the CES-D was similar whether depressive symptoms were measured as a continuous or categorical variable. As expected, patients in this sample had high CES-D scores, with more than 65% of the male

sample and 80% of the female sample scoring more than the screening cutoff of 15. Because depressive symptoms are common for those who may be undergoing withdrawal from alcohol (Anthenelli and Schuckit, 1993) or experiencing physical distress from a medical condition (Clark et al., 1998; Schulberg et al., 1985; Zich et al., 1990), there has been some debate regarding the value of different screening cutoff scores to assess clinically significant depressive symptoms among medical patients (Fechner-Bates et al., 1994; Schulberg et al., 1987; Zich et al., 1990). In the current study, baseline depressive symptoms as measured by continuous or different CES-D cutoff scores were similarly associated with drinks per day and alcohol-related problems. The results do not address the question of whether different cutoff scores are most appropriate for identifying clinical depression in this sample but rather suggest that depressive symptoms have important implications for the development of alcohol-use patterns among male patients with unhealthy alcohol use regardless of whether these symptoms are measured as continuous or categorical variables using screening threshold scores.

The results of this study add to the body of research with urban medical populations (Crum et al., 2001; Jones-Webb et al., 1996; Kim et al., 2003) that have shown that depressive symptoms are associated with alcohol-use patterns among men. Although the expected association between depressive symptoms and alcohol quantity was observed (Graham et al., 2007), the strongest association observed between depressive symptoms during hospitalization and alcohol outcomes was observed with the alcohol-related problems. This is consistent with studies that have shown that a range of indices of distress are directly associated with alcohol problems independent of the effects of actual consumption (see Gaher et al., 2006).

The link between depression and alcohol use and problems over time may be mediated by a number of factors, including reciprocal direct and indirect effects over time and shared underlying causal factors (Ramsey et al., 1997; Swendsen and Merikangas, 2000). We explored the possibility that the association between depression and alcohol outcomes may be a function of economic stressors. A large proportion of the patients was unemployed, and many had experienced some degree of homelessness during the preceding 3 months. Within our sample, however, socioeconomic status, as defined by homelessness and employment, did not account for the observed associations in this study. However, other sources of distress associated with urban poverty such as poor housing, crime, and a lack of opportunity could exacerbate alcohol use over time and impair recovery efforts (Jones-Webb et al., 1996; Kim et al., 2003). These other potential shared causal factors may be important to examine in future work to better understand the nature of the depressive symptom-alcohol association. In addition to shared causal factors, future efforts to identify mechanisms underlying the link between depression and alcohol use and problems among men may benefit from assessment of the distinct functions of alcohol use among those with elevated depressive symptoms such as coping with negative affect (Cooper et al., 2000; Holahan et al., 2003) or providing sources of positive reinforcement (Vuchinich and Tucker, 1988).

Given the findings suggesting that the association between alcohol-use disorders and depression may be higher among women (Kessler et al., 1997), it might be expected that depressive symptoms would also influence alcohol use and problems among women in this sample. There may be a number of factors that might account for the absence of effects among women. Depressive symptoms may influence decisions about changing alcohol use (Schutte et al., 1997) differently for women in this population. Although depressive symptoms may increase motivation to use alcohol to cope (Holahan et al., 2004), they may also increase alcohol-related change processes associated with treatment engagement (Finney and Moos, 1995). These countervailing forces may both influence alcohol use and problems among those with depressive symptoms (Schutte et al., 1997) and do so differently for different populations. Alternatively, the relatively high levels of depressive symptoms for women in this sample of predominantly alcohol-dependent individuals may in part account for their lower value for predicting alcohol use and problems (Sung et al., 2004). Although these effects were observed for men only, it is important to point out that the current stratified analyses demonstrate considerable overlap in confidence intervals of the IRRs for women and men. As the sample size for women was limited, the study may have been underpowered to detect differences within this group. Thus, the lack of a significant association between depressive symptoms and alcohol outcomes for women in the current study warrants future research.

Limitations

It is important to consider limitations of the present study for interpretation and generalizability. First, the present sample was urban, predominantly minority (38% white), poor, and markedly alcohol involved. Although the study enrollment procedures were designed to identify patients who exhibited a range of unhealthy alcohol use, the majority of patients was alcohol dependent. Although this sample may accurately reflect the patient sample in this type of context (Saitz et al., 2006), it will be important to explore the association between depression and alcohol use among patients with different levels of socioeconomic status, environment (e.g., nonurban), medical services, and levels of alcohol involvement in future studies. Another limitation is that this study was conducted in the context of a brief intervention trial (Saitz et al., 2007). Although the effects of

intervention were controlled in the analyses, these findings may not generalize to patients who are not participating in a randomized intervention trial. Finally, the current study used self-report measures of depressive symptoms and alcohol use that required patient recall. Although we used procedures to minimize these sources of bias (Babor et al., 1990), the effects of social desirability and memory must be considered in the context of this sample.

Summary

In sum, results of this study demonstrate that depressive symptoms predict drinks per day and alcohol-related problems over time among male medical inpatients with unhealthy alcohol use. Although the psychological mechanisms that underlie this association remain unspecified, these findings suggest that depressive symptoms may be important factors to consider in the long term clinical management of unhealthy drinking men who are medical inpatients. Better understanding of the mechanisms that underlie the association between depressive symptoms and unhealthy alcohol use may help refine hospital-based intervention strategies for this population.

References

Alati, R., Kinner, S., Najman, J.M., Fowler, G., Watt, K., and Green, D. Gender differences in the relationships between alcohol, tobacco, and mental health in patients attending an emergency department. Alcohol Alcsm **39:** 463-469, 2004.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), Washington, DC, 1994.

Amodei, N., Elkin, B.B., Burge, S.K., Rodriguez-Andrew, S., Lane, P., and Seale, J.P. Psychiatric problems experienced by primary care patients who misuse alcohol. Int. J. Addict. **29:** 609-626, 1994.

Andrews, G. and Peters, L. The psychometric properties of the Composite International Diagnostic Interview. Social Psychiat. Psychiat. Epidemiol. **33:** 80-88, 1998.

Aneshensel, C.S., Frerichs, R.R., and Huba, G.J. Depression and physical illness: A multiwave nonrecursive causal model. J. Hlth Social Behav. **25:** 350-371, 1984.

Aneshensel, C.S. and Huba, G.J. Depression, alcohol use, and smoking over one year: A four-wave longitudinal causal model. J. Abnorm. Psychol. **92:** 134-150, 1983.

Anthenelli, R.M. and Schuckit, M.A. Affective and anxiety disorders and alcohol and drug dependence: Diagnosis and treatment. J. Addict. Dis. **12:** 73-87, 1993.

Babor, T.F., Brown, J., and Del Boca, F.K. Validity of self-report in applied research on addictive behavior: Fact or fiction? Behav. Assess. **12:** 5-31, 1990.

Bobo, J.K., McIlvain, H.E., and Leed-Kelly, A. Depression screening scores during residential drug treatment and risk of drug use after discharge. Psychiat. Serv. **49:** 693-695, 1998.

Booth, B.M., Blow, F.C., and Cook, C.A. Functional impairment and co-occurring psychiatric disorders in medically hospitalized men. Arch. Intern. Med. **158:** 1551-1559, 1998.

Caldwell, T.M., Rodgers, B., Jorm, A.F., Christensen, H., Jacomb, P.A., Korten, A.E., and Lynskey, M.T. Patterns of association between alcohol consumption and symptoms of depression and anxiety in young adults. Addiction **97:** 583-594, 2002.

Carpenter, K.M. and Hasin, D.S. Drinking to cope with negative affect and DSM-IV alcohol use disorders: A test of three alternative explanations. J. Stud. Alcohol **60:** 694-704, 1999.

Charney, D.A., Paraherakis, A.M., Negrete, J.C., and Grill, K.J. The impact of depression on the outcome of addictions treatment. J. Subst. Abuse Treat. **15:** 123-130, 1998.

Chou, S.P., Grant, B.F., and Dawson, D.A. Medical consequences of alcohol consumption—United States, 1992. Alcsm Clin. Exp. Res. **20:** 1423-1429, 1996.

Clark, D.A., Cook, A., and Snow, D. Depressive symptom differences in hospitalized, medically ill, depressed psychiatric inpatients and nonmedical controls. J. Abnorm. Psychol. **107:** 38-48, 1998.

Conigrave, K.M., Hall, W.D., and Saunders, J.B. The AUDIT Questionnaire: Choosing a cut-off score. Addiction **90:** 1349-1356, 1995.

Cooper, M.L., Agocha, V.B., and Sheldon, M.S. A motivational perspective on risky behaviors: The role of personality and affect regulatory processes. J. Pers. **68:** 1059-1088, 2000.

Cottler, L.B., Grant, B.F., Blaine, J., Mavreas, V., Pull, C., Hasin, D., Compton, W.M., Rubio-Stipec, M., and Mager, D. Concordance of DSM-IV alcohol and drug use disorder criteria and diagnoses measured by AUDADIS-ADR, CIDI, and SCAN. Drug Alcohol Depend. **47:** 185-205, 1997.

Crum, R.M., Brown, C., Liang, K.Y., and Eaton, W.W. The association of depression and problem drinking: Analyses from the Baltimore ECA follow-up study. Addict. Behav. **26:** 765-773, 2001.

Curtis, J.L., Millman, E.J., Joseph, M., Charles, J., and Bajwa, W.K. Prevalence rates for alcoholism, associated depression and dementia on the Harlem Hospital Medicine and Surgery Services. Adv. Alcohol Subst. Abuse **6:** 45-64, 1986.

Dawson, D.A., Grant, B.F., Stinson, F.S., and Chou, P.S. Psychopathology associated with drinking and alcohol use disorders in the college and general adult populations. Drug Alcohol Depend. **77:** 139-150, 2005.

Fechner-Bates, S., Coyne, J.C., and Schwenk, T.L. The relationship of self-reported distress to depressive disorders and other psychopathology. J. Cons. Clin. Psychol. **62:** 550-559, 1994.

Finney, J.W. and Moos, R.H. Entering treatment for alcohol abuse: A stress and coping model. Addiction **90:** 1223-1240, 1995.

Fleming, M.F., Barry, K.L., Manwell, L.B., Johnson, K., and London, R. Brief physician advice for problem alcohol drinkers: A randomized controlled trial in community-based primary care practices. JAMA **277:** 1039-1045, 1997.

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.

Gaher, R.M., Simons, J.S., Jacobs, G.A., Meyer, D., and Johnson-Jiminez, E. Coping motives and trait negative affect: Testing mediation and moderation models of alcohol problems among American Red Cross disaster workers who responded to the September 11, 2001 terror attacks. Addict. Behav. **31:** 1319-1330, 2006.

Graham, K., Massak, A., Demers, A., and Rehm, J. Does the association between alcohol consumption and depression depend on how they are measured? Alcsm Clin. Exp. Res. **31:** 78-88, 2007.

Grant, B.F. and Harford, T.C. Comorbidity between DSM-IV alcohol use disorders and major depression: Results of a national survey. Drug Alcohol Depend. **39:** 197-206, 1995.

Greenfield, S.F., Weiss, R.D., Muenz, L.R., Vagge, L.M., Kelly, J.F., Bello, L.R., and Michael, J. The effect of depression on return to drinking: A prospective study. Arch. Gen. Psychiat. **55:** 259-265, 1998.

Greenfield, T.K., Rehm, J., and Rogers, J.D. Effects of depression and social integration on the relationship between alcohol consumption and all-cause mortality. Addiction **97:** 29-38, 2002.

Hartka, E., Johnstone, B., Leino, E.V., Motoyoshi, M., Temple, M.T., and Fillmore, K.M. A meta-analysis of depressive symptomatology and alcohol consumption over time. Brit. J. Addict. **86:** 1283-1298, 1991.

HESSELBROCK, M.N. AND HESSELBROCK, V.M. Gender, alcoholism, and psychiatric comorbidity. In: WILSNACK, R.W AND WISLNACK, S.C. (Eds.) Gender and Alcohol: Individual Differences and Social Perspectives, New Brunswick, NJ: Rutgers Center of Alcohol Studies, 1997, pp. 49-71.

HOLAHAN, C.J., MOOS, R.H., AND HOLAHAN, C.K. Drinking to cope and alcohol use and abuse in unipolar depression: A 10-year model. J. Abnorm. Psychol. **112:** 159-165, 2003.

HOLAHAN, C.J., MOOS, R.H., HOLAHAN, C.K., CRONKITE, R.C., AND RANDALL, P.K. Unipolar depression, life context vulnerabilities, and drinking to cope. J. Cons. Clin. Psychol. **72:** 269-275, 2004.

ILGEN, M. AND MOOS, R. Deterioration following alcohol-use disorder treatment in project MATCH. J. Stud. Alcohol **66:** 517-525, 2005.

JARMAN, C.M. AND KELLETT, J.M. Alcoholism in the general hospital. Brit. Med. J. **25:** 469-472, 1979.

JONES-WEBB, R., JACOBS, D.R., JR., FLACK, J.M., AND LIU, K. Relationships between depressive symptoms, anxiety, alcohol consumption, and blood pressure: Results from the CARDIA study. Alcsm Clin. Exp. Res. **20:** 420-427, 1996.

KENNA, G.A., LONGABAUGH, R., GOGINENI, A., WOOLARD, R.H., NIRENBERG, T.D., BECKER, B., MINUGH, P.A., CARTY, K., CLIFFORD, P.R., AND KAROLCZUK, K. Can the Short Index of Problems (SIP) be improved? Validity, reliability of the three-month SIP in the emergency department sample. J. Stud. Alcohol **66:** 433-437, 2005.

KERTESZ, S.G., HORTON, N.J., FRIEDMANN, P.D., SAITZ, R., AND SAMET, J.H. Slowing the revolving door: Stabilization programs reduce homeless persons' substance use after detoxification. J. Subst. Abuse Treat. **24:** 197-207, 2003.

KESSLER, R.C., CRUM, R.M., WARNER, L.A., NELSON, C.B., SCHULENBERG, J., AND ANTHONY, J.C. Lifetime co-occurrence of DSM-II-R alcohol abuse and dependence with other psychiatric disorders in the National Comorbidity Survey. Arch. Gen. Psychiat. **54:** 313-321, 1997.

KIM, M.T., HAN, H.R., HILL, M.N., ROSE, L., AND ROARY, M. Depression, substance use, adherence behaviors, and blood pressure in urban hypertensive black men. Ann. Behav. Med. **26:** 24-31, 2003.

LI, T.K., HEWITT, B.G., AND GRANT, B.F. Alcohol use disorders and mood disorders: A National Institute on Alcohol Abuse and Alcoholism perspective. Biol. Psychiat. **56:** 718-720, 2004.

MCCULLAGH, P. AND NELDER, J.A. Generalized Linear Models, 2nd Edition, New York: Chapman and Hall, 1989.

MCLELLAN, A.T., KUSHNER, H., METZGER, D., PETERS, R., SMITH, I., GRISSOM, G., AND ARGERIOU, M. The Fifth Edition of the Addiction Severity Index. J. Subst. Abuse Treat. **9:** 199-213, 1992.

MAISTO, S.A. AND SAITZ, R. Alcohol use disorders: Screening and diagnosis. Amer. J. Addict. **12:** S12-S25, 2003.

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, 1995.

MOORE, R.D., BONE, L.R., GELLER, G., MAMON, J.A., STOKES, E.J., AND LEVINE, D.M. Prevalence, detection, and treatment of alcoholism in hospitalized patients. JAMA **261:** 403-407, 1989.

MOSCATO, B.S., RUSSELL, M., ZIELEZNY, M., BROMET, E., EGRI, G., MUDAR, P., AND MARSHALL, J.R. Gender differences in the relation between depressive symptoms and alcohol problems: A longitudinal perspective. Amer. J. Epidemiol. **146:** 966-974, 1997.

RADLOFF, L.S. The CES-D Scale: A self-report depression scale for research in the general population. Appl. Psychol. Meas. **1:** 385-401, 1977.

RAIMO, E.B. AND SCHUCKIT, M.A. Alcohol dependence and mood disorders. Addict. Behav. **23:** 933-946, 1998.

RAMSEY, S.E., ENGLER, P.A., AND STEIN, M.D. Alcohol use among depressed patients: The need for assessment and intervention. Profess. Psychol. Res. Pract. **36:** 203-207, 2005.

REPETTO, P.B., ZIMMERMAN, M.A., AND CALDWELL, C.H. A longitudinal study of the relationship between depressive symptoms and alcohol use in a sample of inner-city black youth. J. Stud. Alcohol **65:** 169-178, 2004.

ROBINS, L.N., WING, J., WITTCHEN, H.-U., HELZER, J.E., BABOR, T.F., BURKE, J., FARMER, A., JABLENSKI, A., PICKENS, R., REIGER, D.A., SARTORIUS, N., AND TOWLE, L.H. The Composite International Diagnostic Interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch. Gen. Psychiat. **45:** 1069-1077, 1988.

ROELOFFS, C.A., WELLS, K.B., ZIEDONIS, D., TANG, L., AND UNUTZER, J. Problem substance use among depressed patients in managed primary care. Psychosomatics **43:** 405-412, 2002.

ROUNSAVILLE, B.J., DOLINSKY, Z.S., BABOR, T.F., AND MEYER, R.E. Psychopathology as a predictor of treatment outcome in alcoholics. Arch. Gen. Psychiat. **44:** 505-513, 1987.

SAITZ, R., Unhealthy alcohol use. New Eng. J. Med. **352:** 596-607, 2005.

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 patients with unhealthy alcohol use: A randomized controlled trial. Ann. Intern. Med. **146:** 167-176, 2007.

SAUNDERS, J.B., AASLAND, O.G., BABOR, T.F., DE LA FUENTE, J.R., AND GRANT, M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Use Consumption-II. Addiction **88:** 791-804, 1993.

SCHULBERG, H.C., MCCLELLAND, M., AND BURNS, B.J. Depression and physical illness: The prevalence, causation, and diagnosis of comorbidity. Clin. Psychol. Rev. 7: 145- 167, 1987.

SCHULBERG, H.C., SAUL, M., MCCLELLAND, M., GANGULI, M., CHRISTY, W., AND FRANK, R. Assessing depression in primary medical and psychiatric practices. Arch. Gen. Psychiat. **42:** 1164-1170, 1985.

SCHUTTE, K.K., HEARST, J., AND MOOS, R.H. Gender differences in the relations between depressive symptoms and drinking behavior among problem drinkers: A three-wave study. J. Cons. Clin. Psychol. **65:** 392-404, 1997.

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. 41-69.

SUNG, M., ERKANLI, A., ANGOLD, A., AND COSTELLO, E.J. Effects of age at first substance use and psychiatric comorbidity on the development of substance use disorders. Drug Alcohol Depend. **75:** 287-299, 2004.

SWENDSEN, J.D. AND MERIKANGAS, K.R. The comorbidity of depression and substance use disorders. Clin. Psychol. Rev. **20:** 173-189, 2000.

VUCHINICH, R.E. AND TUCKER, J.A. Contributions from behavioral theories of choice to an analysis of alcohol abuse. J. Abnorm. Psychol. **97:** 181-195, 1988.

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: Quality Metric, 1998.

WINDLE, M. AND MILLER, B.A. Problem drinking and depression among DWI offenders: A three-wave longitudinal study. J. Cons. Clin. Psychol. **58:** 166-174, 1990.

WORLD HEALTH ORGANIZATION. Composite International Diagnostic Interview: Core Version 2.0, Geneva, Switzerland: World Health Organization, 1996.

WU, L.-T., KOUZIS, A.C., AND LEAF, P.J. Influence of comorbid alcohol and psychiatric disorders on utilization of mental health services in the National Comorbidity Study. Amer. J. Psychiat. **156:** 1230-1236, 1999.

ZICH, J.M., ATTKISSON, C.C., AND GREENFIELD, T.K. Screening for depression in primary care clinics: The CES-D and the BDI. Int. J. Psychiat. Med. **20:** 259-277, 1990.

The American Journal of Drug and Alcohol Abuse, 33: 169–178, 2007

ISSN: 0095-2990 print/1097-9891 online
DOI: 10.1080/00952990601091176

informa
healthcare

Associations Between Alcohol, Heroin, and Cocaine Use and High Risk Sexual Behaviors Among Detoxification Patients

Anita Raj,[1] Richard Saitz,[2,3,4] Debbie M. Cheng,[5] Michael Winter,[6] and Jeffrey H. Samet[2]

[1]Department of Social Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts, USA
[2]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center, Boston, Massachusetts, USA
[3]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA
[4]Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, Massachusetts, USA
[5]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA
[6]Data Coordinating Center, Boston University School of Public Health, Boston, Massachusetts, USA

Abstract: The purpose of this study was to assess associations between substance use (alcohol to intoxication, heroin, and cocaine) and sexual activity, high risk sexual behaviors, and STD among detoxification inpatients (n = 470). Participants were surveyed on past 30 day substance use, past 6 month sexual behaviors, and STD in the past 6 months and/or over 24 months of follow-up. Logistic regression models adjusted for demographics found that cocaine use was significantly associated with being sexually active ($OR_{adj} = 2.3$, 95% CI = 1.1–4.8) and selling sex ($OR_{adj} = 2.6$, 95% CI = 1.3–5.3). Alcohol and heroin were not significantly associated with sexual activity, high risk sexual behaviors or STD in this sample.

Keywords: Detoxification center, high risk sexual behaviors, STD, substance abuse

Address correspondence to Anita Raj, Ph.D., Boston University School of Public Health, 715 Albany St., T2W, Boston, MA 02118. E-mail: anitaraj@bu.edu

INTRODUCTION

Researchers and practitioners (1–3), the National Institute on Drug Abuse (4), and the Centers for Disease Control and Prevention (5) have called for inclusion of human immunodeficiency virus (HIV) prevention efforts including sexual risk reduction in substance abuse treatment settings. Numerous studies demonstrate the high risk for sexually transmitted diseases (STDs) including HIV among patients in detoxification centers and the efficacy of sexual risk reduction prevention programs in these settings (4, 6–12). Yet, in practice, state-mandated sexual risk reduction prevention efforts do not appear to be associated with reduced sexual risk behaviors (13, 14). Use of a ''one-size-fits-all'' approach is likely inadequate for differential sexual risk practices across substance using groups; this undifferentiated approach may be one reason sexual risk reduction efforts proven efficacious in research studies are not promoting safer sexual behaviors among detoxification patients in practice (4, 14).

Certain substances, specifically alcohol, heroin, and cocaine, have been associated with higher prevalence of STD/HIV (4); these 3 substances also represent the primary addiction of 81% of people receiving substance abuse treatment in the United States (15). Thus, study of the associations between use of these substances and sexual activity, including high risk sexual behaviors and STD among detoxification inpatients may be useful in helping identify patients at increased risk for STD/HIV. However, the high prevalence of polysubstance use in this population (15) requires such analyses to consider the independent contributions of these substances. The purpose of the current study was to assess the independent associations between alcohol to intoxication, heroin, and cocaine use and the outcomes of sexual activity, high risk sexual behaviors, and STD among patients in a detoxification center.

METHODS

Subjects

Study subjects were the 470 detoxification patients enrolled in the Health Evaluation and Linkage to Primary care (HELP) study, a randomized, controlled trial to link detoxification patients to primary medical care services (16). HELP subjects were required to a) be English- or Spanish-speaking ambulatory inpatients at recruitment, b) report alcohol, cocaine, or heroin as their first or second drug of choice, and c) have no primary medical care provider. The study sample was 18–60 years old (median age 35 years), 24% female, 46% African-American and 37%

White; 48% had been in detoxification treatment prior to this entry and within the past 6 months. Baseline behavioral data and baseline and follow-up self-reported STD data from this study were used for the current cross-sectional analyses.

Data Collection Procedures

Eligible patients participated in a 90-minute baseline interviewer administered survey during their inpatient detoxification stay, after 48 hours of arrival; they were again interviewed every 6 months over 24 months. Written informed consent was obtained from each subject prior to their participation. Subjects were compensated modestly for their participation. The Institutional Review Board of Boston Medical Center approved the study and, for additional protection, a certificate of confidentiality was obtained from the National Institute on Alcohol Abuse and Alcoholism.

Measures

Demographics including age, gender, and race were assessed via single survey items.

Past 30 day *alcohol use to intoxication* (defined as having more than 3 drinks a day or "feeling the effects"), *heroin use, and cocaine use*, the 3 main independent variables in this study, were measured using the Addiction Severity Index (ASI) (17). Subjects could report use of more than one of these substances.

High risk sexual behavior variables were measured using items from the Risk Assessment Battery (RAB) (18). These outcomes were assessed for the past 6 months and dichotomized to yield the following variables: *sexually active*, *inconsistent or no condom use*, *multiple sex partners*, *buying sex with money or drugs*, and *selling sex for money or drugs*. Any *self-reported STD* in the past 6 months was assessed at baseline and in each follow-up interview; STDs assessed included syphilis, gonorrhea, chlamydia, genital warts, genital herpes, other STDs (defined as "crabs, Hepatitis B, other, not HIV"), and pelvic inflammatory disease (women only).

Data Analyses

Initial analyses assessed the frequency of each of the substance use variables and each of the sexual risk and STD outcome variables. To assess associations between each substance and the outcome variables, 6

separate logistic regression models, adjusted for demographics, were fit for each outcome. Initial analyses included interactions between the 3 substance abuse variables to assess whether the effect of using a particular substance depended on use of the others. Significant interactions were not observed and the interactions were therefore removed in the final models. Associations were assessed using odds ratios and 95% confidence intervals.

RESULTS

Thirty-Day Prevalence of Alcohol to Intoxication, Heroin, and Cocaine Use

At baseline, 80% of the overall sample (n = 375/469) reported drinking to intoxication; 38% (n = 177/470) reported heroin use; and 65% (n = 304/470) reported cocaine use. Twenty percent of the overall sample (n = 93/469) reported alcohol to intoxication but no heroin or cocaine use; 7% (n = 33/469) reported heroin but no alcohol to intoxication or cocaine; and 6% reported cocaine use but no alcohol to intoxication or heroin. (Note: One of 470 respondents did not answer questions on use of alcohol to intoxication or heroin but did answer the question on cocaine use; thus, denominators differ slightly across these 3 variables).

Polysubstance use was common, with 66% of our sample (n = 309/469) reporting some combination of alcohol to intoxication, heroin, and cocaine in the past 30 days. Among subjects reporting any alcohol use to intoxication, heroin and cocaine were used a mean of 6.4 and 10.8 days, respectively. Among subjects reporting any heroin use, alcohol to intoxication and cocaine were used a mean of 11.2 and 8.7 days, respectively. Among subjects reporting any cocaine use, alcohol use to intoxication and heroin were used a mean of 14.8 and 7.9 days, respectively. Eighteen percent of our total study sample (n = 84/469) used all three substances (i.e., alcohol to intoxication, cocaine, and heroin).

Six-Month Prevalence of Sexual Risk Behaviors and STD

Ninety percent of subjects (422/470) engaged in sex in the past 6 months. These sexually active subjects (n = 422) reported large proportions of high risk sexual behaviors: 55% reported multiple partners; 68% reported inconsistent or no condom use; 19% bought sex for drugs or money; and 17% sold sex for drugs or money. Consistent with this level of high risk sexual activity, 12% reported an STD at some point during the 30-month assessment period. The proportions of subjects reporting

sexual activity, high risk sexual behaviors, and STD were similar for the total sample and across the sub-samples of subjects reporting each substance assessed.

Associations Between Substance Use and Sexual Risk

After controlling for the effects of alcohol to intoxication and heroin, among detoxification patients, cocaine users were significantly more likely than non-cocaine users to report sexual activity in the past 6 months ($OR_{adj} = 2.3$, 95% CI = 1.1–4.8) (Table 1). Among sexually active subjects, cocaine users were also significantly more likely than non-cocaine users to report selling sex ($OR_{adj} = 2.6$, 95% CI = 1.3–5.3). Among sexually active subjects, nonsignificant but notable relationships were also observed between alcohol use to intoxication and sex bought ($OR_{adj} = 1.8$, 95% CI = .8–4.2), cocaine use and sex bought ($OR_{adj} = 1.9$, 95% CI = 1–3.6), and cocaine use and STD ($OR_{adj} = 2$, 95% CI = .8–4.8).

As females are more likely than males to sell sex and males more likely than females to purchase sex (19–21), gender-stratified analyses were conducted to further assess associations between substance use and selling and buying sex. Selling sex was reported by 11% of males (95% CI = 7%–14%; n = 322) and 36% of females (95% CI = 27%–45%; n = 100); buying sex was reported by 23% of males (95% CI = 18%–28%; n = 321) and 5% of females (95% CI = 2%–11%; n = 100). Among sexually active subjects, logistic regression analyses including all substance use variables and controlling for age and race demonstrated that cocaine was significantly associated with selling sex for females ($OR_{adj} = 9.1$, 95% CI = 2.3–36.4) but not males ($OR_{adj} = 1.5$, 95% CI = .6–3.4). Given the small numbers of females reporting sex bought, associations between substance use and sex bought could only be assessed separately for males. Analyses with males demonstrated that cocaine use was significantly related to sex bought ($OR_{adj} = 2.4$, 95% CI = 1.2–4.8); a nonsignificant but notable association between alcohol to intoxication and sex bought was also observed for males ($OR_{adj} = 2$, 95% CI = .8–4.9).

DISCUSSION

In this cohort of detoxification inpatients, a substantial proportion of individuals with alcohol, heroin, or cocaine use was sexually active (90%) and reported high risk sexual behaviors. Over half of sexually active subjects reported multiple sex partners, and more than two-thirds did not use condoms or used them inconsistently. Self-reported STD

Table 1. Association between substance use (past 30 days) and sexual behaviors and STD

Substance use	Sexually active OR_{adj} (95% CI)	Multiple partners* OR_{adj} (95% CI)	Inconsistent condom use* OR_{adj} (95% CI)	Sex bought* OR_{adj} (95% CI)	Sex sold* OR_{adj} (95% CI)	STD** AOR (95% CI)
Alcohol to intoxication	1.2 (.5–2.9)	1.3 (.8–2.1)	.8 (.5–1.5)	1.8 (.8–4.2)	1.2 (.6–2.5)	.9 (.4–2.1)
Heroin	1.2 (.5–2.4)	1.4 (.9–2.1)	.9 (.5–1.4)	.7 (.4–1.3)	1.1 (.6–2)	1 (.5–2.1)
Cocaine	2.3 (1.1–4.8)	1.4 (.9–2.2)	.9 (.5–1.4)	1.9 (1–3.6)	2.6 (1.3–5.3)	2 (.8–4.8)

AOR (95%CI) = Adjusted Odds Ratio and 95% Confidence Interval.
Models include all substance use variables, age, race, and gender.
*Includes only those reporting any sex in the past 6 months (n = 422).
**Includes only those who reported sex in past 6 months at baseline and had at least 1 follow-up interview (n = 349).

within the wider assessment period of up to 30 months was reported by 12% of our sample. Notably, proportions of sexual activity, high risk sexual behaviors, and STD prevalence were similar for subjects reporting alcohol to intoxication, heroin, and cocaine; this may be partly due to the overlap of substances used by our sample. Overall, findings confirm previous research demonstrating high risk sexual behaviors and a substantial proportion with an STD diagnosis over time among detoxification patients (4), confirming the need for sexual risk reduction programs for this population.

While STD/HIV risk is an important issue for all detoxification patients, results from this study do suggest that those reporting recent cocaine use are at even greater risk for sexual activity, sex trade involvement (i.e., buying sex for males and selling sex for females), and contracting an STD, as compared with those not reporting recent cocaine use. These findings are consistent with previous research with detoxification patients documenting greater sex trade and STD diagnosis among cocaine users compared with those reporting solely alcohol use (22, 23). Study findings suggest a need to prioritize sexual risk assessment and intervention, including consideration of the issue of sex trade, for detoxification patients reporting recent cocaine use.

Several limitations of the current study deserve mentioning. Use of cross-sectional analyses prohibits assumptions of causality; however, the associations found do inform us regarding levels of risk in detoxification patients at program entry. Reliance on self-report potentially results in social desirability and recall biases, although recall biases are likely minimal as timeframes for behavior assessment were short (i.e., past 1 to 6 months), and social desirability bias would have led to underestimates rather than overestimates of risk behavior and STD prevalence. Our STD diagnosis variable is not indicative of when the STD was contracted but only indicative of when diagnosis occurred; however, at worst, this is nondifferential misclassification that would not impact differences observed across substances. Finally, as subjects often combined the drugs of interest, pure analysis by drug was not possible; however, drugs are commonly used in combination, making our analyses more relevant in terms of characterizing the reality of patients in detoxification centers.

Additional limitations include use of a single urban detoxification center, limiting generalizablity. The use of a relatively small sample size precluded gender-stratified analyses and may have led to lack of statistically significant findings for some associations, particularly low-prevalence variables like STD. Research with larger samples and multiple diverse detoxification sites could further examine associations identified in the current work. Additionally, as the focus of the current study was

on associations between substance use and sexual risk for STD/HIV, analyses did not include injection drug use behaviors. While this was the appropriate approach for our research question, our current analyses are limited in their assessment of overall risk for STD/HIV among our injection drug-using participants (22% of our sample); additional research is needed to explore sexual and injection drug use related risk for STD/HIV among this subpopulation of patients.

Implications for Practice. In order to achieve more effective STD/HIV prevention among high risk substance abuse treatment patients, more intensive and better-tailored efforts will be needed to promote sexual risk reduction. Resource limitations may require identification of patients at high risk for STD/HIV for these more tailored efforts. Study findings suggest that detoxification patients reporting cocaine use should be prioritized for intervention due to their heightened risk for sex trade involvement and STD as compared with nonusers of cocaine. However, high proportions of polysubstance use and high risk sexual behaviors for the sample as a whole, suggest that assessment of sexual risk should also occur at program entry, to ensure provision of tailored interventions for patients at greater risk for STD/HIV regardless of reported substance used.

ACKNOWLEDGMENTS

We thank the supportive staff members of River Street Detoxification Unit. This research was supported by the National Institute on Drug Abuse (R01-DA10019) and the National Institute on Alcohol Abuse and Alcoholism (R01-AA10870). This research was conducted in part in the General Clinical Research Center at Boston Medical Center, USPHS grant M01 RR00533.

REFERENCES

1. Flanzer J. Health services research: Drug use and human immunodeficiency virus in the United States. Clin Infect Dis 2003; 37(5):439–444.
2. Woods WJ, Lindan CP, Hudes ES, Boscarino JA, Clark WW, Avins AL. HIV infection and risk behaviors in two cross-sectional surveys of heterosexuals in alcohol treatment. J Stud Alcohol 2000; 61(2):262–266.
3. Clapp JD, Burke AC. Supervisor ideology and organizational response: HIV/AIDS prevention in outpatient substance abuse treatment units. Adm Soc Work 1997; 21(1):49–64.
4. Brown B. Health Services Resource: HIV/AIDS and drug abuse treatment services—Parts A and B. Sept 1998. Available at: http://165.112.78.61/

about/organization/DESPR/HSR/da-tre/BrownHIV.html. (Accessed July 18, 2004)
5. CDC. Linking HIV prevention services and substance abuse treatment programs. Feb 2002. Available at: http://www.cdc.gov/idu/facts/HIVPrevServ-Fin.pdf (Accessed April 30, 2004)
6. Semaan S, Des Jarlais DC, Sogolow E, Johnson WD, Hedges LV, Ramirez G, Flores SA, Norman L, Sweat MD, Needle R. A meta-analysis of the effect of HIV prevention interventions on the sex behaviors of drug users in the United States. J Acquir Immune Defic Syndr 2002; 30:73–93.
7. Margolin A, Avants SK, Warburton LA, Hawkins KA, Shi J. A randomized clinical trial of a manual-guided risk reduction intervention for HIV-positive injection drug users. Health Psychol 2003; 22:223–228.
8. Metzger DS, Navaline H, Woody GE. Drug abuse treatment as AIDS prevention. Public Health Rep 1998; 113(1):97–106.
9. Prendergast ML, Podus D, Chang E, Urada D. The effectiveness of drug abuse treatment: A meta-analysis of comparison group studies. Drug Alcohol Depend 2002; 67:53–72.
10. Sorensen JL, Copeland AL. Drug abuse treatment as an HIV prevention strategy: A review. Drug Alcohol Depend 2000; 59:17–31.
11. Avins AL, Lindan CP, Woods WJ, Hudes ES, Boscarino JA, Kay J, Clark W, Hulley SB. Changes in HIV-related behaviors among heterosexual alcoholics following addiction treatment. Drug Alcohol Depend 1997; 44:47–55.
12. Woods WJ, Guydish JR, Sorensen JL, Coutts A, Bostrom A, Acampora A. Changes in HIV-related risk behaviors following drug abuse treatment. AIDS 1999; 13:2151–2155.
13. Massachusetts Department of Public Health (MDPH). Substance Abuse Treatment. Available at: http://www.state.ma.us/dph/bsas/treatment/treatment.htm. (Accessed July 18, 2004)
14. Palepu A, Raj A, Horton NJ, Tibbets N, Meli S, Samet JH. Substance abuse treatment and risk behaviors among HIV-infected persons with alcohol problems. J Subst Abuse Treat 2005; 28(1):3–9.
15. Substance Abuse and Mental Health Services Administration (SAMHSA). Office of Applied Statistics. 2003 National Survey on Drug Use and Health: Results. Available at: http://oas.samhsa.gov/NHSDA/2k3NSDUH/2k3results.htm#ch7 (Accessed November 12, 2004)
16. Samet JH, Larson MJ, Horton NJ, Doyle K, Winter M, Saitz R. Linking alcohol- and drug-dependent adults to primary medical care: A randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. Addiction 2004; 98:509–516.
17. McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grisson G, Pettinati H, Argeriou M. The fifth edition of the addiction severity index. J Subst Abuse Treat 1992; 9:199–213.
18. 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; 10:281–283.

19. Latki CA, Hua W, Forman VL. The relationship between social network characteristics and exchanging sex for drugs or money among drug users in Baltimore, MD, USA. Intl J STD AIDS 2003; 14(11):770–775.
20. Pitts MK, Smith AM, Grierson J, O'Brien M, Misson S. Who pays for sex and why? An analysis of social and motivational factors associated with male clients of sex workers. Arch Sex Behav 2004; 33(4):353–358.
21. Vanwesenbeeck I. Another decade of social scientific work on sex work: a review of research 1990–2000. Ann Rev Sex Res 2001; 12:242–289.
22. Ross L, Kohler CL, Grimley DM, Bellis J. Intention to use condoms among three low-income, urban African-American subgroups: Cocaine users, non-cocaine drug users, and non-drug users. J Urban Health 2003; 80(1):147–160.
23. Windle M. The trading of sex for money or drugs, sexually transmitted diseases (STDs), and HIV-related risk behaviors among multisubstance using alcoholic inpatients. Drug Alcohol Depend 1997; 49(1):33–38.

Exploring Racial and Sociodemographic Trends in Physician Behavior, Physician Trust and Their Association with Blood Pressure Control

Mustafa M. Rawaf, MA and Nancy R. Kressin, PhD

Financial support: This research was supported by grants from the Department of Veterans Affairs (VA) Health Services Research and Development Service (TRH01-038, N. Kressin, P.I). Dr. Kressin is a research career scientist in the Department of Veterans Affairs, Health Services Research & Development (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.

Racial disparities in several facets of healthcare have been widely documented, showing that African Americans face disproportionately high health risks when compared to whites. With respect to hypertension, 40% of the ≥36 million African Americans are affected. We examined the correlation between the patient–physician relationship and the racial disparities in healthcare. We hypothesized that increased physician counseling would lead to higher patient trust and, thus, a greater likelihood of having controlled blood pressure. Four-hundred-sixty black and 333 white Veteran Affairs (VA) patients previously diagnosed with hypertension were included. Patients with a systolic reading ≥140 mmHg and/or a diastolic reading ≥90 mmHg at a recent doctor visit were considered to have uncontrolled blood pressure. By using patient exit interviews (PEIs), we quantified the number of counseling behaviors performed by physicians. Patient trust in physician was measured by validated questions answered on a 1–5 agreement scale. Results showed no racial disparity in blood pressure control. While blacks were found to receive more counseling, whites reported higher trust. Controlling for sociodemographic factors, we found that regardless of race, higher PEI scores were associated with higher trust; however, they were also associated with uncontrolled blood pressure. The association of physician behavior with blood pressure was not mediated by trust. We were unable to make direct cause-and-effect conclusions because the measures were recorded from a one-time questionnaire. Future research should focus on uncovering causal relationships, allowing physicians to work towards ending the established healthcare disparities.

Key words: race/ethnicity ■ health disparities ■ hypertension

 From Boston University School of Medicine (Rawaf, submitted in partial fulfillment of the requirements for the degree of master of arts, 2006); Center for Health Quality, Outcomes and Economic Research (a VA Health Services Research and Development National Center of Excellence), Bedford VA Medical Center, Bedford, MA (Kressin); and Division of General Internal Medicine, Boston University School of Medicine, Boston, MA (Kressin). Send correspondence and reprint requests for *J Natl Med Assoc.* 2007;99:1248–1254 to: Mustafa M. Rawaf, 4113 Ironwood Drive, Chino Hills, CA 91709; phone: (714) 514-4865; e-mail: mustafa.rawaf@gmail.com

INTRODUCTION

Racial disparities in American healthcare have been broadly documented in the medical literature.[1-5] The effects of these disparities have been shown to be far reaching, with deep economic, personal, and social implications. Minorities utilizing the American healthcare system are less likely to receive high-quality care and are thus less likely to have their health conditions appropriately treated, thereby leading to the development of future health problems. This creates a trend of higher healthcare costs not only for the patients involved but the taxpayers or employers who are helping support the healthcare service.[6] At a personal level, healthcare disparities can lead to increased morbidity, disability and lost productivity.[7]

A spotlight was placed on these racial disparities beginning in the late 1980s and early 1990s, prompting several reforms by the government to deal with the emerging issue. Examples of proposed solutions included increasing funding for treatment of minorities, ensuring adequate minority representation in clinical trials and setting up offices of minority health in several states.[5] Additionally, publicized programs were implemented including the Racial and Ethnic Approaches to Community Health (REACH 2010).[8] Launched in 1999, this initiative was designed to eliminate health disparities in six areas: cardiovascular disease, immunizations, breast and cervical cancer screening and management, diabetes, HIV/AIDS and infant mortality. Despite these extensive efforts, it is evident that racial disparities in the American healthcare system continue.[5]

With a population of >36 million, African Americans make up 13–14% of the U.S. population.[9] Several studies have documented the disadvantages they face

when seeking medical attention.[10-17] According to statistics from the American Heart Association, the leading cause of death for African Americans is cardiovascular disease, which kills 36.4% of the ≥290,000 African Americans who die each year. When compared to whites, African Americans are more likely to have high blood pressure, less likely to engage in physical activity, more likely to have diabetes, and more likely to be overweight or obese. All of these factors increase the risk of cardiovascular disease. Forty percent of African Americans are affected with high blood pressure, which ranks among the highest rates of hypertension for blacks in the world.[18,19] The racial disparity surrounding hypertension control has been shown to be a significant contributor to the difference in life expectancy between whites and blacks.[20] Unfortunately, the extensive amount of literature on racial differences in healthcare has yet to lead us to the roots of this ever-growing problem. In order to effectively work towards a healthcare system which has no racial bias, the sources of the disparities should be identified and addressed methodically.

It has been established that the patient–physician relationship affects outcomes of medical encounters.[21,22] Thus, we sought to examine the role of the patient–physician relationship in racial disparities currently plaguing healthcare. In other words, do African Americans and whites experience a visit with their physician differently? If so, does that difference lead to variability in blood pressure control? Two key components of the physician–patient relationship are the physician's behavior and the patient's trust in the physician.[21,23-25] In order to study these issues among a sample where differences in access to healthcare are minimized, we utilized data from the Veteran Affairs (VA) medical system.

Trust in Physician

The concept of trust in one's physician has gained prominence over recent years as a meaningful way to assess patient–physician relationships. Trust is an important factor governing the strength of everyday relationships. Within the healthcare realm, trust has been shown to be a major component of a strong patient–physician relationship.[23-25] Studies have shown that a patient's trust in his/her physician is associated with continuity of care, self-reported adherence to medication, satisfaction and improved health.[25-28] With regard to racial disparities, some studies have shown black patients to hold less trust in their physicians[29-31] than whites, while other studies have shown their trust levels to be at least as high as those of whites.[32-34] As the concept of trust has come to the forefront of the patient–doctor relationship, we felt it important to analyze how trust may play a role in the aforementioned racial disparities in treatment and outcomes of hypertension care.

Physician Behavior

The effectiveness of a physician is influenced by the way he/she behaves in the presence of his/her patient. It has been noted that the manner in which a physician behaves during a patient visit has implications on how much that patient trusts their doctor.[35] High levels of trust are reported when patients feel their physician attempts to understand their experiences, share power, communicate clearly and obtain referrals.[21,36-37] Specific physician characteristics which have been shown to be associated with higher levels of trust include caring, comfort, technical competency and communication.[37]

In summary, the purpose of this study is to examine racial differences in the patient–physician relationship and how they relate to specific experiences with trust in the physician, and ultimately, if they are associated with blood pressure control. We will be assessing the patient–physician relationship via measurements of physician behavior and patient trust taken at one point in time. We believe more hypertension counseling behaviors will lead to higher ratings of trust and, thus, lead to more controlled blood pressure. We hypothesize that African Americans, in comparison to whites, will experience a lower frequency of positive physician behaviors, leading to lower levels of trust in their physician and higher rates of uncontrolled blood pressure.

METHODS

Data were obtained from a study funded by the Department of Veteran Affairs Health Services Research and Development Service (HSR&D) entitled "Improving Hypertension Control: A Physician Intervention." White and African-American patients diagnosed with hypertension were identified at three urban tertiary Department of VA medical centers. Patients with hypertension diagnoses appearing in their records on ≥2 separate visits over a one-year period (ICD9 diagnosis codes: 401: essential hypertension, 401.0: malignant, 401.1: benign, 401.9: unspecified, 405: secondary hypertension, 405.0: malignant, 405.1: benign, 405.11: renovascular, 405.19: other, 405.9: unspecified, 405.91: renovascular, 405.99: other) were eligible to participate. After applying the above-mentioned criteria, there were 11,528 hypertensive patients who were consistent users of primary care services within these three sites in the Department of VA. Research assistants tracked these patients' primary care visits over a 14-month recruitment period, during which 1,210 of them presented to the clinics and were asked to participate in the study. Two-hundred-fifteen patients (18% of 1,210) refused participation, and a total of 204 (17% of 1,210) were excluded due to being non-African American or nonwhite (n=18), having poor mental status (n=41), denying hypertension (n=59), participation in another study regarding hypertension care (n=6), or miscellaneous reasons (n=80). Thus, 793 patients were included in the final cohort, representing 78.6% of the 1,006 eligible patients.

MEASURES

Sociodemographic Characteristics

Participating patients were asked to complete a one-time research assistant-administered interview which included questions about sociodemographic characteristics, including date of birth, highest grade completed in school, employment status, marital status and income.

Experiences with Provider

We adapted a series of questions developed by Ockene et al.[38] to assess the content of the patient–doctor interaction that focused on hypertension. Such patient exit interviews (PEIs) are valid ways to measure the actual content of clinic visits.[39] After the patient was seen by the physician, we conducted a one-time exit interview with him/her, covering the following areas:

1. Did the physician discuss hypertension and blood pressure medications during the visit?
2. Did the physician ask if the patient was taking blood pressure medications as prescribed?
3. Did the physician discuss how important blood pressure meds are for controlling blood pressure?
4. Did the physician discuss other health problems that might develop if the patient doesn't take his/her blood pressure meds?
5. Did the physician advise the patient to take blood pressure meds as prescribed?
6. Did the physician discuss the patient's prior efforts to manage his/her blood pressure meds?
7. Did the physician discuss things that get in the way of the patient taking his/her blood pressure meds?
8. Did the physician discuss things one can do to make it easier to take blood pressure meds?
9. Did the physician discuss any specific goals regarding taking blood pressure meds as prescribed?
10. Did the physician and patient agree on any specific goals for taking blood pressure meds?
11. Did the physician ask the patient to make another appointment to discuss blood pressure?
12. Did the physician provide written materials about hypertension?

Possible responses to these questions were dichotomous (yes/no). Following Ockene,[38] these answers were summed to create one scale score ranging from 1–12 assessing physicians' use of patient-centered counseling strategies.

Trust in Physician

To measure the patients' trust in physician, an 11-item scale adapted from Anderson and Dedrick[40] and validated by Thom et al.[25] was utilized. The 11 items appeared on the same questionnaire administered once following the patient visit. The items appear as the following statements:

1. I doubt that my provider really cares about me as a person.
2. My provider is usually considerate of my needs and puts them first.
3. I trust my provider so much I always try to follow his/her advice.
4. If my provider tells me something is so, it must be true.
5. I sometimes distrust my provider's opinions and would like a second one.
6. I trust my provider's judgments about my medical care.

Table 1. Sociodemographic characteristics with patient exit interview, trust in physician and blood pressure

	Whites	African Americans	P
Number of Participants	333 (42%)	460 (58%)	
Mean Age (SD)	67.33 (±10.39)	64.43 (±10.19)	0.0001
Mean Completed Years of Education (SD)	12.25 (±2.51)	12.12 (±2.33)	0.447
Employment Status			0.146
Employed	45 (13.6%)	81 (17.6%)	
Unemployed	287 (86.4%)	378 (82.4%)	
Annual Income			0.117
<$20k	159 (51.0%)	224 (51.7%)	
≥$20k	146 (46.8%)	187 (43.2%)	
Don't Know	7 (2.2%)	22 (5.1%)	
Marital Status			<0.0001
Married	181 (54.5%)	172 (37.4%)	
Separated, divorced, widowed, single	151 (45.5%)	288 (62.6%)	
PEI Score (SD)	5.81 (±3.70)	6.57 (±3.68)	0.0061
Trust Score (SD)	82.47 (±10.01)	78.19 (±10.54)	<0.0001
Blood Pressure Status			0.4489
Normal	138 (43.4%)	208 (46.4%)	
High	180 (56.6%)	240 (53.6%)	

7. I feel my provider does not do everything he/she should about my medical care.
8. I trust my provider to put my medical needs above all other considerations when treating my medical problems.
9. My provider is well qualified to manage (diagnose and treat or make an appropriate referral) medical problems like mine.
10. I trust my provider to tell me if a mistake was made about my treatment.
11. I sometimes worry that my provider may not keep the information we discuss totally private.

The patients were told to respond to each statement on a five-point scale (1 = totally disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = totally agree). Negative statements (1, 5, 7 and 11) were reverse coded, and the average score for each patient was calculated and converted to a 100-point scale with higher scores representing higher trust in the physician. The end result is a trust in physician (TIP) score for each patient after a particular visit with their provider.

Blood Pressure Control

Each patient's blood pressure was recorded at the time of the office visit taking place on the same day the questionnaire was administered. Thus, it was ensured that the blood pressure readings used in the study were taken after a minimum of 14 months of continuous primary care. Based on the blood pressure reading, each patient was placed in one of two groups: normal/controlled blood pressure or high/uncontrolled blood pressure. Group designations were based on the 7th report of the Joint National Committee on prevention, detection, evaluation and treatment of hypertension (JNC-7). The JNC-7 hypertension guidelines define high blood pressure as a systolic reading of 140 mmHg or higher and/or a diastolic reading of ≥90 mmHg.

ANALYSES

Statistical analyses were conducted using Microsoft® Excel® 2002 modified by Analyse-it®[41] statistical package. Bivariate analyses were conducted using one-way ANOVAs, Pearson correlations and Chi-squared tests. Multivariate results were produced via linear regressions.

RESULTS

Table 1 provides the participant demographic characteristics. Of the 793 participants, 333 (42%) were white and 460 (58%) were African American. Significant differences were found between the white and African-American participants in two categories: age and marital status. With regard to age, whites averaged 67.33 years, while blacks were younger, with an average of 64.43 years (p=0.0001). Analysis of marital status showed that 181 (50.8%) of white participants were married, while only 172 (37.4%) of black participants were married (p<0.0001). Other demographic factors such as annual income, employment status and years of education were not found to be significantly different between the groups.

Trust Score

One-way between subjects ANOVAs, Chi-squared tests, and Pearson correlations were employed to detect significant associations between demographic factors and the TIP score (Table 2). Significantly higher rates of trust were found among white individuals versus blacks (white mean 82.47 versus black mean 78.19, p<0.0001), and the employed versus unemployed (employed mean 82.31 versus unemployed mean 79.52, p=0.0067). Having a higher level of education was shown to have a marginally significant association with higher trust (p=0.056).

PEI Score

Similarly, one-way ANOVAs, Chi-squared tests and Pearson correlations were used to examine demographic

Table 2. Bivariate results with patient exit interview, trust in physician and blood pressure

	Race		Marriage		Employment		Income		Age	Ed.
	Black	White	Marr.	NM	Emp.	UE	<$20k	>$20k	CC	
Trust Score									-0.03	0.07
Mean	78.19	82.47	80.58	79.51	82.31	79.52	79.82	80.96		
(SD)	(10.54)	(10.01)	(10.17)	(10.80)	(10.89)	(10.41)	(10.58)	(10.47)		
P value	<0.001		0.1573		0.0067		0.1489		0.3734	0.0575
PEI Score									-0.04	0.01
Mean	6.57	5.81	6.2	6.29	6.44	6.21	6.34	6.14		
(SD)	(3.68)	(3.70)	(3.54)	(3.83)	(3.41)	(3.76)	(3.81)	(3.57)		
P value	0.0062		0.7507		0.536		0.4962		0.2783	0.7746
Blood Pressure									Mean	
Controlled (n)	208	138	143	203	53	292	165	144	64.74	12.08
Uncontrolled (n)	240	180	195	224	70	349	210	174	66.51	12.24
P value	0.4489		0.1704		0.6861		0.7934		0.0195	0.3415

Marr.: married; NM: Not married; Emp.: employed; UE: unemployed; CC: Pearson correlation coefficient; Ed.: education

associations with PEI scores (Table 2). A significant association was detected between race and PEI score. It was found that blacks had significantly higher PEI scores than whites (black mean 6.57 versus white mean 5.81, p=0.0062). No other demographic factors had significant associations with the PEI score.

Blood Pressure Control

The only demographic factor found to be associated with blood pressure control was age. Not surprisingly, older patients were more likely to have high blood pressure. A one-way ANOVA found that participants with high blood pressure averaged 66.5 years of age, while those with normal blood pressure averaged 64.7 years (p=0.0195). No other participant characteristics were significantly associated with blood pressure control.

Multivariate Analysis

Linear regressions were carried out in order to view specific associations between independent and dependent variables, adjusting for sociodemographic factors (Table 3). The regressions were organized to follow a logical path described by our hypothesis and depicted in Figure 1.

The first regression was designed to examine the association of the PEI score with the patient's blood pressure status. Findings from regression #1 showed that, when controlling for sociodemographic factors, higher PEI scores were associated with high blood pressure (p=0.0004). The R^2 value of 0.03 signifies that the factors included in the regression explain 3% of the variation in the blood pressure status of patients in the study. Regressions #2 and 3 were designed to examine the stepwise effects of PEI on trust, and trust on blood pressure, respectively.

Regression #2 revealed a significant direct association between PEI and trust, meaning a higher number of physician counseling behaviors were associated with higher trust scores from the patient (p<0.0001). Other notable results from regression #2 reiterate the findings from the previous bivariate results—while controlling for the other sociodemographic factors, being white was associated with having higher trust in the physician (p<0.0001), and employed individuals were more likely to have higher trust (p<0.0001). The R^2 for regression #2 was 0.10. Regression #3 showed that while controlling for the sociodemographic factors, blood pressure status was not significantly associated with trust in physician. Regression #4 was designed to see the combined effect of PEI and trust on blood pressure status. We observed that once again, higher PEI scores were associated with high blood pressure (p=0.002). Interestingly, when comparing the PEI coefficients between regressions #4 (0.0175) and #1 (0.0193), we see that the inclusion of trust slightly dampens the effect of reported physician behavior on blood pressure status. In order to more closely observe the difference in the PEI/TIP relationship along racial lines, regression #5 was conducted. The results show that PEI is directly associated with trust in both whites (p=0.0018) and blacks (p<0.0001). Interestingly, this association was stronger for blacks (0.6985) than whites (0.5102).

DISCUSSION

We sought to examine the role of the patient–physician relationship on racial disparities in blood pressure control within the VA medical system. We hypothesized that white patients would experience more hypertension counseling behaviors from their doctors than blacks, and this would be associated with stronger feelings of trust in their physician and more controlled blood pressure.

Based on our findings, we did not see a significant difference in the blood pressure status between the black and white patients in our study. This could be a result of the equal access to healthcare presented within the VA system. Similar results were recently documented by Rehman et al., who found there is a greater percentage of African-American men with controlled blood pressure in the VA system as compared with non-VA clinics.[42]

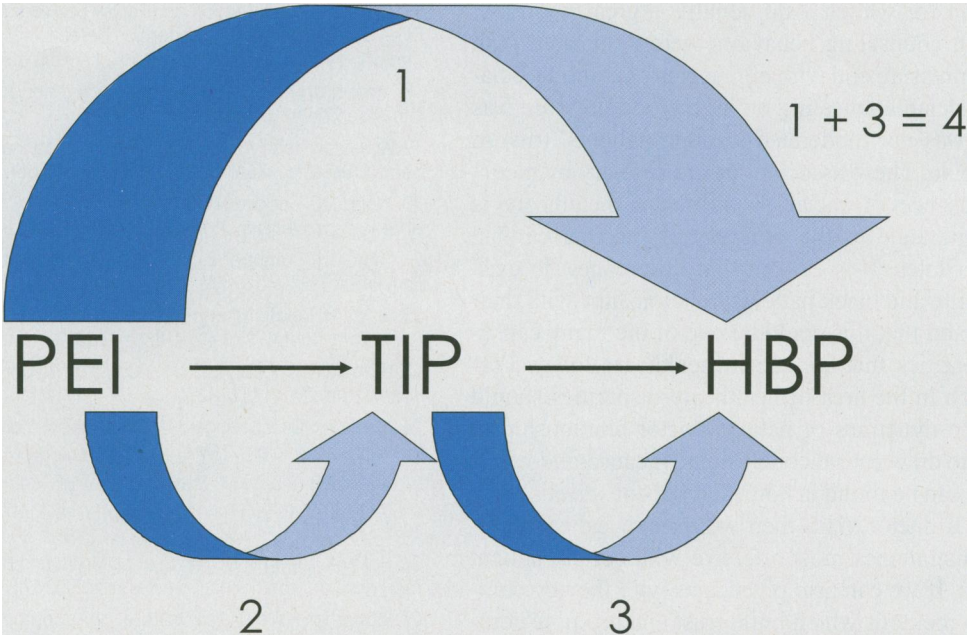
With regard to the PEI data measuring physician behavior, we found black patients to report more hypertension counseling behaviors than whites. This was a surprising finding because based on the current state of racial disparities in healthcare, we hypothesized that whites would experience more counseling behaviors.

Table 3. Linear regression results

	Independent Variables									
Dependent Variables	**PEI**	**TIP**	**Race**	**Age**	**Income**	**Mar.**	**Emp.**	**Educ.**	**R^2**	**F**
1) HBP	0.019*	–	-0.043	0.004	-0.033	-0.023	-0.015	0.008	0.03	2.51*
2) TIP	0.609*	–	-4.813*	-0.032	-0.012	-0.939	-2.786*	-0.001	0.10	9.57*
3) HBP	–	0.002	-0.020	0.005*	-0.043	-0.014	-0.046	0.013	0.02	1.53
4) HBP	0.018*	0.001	-0.036	0.003	-0.039	-0.016	-0.009	0.008	0.03	2.00*
5) TIP										
Whites	0.510*	–	–	-0.163*	0.109	-0.442	-2.732	0.120	0.08	3.80*
Blacks	0.699*	–	–	0.076	-0.031	-1.483	-2.634	-0.024	0.08	4.79*

HBP: high blood pressure; * p<0.05; TIP: trust in physician; PEI: patient exit interview; Mar.: marital status; Emp.: employment status; Educ.: education level

Figure 1. Hypothesis flow chart with regression designations



This discrepancy may be a result of the increased knowledge about blacks and their high risk for hypertension and cardiovascular disease (CVD). In other words, physicians at the VA may have been making an extra effort to counsel their hypertensive black patients because of their disproportionately higher death rates from CVD.

We also found that after controlling for sociodemographic factors, higher PEI scores were associated with higher reported trust in physicians. We also noted that this association varied by race, such that a stronger link between these variables was observed for African Americans. The former findings were in agreement with our hypothesis, as we believed more counseling by the physicians would lead to stronger feelings of trust. Similar findings were recently reported by Fiscella et al., who found that physicians who spent more time with their patients and thoroughly explored their illness were trusted more.[35]

Unexpectedly, we found higher PEI scores to be associated with higher blood pressure. This finding might be due to a reverse relationship in which patients with uncontrolled blood pressure may have received more attention from their physicians and thus more verbal reinforcement through additional physician counseling.

Similar to previous work with physician trust,[29-31] we found white patients have significantly higher trust when compared with blacks. This result remained, even after controlling for all other recorded sociodemographic characteristics. We also found that, regardless of other factors, patients who were employed reported higher trust than those who were unemployed. We did not find any significant association between trust and blood pressure control.

While there were associations of PEI with blood pressure, and PEI with trust, our results do not support the notion that PEI was linked with blood pressure through trust. Interestingly, we found that higher PEI scores were associated with higher trust among members of both races, and this association was stronger for African Americans. Despite this slightly stronger relationship, blacks still reported significantly lower trust in their physicians than did whites.

While there were meaningful findings from this study, there were some limitations which should be noted. First, the participating patient population is not representative of the entire American population due to their having equal access to medical care and being almost exclusively male (N=783, 98.7%). Additionally, this group of 793 patients was drawn from three VA sites and the results, therefore, cannot be generalized to the 25 million people eligible for VA care. Also, we could not judge the length of the relationship a patient had with a particular provider, which could have affected trust scores. In addition, the data collected in this study were from a one-time questionnaire, including all of the measures used in this study, and therefore represent a snapshot of the patients' feelings toward their physicians and blood pressure status. As a result, we are unable to conclude any causal relationships among physician behavior, trust in physician and blood pressure control.

In summary, we found no racial disparity in blood pressure control in our VA patient sample. Physician counseling behaviors were greater for African-American patients, but white patients reported a higher level of trust in their physician than did blacks. In multivariate models controlling for other sociodemographic factors, we observed

that there was a stronger association between physician counseling and trust in physician for African-American patients than for whites. Additionally, a greater number of physician counseling behaviors were associated with high blood pressure and with higher trust, but the association of physician counseling on high blood pressure was direct and was not moderated through patients' trust in their physician. These associations did not vary by race.

The presence of racial disparities in healthcare is a major hindrance to the progress of medical care in the United States. It is evident that differences do exist between white and black patients' relationships with their physicians and that this could be one of the many causes of the disparities that we see in healthcare today. Further research in the area of healthcare disparities should focus on the dynamics of patient–doctor relationship in an attempt to draw causal conclusions. If cause-and-effect differences can be found in how members of various races experience a doctor visit, then we may reveal the communication strategies most effective with certain patient populations. If we can arm physicians with the key communication methods which build trust and promote compliance, we can begin to break down portions of the racial disparities which plague America's healthcare system.

REFERENCES

1. Wenneker JB, Epstein AM. Racial inequalities in the use of procedures for patients with ischemic heart disease in Massachusetts. *JAMA*. 1989;261:253-257.

2. Carlisle DM, Leake BD, Shapiro MF. Racial and ethnic differences in the use of invasive cardiac procedures among cardiac patients in Los Angeles County, 1986 through 1988. *Am J Public Health*. 1995;85:352-356.

3. Williams DR, Collins C. U.S. socioeconomic and racial differences in health: patterns and explanations. *Annu Rev Sociol*. 1995;21:349-386.

4. Schoenbaum M, Waidmann T. Race, socioeconomic status, and health: accounting for race differences in health. *J Gerontol B Psychol Sci Soc Sci*. 1997;52:61-73.

5. Jha AK, Fisher ES. Racial trends in the use of major procedures among the elderly. *N Engl J Med*. 2005;353:683-691.

6. "Protecting the Health of Minority Communities." U.S. Department of Health and Human Services *Fact Sheet*, dated 09/24/02.

7. *National Healthcare Disparities Report: Summary*, 2004. Agency for Healthcare Research and Quality, Rockville, MD.

8. U.S. Department of Health and Human Services: Office of Minority Health. www.omhrc.gov/healthdisparities/index.htm. Accessed 12/01/05.

9. U.S. Census Bureau: Current Population Survey, March 2002. http://www.census.gov/population/www/socdemo/race/black.html. Accessed 11/14/05.

10. Kjellstrand CM. Age, sex, and race inequality in renal transplantation. *Arch Intern Med*. 1988;148:1305-1309.

11. Yergan J, Flood AB, LoGerfo JP, et al. Relationship between patient race and intensity of hospital services. *Med Care*. 1987;25:592-603.

12. Blendon RJ, Aiken LH, Freeman HE, et al. Access to medical care for black and white Americans: a matter of continuing concern. *JAMA*. 1989;261:278-281.

13. Schwartz E, Kofie VY, Rivo M, et al. Black/white comparisons of deaths preventable by medical intervention: United States and the District of Columbia 1980–1986. *Int J Epidemiol*. 1990;19:591-598.

14. Weissman JS, Stern R, Fielding SL, et al. Delayed access to health care: risk factors, reasons, and consequences. *Ann Intern Med*. 1991;114:325-331.

15. Pappas G, Queen S, Hadden W, et al. The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *N Engl J Med*. 1993;329:103-109.

16. Levine RS, Foster JE, Fullilove RE, et al. Black-white inequalities in mortality and life expectancy, 1933–1999: implications for healthy people 2010. *Public Health Rep*. 2001;116:474-483.

17. Hypertension Detection and Follow-up Program Cooperative Group. Five-year findings of the Hypertension Detection and Follow-up Program, II: mortality by race, sex and age. *JAMA*. 1979;242:2572-2577.

18. American Heart Association. www.americanheart.org/presenter.jhtml?identifier=2150. Accessed 11/14/05.

19. American Heart Association. www.americanheart.org/downloadable/heart/1106668161495AllAmAfAmHeartFacts05.pdf. Accessed 11/21/05.

20. Wong MD, Shapiro MF, Boscardin WJ, et al. Contribution of major disease to disparities in mortality. *N Engl J Med*. 2002;347:1585-1592.

21. Thom DH, Campbell B. Patient–physician trust: an exploratory study. *J Fam Pract*. 1997;44:169-176.

22. Brody H. Relationship-centered care: beyond finishing school. *J Am Board Fam Pract*. 1995;8:416-418.

23. Leopold N, Cooper M, Clancy C. Sustained partnership in primary care. *J Fam Pract*. 1996;42:129-137.

24. Mechanic D, Schlesinger M. The impact of managed care on patients' trust in their physicians. *JAMA*. 1996;275:1693-1697.

25. Thom DH, Ribisl KM, Stewart AL, et al. Validation of a measure of patients' trust in their physician: the Trust in Physician Scale. *Med Care*. 1999;37:510-517.

26. Hall MA, Zheng B, Dugan E, et al. Measuring patients' trust in their primary care providers. *Med Care Res Rev*. 2002;59:293-318.

27. Mostashari F, Riley E, Selwyn PA, et al. Acceptance and adherence with antiretroviral therapy among HIV-infected women in a correctional facility. *J Acquir Immune Defic Syndr*. 1998;18:341-348.

28. Safran DG, Taira DA, Rogers WH, et al. Linking primary care performance to outcomes of care. *J Fam Pract*. 1998;47:213-220.

29. Doescher MP, Saver BG, Franks P, et al. Racial and ethnic disparities in perceptions of physician style and trust. *Arch Fam Med*. 2000;9:1156-1163.

30. Cooper-Patrick L, Gallo JJ, Gonzales JJ, et al. Race, gender, and partnership in the patient–physician relationship. *JAMA*. 1999;282:583-589.

31. Kao AC, Green DC, Zaslavsky AM, et al. The relationship between method of physician payment and patient trust. *JAMA*. 1998;280:1708-1714.

32. Murray-Garcia JL, Selby JV, Schmittdiel J, et al. Racial and ethnic differences in a patient survey: patients' values, ratings, and reports regarding physician primary care performance in a large health maintenance organization. *Med Care*. 2000;38:300-310.

33. Taira DA, Safran DG, Seto TB, et al. Do patient assessments of primary care differ by patient ethnicity? *Health Serv Res*. 2001;36:1059-1071.

34. Heisler M, Smith DM, Hayward RA, et al. Racial disparities in diabetes care processes, outcomes, and treatment intensity. *Med Care*. 2003;41:1221-1232.

35. Fiscella K, Meldrum S, Franks P, et al. Patient Trust: is it related to patient-centered behavior of primary care physicians? *Med Care*. 2004;42:1049-1055.

36. Grumbach K, Selby JV, Damberg C, et al. Resolving the gatekeeper conundrum: what patients value in primary care and referrals to specialists. *JAMA*. 1999;282:261-266.

37. Thom DH, Stanford Trust SP. Physician behaviors that predict patient trust. *J Fam Pract*. 2001;50:323-328.

38. Ockene IS, Herbert JR, Ockene JK, et al. 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.

39. Pbert L, Adams A, Quirk M, et al. Luippold RS. The patient exit interview as an assessment of physician-delivered smoking intervention: a validation study. *Health Psychol*. 1999;18(2):183-188.

40. Anderson LA, Dedrick RF. Development of the trust in physician scale: a measure to assess interpersonal trust in patient–physician relationships. *Psychol Rep*. 1990;67:1091-1100.

41. Analyse-it for Microsoft Excel, Leeds, UK. www.analyse-it.com.

42. Rehman SU, Hutchison FN, Hendrix K, et al. Ethnic differences in blood pressure control among men at Veteran Affairs clinics and other health care sites. *Arch Intern Med*. 2005;165:1041-1047. ■

Predictors of Aggressive Therapy for Nonmetastatic Prostate Carcinoma in Massachusetts From 1998 to 2002

Adam J. Rose, MD,† Bertina M. Backus, MPH,‡ Susan T. Gershman, PhD,‡ Palmira Santos, PhD,§ Arlene S. Ash, PhD,* and Tracy A. Battaglia, MD, MPH**

Background: Most studies have found that black men are less likely to receive aggressive therapy for nonmetastatic prostate cancer, even after controlling for covariates. However, previous studies have not accounted for the clustering of outcomes by facility.

Objective: We sought to compare the proportions of black and white men receiving aggressive therapy for newly diagnosed nonmetastatic prostate cancer between 1998 and 2002, accounting for the clustering of outcomes by facility.

Methods: We used the Massachusetts Cancer Registry of all cancer diagnosed in residents of Massachusetts. We used logistic regression, clustering by the facility where the tumor was diagnosed, to predict the probability that a patient would receive any aggressive therapy, and the specific therapeutic choices of radical prostatectomy, external-beam radiation therapy, and brachytherapy. Predictors included race, age, poverty, insurance status, marital status, year of diagnosis, and tumor grade.

Results: Black men were similarly likely to receive aggressive therapy compared with white men (odds ratio [OR] 0.79, 95% confidence interval [CI] 0.62–1.01). However, there was a racial difference in the receipt of particular types of therapy: black men were significantly more likely to receive radiation therapy (OR 1.39, 95% CI 1.16–1.68) and less likely to receive radical prostatectomy (OR 0.53, 95% CI 0.38–0.74).

Conclusions: Among men diagnosed with nonmetastatic prostate cancer in Massachusetts from 1998 to 2002, black men received aggressive therapy at rates approaching those of whites. However, they were more likely to receive radiation therapy and less likely to receive radical prostatectomy.

Key Words: prostate cancer, African American, health inequality, cancer control, health services

(*Med Care* 2007;45: 440–447)

The choice of primary therapy for prostate cancer is without clear guidelines, and much depends upon patient and tumor characteristics, and patient preference.[1–3] The decision is informed by the slow-growing nature of this tumor, and the many competing causes of mortality among mostly older men who are treated for it. Many studies have considered radical prostatectomy, external beam radiotherapy, or brachytherapy to represent aggressive therapy, whereas any treatment strategy not involving 1 of these 3 is considered conservative management.[4–11]

A recent analysis of Surveillance, Epidemiology, and End Results (SEER)-Medicare data between 1986 and 1999[4] found that non-Hispanic white men were more likely to receive aggressive management for newly diagnosed prostate cancer than non-Hispanic black men. In the absence of evidence that such a difference is causing harm to 1 of the 2 groups, this remains a difference rather than an actual disparity in health care.[12] However, black men are not only diagnosed with prostate cancer at a higher rate, but also die of it at a higher rate[13]; the fact that they have received less aggressive management could be part of the reason for the higher mortality.

Most previous analyses have found that, after adjusting for covariates, black men are less likely than white men to receive aggressive therapy for nonmetastatic prostate cancer. However, although a growing body of research suggests that it is necessary to adjust for the site of care when examining racial disparities, previous analyses have not accounted for the clustering of outcomes by the facility providing care.[4,6,10,14–16] Therefore, we analyzed data from the Massachusetts Cancer Registry with regard to the primary therapy received for prostate cancers diagnosed between January 1, 1998 and December 31, 2002. We hypothesized that, after accounting for the clustering of outcomes by the facility reporting the tumor to the registry, black and white men would be equally likely to receive aggressive management.

From the *Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston Medical Center, Boston, Massachusetts; †Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Bedford, Massachusetts; ‡Massachusetts Cancer Registry, Massachusetts Department of Public Health, Boston, Massachusetts; and the §Massachusetts Comprehensive Cancer Control Coalition, Boston, Massachusetts.

Supported by a grant from the Department of Veterans Affairs Office of Academic Affairs (to A.J.R.); a Physician Training Award (PTAPM-97-185-04) from the American Cancer Society (to A.J.R. and T.A.B.); and the Centers for Disease Control and Prevention Cooperative Agreement U55/CCU521937-04 awarded to the Massachusetts Department of Public Health (to B.M.B. and S.T.G.).

Reprints: Adam J. Rose, Department of Medicine, Boston University School of Medicine, 91 E. Concord Street, Suite 200, Boston, MA 02118-2644. E-mail: adamrose@bu.edu.

The opinions expressed are the responsibility of the authors and do not necessarily represent the official views of the Department of Veterans Affairs or the Centers for Disease Control and Prevention.

ISSN: 0025-7079/07/4505-0440

METHODS

The Database

We selected all men from the Massachusetts Cancer Registry database with a diagnosis of "adenocarcinoma of the prostate" between January 1, 1998 and December 31, 2002. The Massachusetts Cancer Registry is a population-based cancer registry that collects information on newly diagnosed cases of cancer among residents of the state.[17] The Institutional Review Board of Boston University Medical Center approved this study.

We collected information on the tumor stage, using the tumor, node, metastasis (TNM) staging system. We excluded patients with stage 4, or metastatic, disease from our analyses, because the treatment decision for a patient with metastatic prostate cancer at the time of diagnosis is qualitatively different than the decision for a patient with nonmetastatic disease.

Independent Variable and Exclusion Criteria: Patient Race

We began with the entire cohort of men diagnosed with prostate cancer between January 1, 1998 and December 31, 2002. We excluded all patients from the database whose race and ethnicity were other than non-Hispanic white and non-Hispanic black, because their numbers were too small to allow for meaningful comparisons.

Dependent Variable and Exclusion Criteria: Therapy Received

Our primary outcome was the receipt of any aggressive therapy, defined as any therapeutic plan containing radical prostatectomy, external beam radiotherapy ("radiation therapy"), or brachytherapy. Secondary outcomes included the specific receipt of radical prostatectomy, radiation therapy, or brachytherapy. These outcomes were chosen in part to enable our results to be compared with the results of other analyses.[4–11] Because our outcome variable was the primary treatment received, we excluded all men who did not have any recorded information under "Date of First Course Treatment."

Covariates

To account for the clustering of outcomes, we adjusted for the facility that sent the initial report to the Registry about the patient. Although we recognize that patients may not be treated at the facility where they are diagnosed, the facility where the tumor is diagnosed represents the patient's entry point into the medical system and therefore may impact the treatment options that are offered to the patient. Because there were facilities that reported fewer than 150 cases of prostate cancer, or less than 30 cases per year, we combined all such facilities into 1 group, which contained 50 facilities and 2177 patients (9.9%). The other group, which contained 44 facilities, reported the remainder of the cases. Thus, our cluster analysis had 45 levels, one for all of the small centers combined and 44 for each of the remaining facilities.

We collected each patient's census tract of residence, as a proxy for socioeconomic status. In accordance with data and recommendations published by the Public Health Disparities Geocoding Project, we used the percentage of people in each census tract living below the federal poverty line to stratify the census tracts in Massachusetts with regard to socioeconomic status.[18,19] We divided the patients into 20 equal quantiles with respect to poverty in the census tract of residence.

We also collected the primary payer at the time of diagnosis, to control for the impact of insurance status on the therapy received. This variable was divided into the following groups: Medicaid, uninsured, and all others (Medicare, HMO, PPO, and Federal Employees' Insurance). Male patients must be severely disabled to receive Medicaid benefits in Massachusetts[20]; therefore, we adjusted for Medicaid insurance as a proxy measure for disability.

Age, a proxy for life expectancy, plays a major role in treatment decisions for prostate cancer. We classified age at the time of diagnosis into 9 strata: younger than 50, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, and older than 84. Married patients have been shown to be more likely to receive aggressive therapy for prostate cancer[14]; we dichotomized marital status into married versus all others.

We collected the Gleason Score to account for tumor grade, a measure of histologic aggressiveness. Gleason scores of 4 or less were recorded as "well differentiated," scores of 5 and 6 were recorded as "moderately differentiated," and scores of 7 and above were recorded as "poorly differentiated."

Analyses

We compared the primary therapy received by 2 racial groups: non-Hispanic whites and non-Hispanic blacks. We used a χ^2 analysis to compare demographics between racial groups, and to compare the proportion of each group receiving each treatment over the entire 5-year period. Then we conducted logistic regression analyses to predict receipt of radical prostatectomy, external-beam radiotherapy (radiation therapy), brachytherapy, or any aggressive therapy.

The independent variable of interest was the patient's race: white or black. Covariates of interest included age, year of diagnosis, tumor grade (Gleason score), poverty quantile, insurance status, and marital status. In our multivariable analyses, an odds ratio >1 indicates an increased chance of receiving a therapy. For some variables, including the year of treatment, the poverty category, and the age category, the odds ratio should be interpreted as "per increasing category of the variable." For the other variables, a reference group was used for comparisons. For the racial comparison, white non-Hispanic race was the reference group. For Gleason score, moderately-differentiated tumors was the reference group, as most tumors fall into this category. For marital status, unmarried was the reference group. For primary payer, the group containing Medicare and private insurance was the reference group.

Initially, we conducted logistic regressions without accounting for the clustering of outcomes by the facility reporting the tumor, to illustrate the results that might have been obtained by performing the analysis without a cluster adjustment. Then, we repeated our analyses using generalized estimating equations (GEE), taking into account the clustering of outcomes by the facility reporting the tumor. The GEE

analyses properly produce larger estimates of the standard error, and hence wider confidence intervals, than standard logistic regression analyses.

We also wished to examine the practical importance of whatever odds ratio we found regarding a racial difference. We began with the chance of an average white patient, age 62 or 77, to receive any aggressive therapy. We then converted this chance of receiving any aggressive therapy into odds. We then multiplied by the odds ratio for the receipt of aggressive therapy for a black patient, and converted the odds back into the percent chance for a black patient of similar age to receive aggressive therapy.

All analyses were conducted using SAS, version 9.0.

RESULTS

Descriptive Statistics

The Massachusetts Cancer Registry contains 25,465 cases of prostate cancer diagnosed between January 1, 1998 and December 31, 2002. Of these, there were 584 Hispanic men and 849 who belonged to other racial groups than white or black. In addition, there were 1942 men (9.2%) with incomplete data regarding treatment. The proportions with missing treatment information among white and black patients respectively were 8.2% and 11.8%. Finally, there were 1036 men whose disease was stage 4, or metastatic, at the time of diagnosis. The rate of metastatic disease at diagnosis was 6.0% among blacks and 4.2% among whites. After the exclusion of these 4361 patients, our final cohort contained 21,104 non-Hispanic white or black men with nonmetastatic disease at the time of diagnosis and complete treatment information.

According to US Census data from 2004, among men age 60 and older in Massachusetts, 93.2% of the population was white, 3.3% was black, and 3.5% were from other groups.[21] In the prostate cancer cohort of the Massachusetts Cancer Registry database, 91.4% patients were white, 5.3% were black, and 3.4% were from other groups. Thus, black men were represented in our prostate cancer database out of proportion to their percentage in the general population, which is consistent with what is known about the relative incidence of prostate cancer in black and white men.[13]

Descriptive statistics for our study cohort are in Table 1. Black men were less likely to be married (63% vs. 78%) and more likely to have Medicaid insurance (8% vs. 1%). Younger black men were more likely to be diagnosed with prostate cancer than younger white men; for example, 6% of black patients and 3% of white patients were diagnosed before the age of 50.

TABLE 1. Characteristics of All Non-Hispanic Black and White Men Diagnosed With Nonmetastatic Prostate Cancer in the Massachusetts Cancer Registry, 1998–2002

	White (n = 20,022)	Black (n = 1082)	*P*
Marital status			<0.001
Married	15,560 (78)	686 (63)	
All others	4462 (22)	396 (37)	
Year of diagnosis			0.007
1998	3572 (18)	183 (17)	
1999	3745 (19)	179 (17)	
2000	3984 (20)	206 (19)	
2001	4415 (22)	290 (27)	
2002	4306 (22)	224 (21)	
Age category, yrs			<0.001
<50	528 (3)	61 (6)	
50–54	1396 (7)	126 (12)	
55–59	2480 (12)	171 (16)	
60–64	3298 (16)	192 (18)	
65–69	4133 (21)	236 (22)	
70–74	4107 (21)	154 (14)	
75–79	2694 (13)	104 (10)	
80–84	941 (5)	30 (3)	
>84	445 (2)	8 (1)	
Insurance status			<0.001
Medicaid	169 (1)	90 (8)	
Uninsured	98 (1)	21 (2)	
All others	19,755 (98)	971 (90)	
Gleason score			<0.001
Well-differentiated	732 (4)	44 (4)	
Mod-differentiated	15,842 (79)	824 (76)	
Poorly-differentiated	2731 (14)	188 (17)	
Unknown	717 (4)	26 (2)	

Unadjusted Analyses

In unadjusted analyses, the age of the patient was highly predictive of the treatment received (Fig. 1). Younger age predicted treatment with radical prostatectomy and with aggressive therapy in general. Men in their 70s were the most likely to receive radiation therapy. The use of brachytherapy was less common in the youngest and the oldest patients.

Compared with non-Hispanic white men, non-Hispanic black men were more likely to receive radiation therapy (34.8% vs. 29.3%) and less likely to receive radical prostatectomy (32.0% vs. 37.8%; $P < 0.001$ for both analyses). Overall, there was no statistically significant difference in terms of the receipt of any aggressive therapy (78.5% vs. 79.4%; $P = 0.47$) or the receipt of brachytherapy (13.2% vs. 11.8%; $P = 0.17$).

Multivariable Analyses

Tables 2–4 show the results of our multivariable regressions to predict the receipt of any aggressive therapy, radical prostatectomy, and external-beam radiation therapy. In our initial logistic regression without accounting for the clustering of outcomes, black race was predictive of a lower odds of receiving any aggressive therapy (OR 0.79, 95% CI 0.67–0.94). However, after using GEE to account for the clustering of outcomes by the facility reporting the tumor to the registry, this effect was of marginal statistical significance (OR 0.79, 95% CI 0.62–1.01). Among different types of aggressive therapy, black men were significantly less likely to receive radical prostatectomy (OR 0.53, 95% CI 0.38–0.74) and more likely to receive radiation therapy (OR 1.39, 95% CI 1.16–1.68). There was no racial difference in the receipt of brachytherapy (OR 1.00, data not shown).

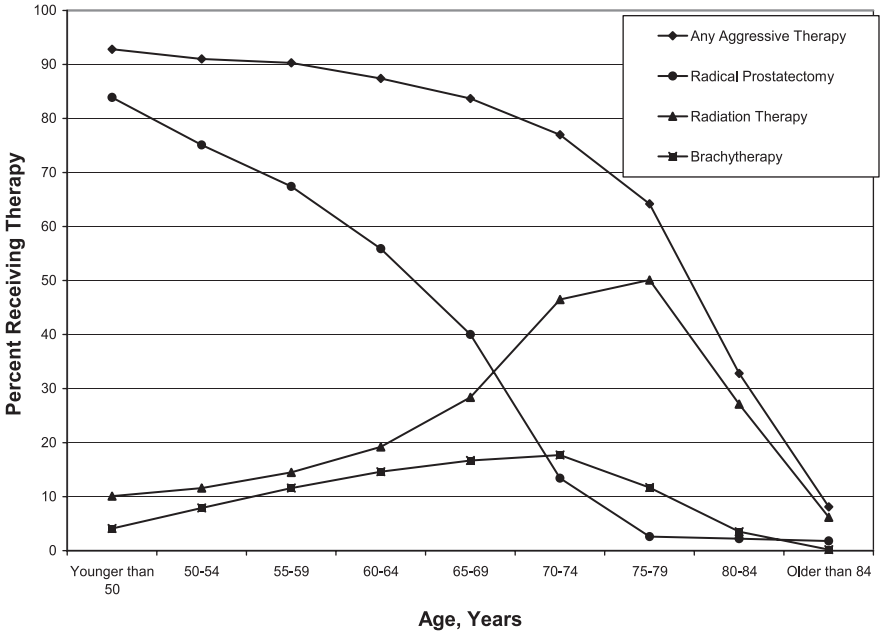


FIGURE 1. Different types of treatment of nonmetastatic prostate cancer, stratified by age group, in Massachusetts from 1998 to 2002.

TABLE 2. Multivariable Analysis: Odds Ratios and Their 95% Confidence Limits From a Logistic Regression Analysis to Predict Any Aggressive Therapy for Prostate Cancer in Massachusetts From 1998–2002

		95% CI		
Predictor Variable	**Odds Ratio**	**Before Cluster Adjustment**	**After Cluster Adjustment**	***P* After Cluster Adjustment**
Black race	0.79	0.67–0.94	0.62–1.01	0.06
Age, per 5 yrs	**0.58**	**0.57–0.60**	**0.55–0.62**	**<0.001**
Per increasing year of diagnosis	**0.82**	**0.80–0.85**	**0.78–0.87**	**<0.001**
Married	**1.46**	**1.34–1.59**	**1.31–1.63**	**<0.001**
Per increasing poverty quantile (20 groups)	0.994	0.987–1.000	0.974–1.014	0.56
Insurance status				
Medicare/private insurance (reference)	1.00	—	—	—
Medicaid	**0.61**	**0.44–0.84**	**0.42–0.88**	**0.008**
Uninsured	**0.44**	**0.29–0.69**	**0.28–0.71**	**<0.001**
Gleason score				
Moderately differentiated (reference)	1.00	—	—	—
Well differentiated	**0.32**	**0.27–0.37**	**0.23–0.43**	**<0.001**
Poorly differentiated	**0.76**	**0.69–0.84**	**0.67–0.87**	**<0.001**

Boldfaced variables are significant at the 0.05 level.

Our main result, the finding that black men were less likely to receive aggressive therapy, approached statistical significance ($P = 0.06$). Therefore, we examined the practical implications of an odds ratio of 0.79 for a black patient to receive any aggressive therapy. In our sample, the average white patient of age 62 had a 90% overall chance of receiving aggressive therapy (Fig. 1). After multiplying by the odds ratio of 0.79, an otherwise similar black patient of the same age would be 88% likely to receive aggressive therapy. Similarly, whereas a 77-year-old white patient had a 65% chance of receiving aggressive therapy, a 77-year-old black patient would have a 60% chance.

After adjustment for covariates, poverty was not a statistically significant predictor of the receipt of aggressive therapy (OR = 0.994, $P = 0.56$). However, increasing poverty did predict a lower odds of receiving a radical prostatectomy (OR = 0.978 per poverty quantile, 95% CI 0.966–0.991). Patients with Medicaid insurance or no insurance were less likely to receive aggressive therapy than the reference group (OR for Medicaid 0.61, OR for uninsured 0.44; $P = 0.008$ and $P < 0.001$). In addition, married patients were more likely to receive aggressive therapy (OR 1.46, 95% CI 1.31–1.63).

Over time, there was a decreasing use of aggressive therapy for prostate cancer. In later years, patients were less likely to receive aggressive therapy for similar tumors, as reflected in the OR per increasing year of diagnosis (0.82, 95% CI 0.78–0.87). In part, this change was due to a

TABLE 3. Multivariable Analysis: Odds Ratios and Their 95% Confidence Limits From a Logistic Regression Analysis to Predict Radical Prostatectomy for Prostate Cancer in Massachusetts From 1998–2002

Predictor Variable	Odds Ratio	95% CI		*P* After Cluster Adjustment
		Before Cluster Adjustment	After Cluster Adjustment	
Black race	**0.53**	**0.45–0.63**	**0.38–0.74**	**<0.001**
Age, per 5 yrs	0.43	0.42–0.44	0.41–0.45	<0.001
Per increasing year of diagnosis	0.95	0.92–0.98	0.91–1.00	0.04
Married	1.68	1.54–1.83	1.50–1.87	<0.001
Per increasing poverty quantile (20 groups)	**0.978**	**0.972–0.984**	**0.966–0.991**	**0.001**
Insurance status				
Medicare/private insurance (reference)	1.00	—	—	—
Medicaid	0.61	0.45–0.83	0.35–1.06	0.08
Uninsured	**0.48**	**0.31–0.75**	**0.29–0.78**	**0.003**
Gleason score				
Moderately differentiated (reference)	1.00	—	—	—
Well differentiated	**0.36**	**0.29–0.45**	**0.27–0.50**	**<0.001**
Poorly differentiated	1.08	0.97–1.20	0.89–1.31	0.45

Boldfaced variables are significant at the 0.05 level.

TABLE 4. Multivariable Analysis: Odds Ratios and Their 95% Confidence Limits From a Logistic Regression Analysis to Predict Radiation Therapy for Prostate Cancer in Massachusetts From 1998–2002

Predictor Variable	Odds Ratio	95% CI		*P* After Cluster Adjustment
		Before Cluster Adjustment	After Cluster Adjustment	
Black race	1.39	1.21–1.60	1.16–1.68	<0.001
Age, per 5 yrs	1.35	1.32–1.37	1.24–1.46	<0.001
Per increasing year of diagnosis	0.89	0.87–0.91	0.85–0.93	<0.001
Married	1.02	0.94–1.09	0.90–1.14	0.80
Per increasing poverty quantile (20 groups)	**1.016**	**1.011–1.022**	**1.000–1.033**	**0.05**
Insurance status				
Medicare/private insurance (reference)	1.00	—	—	—
Medicaid	1.17	0.88–1.56	0.76–1.82	0.48
Uninsured	1.32	0.88–1.98	0.86–2.04	0.21
Gleason score				
Moderately differentiated (reference)	1.00	—	—	—
Well differentiated	**0.55**	**0.46–0.66**	**0.41–0.74**	**<0.001**
Poorly differentiated	**1.43**	**1.31–1.56**	**1.21–1.68**	**<0.001**

Boldfaced variables are significant at the 0.05 level.

decreasing use of radiation therapy (OR for radiation therapy per increasing year of diagnosis 0.89, 95% CI 0.85–0.93).

DISCUSSION

We used the Massachusetts Cancer Registry to examine racial differences in the primary treatment received for prostate cancer from 1998 to 2002. In a logistic regression that did not account for the clustering of outcomes, non-Hispanic blacks were less likely to receive aggressive therapy than non-Hispanic whites. However, after accounting for clustering, the confidence interval for this finding crossed the null value. This clustering procedure, by appropriately widening confidence intervals, changed several results in our study from significant to nonsignificant, underscoring the importance of this maneuver.

Several previous analyses of the SEER-Medicare database have found mixed results regarding our primary question; however, none of these studies extends past 1999. In an analysis of a SEER-Medicare database between 1994 and 1996, Shavers et al[9] found that African Americans and Hispanics were less likely to receive aggressive management, even after adjustment for comorbidity, life expectancy, and census tract of residence. In another SEER-Medicare analysis from 1991 to 1999, Zeliadt et al[11] found that in an adjusted analysis, African American men were 26% less likely to receive aggressive management for their prostate cancer.

However, in an analysis of the SEER-Medicare database between 1995 and 1999, Denberg et al[14] showed that although black men were as likely as white men to receive aggressive management, their aggressive management was more likely to include radiation therapy and less likely to include radical prostatectomy.

In our study, non-Hispanic blacks were only modestly less likely to receive aggressive therapy, and the difference did not attain statistical significance. Our data is more recent than that used in the other studies. In addition, our study is the first of its kind to account for the clustering of outcomes. This procedure appropriately widened our confidence intervals, especially for the race and poverty variables, suggesting that variations in practice between facilities in this regard are large. When outcomes cluster by the site of care, then a failure to account for such clustering may produce a deceptively narrow confidence interval.

We translated an odds ratio of 0.79 into rates and found that actual differences in treatment between racial groups are small. For example, regarding 62-year-old men, white and black men had a 90% and 88% chance of receiving aggressive therapy, respectively. Other studies have found larger effect sizes than an odds ratio of 0.79 for this racial difference, but even these yield remarkably minor differences after translation into rates. Using a point estimate of 0.74, as found by Zeliadt et al,[11] a 62-year-old black man would have an 87% chance of receiving aggressive therapy. Even with a point estimate of 0.64, as was found by Underwood et al,[10] that man would still have an 85% chance of receiving aggressive therapy. Whereas a 5% difference between racial groups might constitute a public health concern, a 2% difference between groups seems quite small, especially when differences between different facilities overwhelm its statistical significance.

This study did not find a statistically significant difference between racial groups in the receipt of aggressive therapy. However, previous studies only found racial disparities in certain subgroups of the population. In an analysis of the Prostate Cancer Outcomes Study, Hoffman et al[7] found that among patients with more aggressive tumors, black patients received aggressive management 61% of the time, compared with 84% for white patients. In another analysis of the same data, Harlan et al[6] found that when the adjusted analysis was stratified by age, black men above the age of 60 were less likely to receive aggressive treatment.

To assess whether we would also find racial differences in certain subpopulations, we repeated our logistic regression (with adjustment for clustering) to predict the receipt of aggressive therapy, stratifying by whether the tumor exhibited aggressive characteristics (poorly-differentiated histology), as in the study by Hoffman et al.[7] Race was a significant predictor of aggressive therapy for the less-aggressive tumors (OR for black race = 0.78, 95% CI 0.61–0.99). However, race was not a significant predictor for the aggressive subset of tumors (OR = 0.81, 95% CI 0.50–1.30). Our findings in this regard are the reverse of what was found by Hoffman et al.[7] However, these 2 point estimates (0.78 and 0.81) are in fact similar, suggesting that these 2 groups may receive similar care.

We also repeated our logistic regression (with adjustment for clustering) to predict the receipt of aggressive therapy, stratifying by age as in the article by Harlan et al.[6] In the 60-and-over group, black race was not an independent predictor of the receipt of aggressive therapy (OR = 0.84; $P = 0.22$). However, in the younger age group, there was a nonsignificant trend toward blacks receiving less aggressive therapy (OR = 0.72; $P = 0.053$), the opposite of the result found by Harlan et al. The point estimates diverge considerably between these older and younger groups, suggesting that there may be a larger racial difference in the receipt of aggressive therapy among younger patients.

Despite the strengths mentioned above, our study does have limitations. Our ability to demonstrate a small difference in the face of a strong clustering effect based on the facility reporting the tumor was limited by our sample size, as we had only 1082 blacks in our database. As discussed earlier, 8.4% of our sample had missing treatment information. The proportion with missing treatment information did differ between white and black men (8.2% vs. 11.8%). If the absence of treatment data is more often associated with conservative management, this might have biased our results toward the null.

We did not control for pathologic tumor stage in our analyses. Radical prostatectomy may result in the "up-staging" of a tumor from stage 1 to stages 2 or 3, whereas other treatment options do not result in additional staging information. Thus, with regard to therapeutic choices, the role of pathologic tumor stage is confounded by observation bias,[22,23] in this case the differential ascertainment of pathologic stages 2 and 3 when performing a radical prostatectomy. Therefore, the estimates for the effect of different tumor stages on the therapy received are likely to be misleading and uninformative. Nevertheless, we did repeat our analyses, controlling for pathologic tumor stage, and none of our estimates for the other variables changed appreciably.

We could not adjust for the patient's level of prostate-specific antigen (PSA) at the time of diagnosis, as these data were not available. The Prostate Cancer Outcomes Study has demonstrated that a PSA score above 50 at the time of diagnosis independently predicts increased odds of receiving conservative management (OR = 3.3).[6] If black patients had a higher PSA score at the time of diagnosis in our cohort, as has been found elsewhere,[24–30] then adjusting for PSA at the time of diagnosis might have moved the point estimates for the odds of receiving aggressive therapy even closer to the null value. Therefore, this omission is likely to strengthen rather than weaken our results.

In addition, we did not have data regarding medical comorbidities. Whereas the patient's age does tell us something about his life expectancy, it clearly does not tell us everything, and life expectancy goes to the heart of the treatment decision for prostate cancer. Our analysis would have been improved by the addition of some mechanism to adjust for the effect of medical comorbidities.

Although we adjusted for Medicaid insurance, we did so primarily as a way to adjust for severe disability by proxy. Male patients in Massachusetts must demonstrate severe physical or mental disability to qualify for Medicaid, rather than merely demonstrating a certain degree of poverty.[20] Therefore, it is likely that considerations of life expectancy, expected quality of life, and ability to tolerate aggressive therapies, rather than issues of insurance and reimbursement, informed the decision to provide less aggressive treatment of these patients.

Finally, it would be an oversimplification to imply that the patient is merely a passive participant in the choice of therapy for prostate cancer. Whereas it is the physician's job to summarize the medical evidence as he or she understands it, the patient brings his own values and preferences to the decision, which will inform the decision regarding whether to pursue aggressive therapy at all, and if so, which therapy to choose.[3,31,32] In addition, clinicians may not offer the same treatment options to different racial groups; Denberg et al have shown that urologists may be less likely to recommend radical prostatectomy to black patients who are socially vulnerable.[33] As with any analysis of a large database, our study cannot differentiate between treatment disparities that arise as a result of the stated values of the patient and treatment disparities that result from differences in access to care or unrecognized bias on the part of the clinician.

Our study suggests that racial differences are no longer an important target for remediation in the treatment received for prostate cancer, at least in Massachusetts. However, black men continue to have excess mortality after a diagnosis of prostate cancer.[13] In our sample, black patients were 1.4 times as likely to have metastatic disease at diagnosis; this might account for part of the mortality gap. Further work is needed to determine whether new screening procedures or more judicious use of current screening procedures will help to narrow this difference in outcomes.

The widening of confidence intervals after accounting for clustering suggests that whereas racial differences may be small in the overall population, they may be larger at certain medical centers. To the extent that certain facilities can be demonstrated to treat different racial groups in different ways, this suggests a target for remediation.

In addition, the broad categories of treatment investigated in this study may not tell the entire story regarding racial differences in the management of prostate cancer. Shavers et al[34] have shown that black and Hispanic patients treated conservatively for prostate cancer may receive less vigilant monitoring than white patients. Therefore, even within a category such as "conservative management," meaningful racial differences may still exist. Much work remains to be done in describing and reducing these subtler differences in treatment.

In conclusion, we studied non-Hispanic black and non-Hispanic white men diagnosed with nonmetastatic prostate cancer in Massachusetts during the period from 1998 to 2002. In logistic regression analyses that did not account for the clustering of outcomes, black patients were slightly less likely to receive aggressive therapy. However, after accounting for the clustering of outcomes by the facility reporting the tumor, this difference was no longer statistically significant, underscoring the importance of this maneuver. Racial differences in treatment do not explain why black men continue to experience excess mortality from prostate cancer.

ACKNOWLEDGMENTS

The authors thank Nancy Kressin, Dan Berlowitz, and Howard Cabral for their valuable contributions to this manuscript.

REFERENCES

1. Kattan MW, Cowen ME, Miles BJ. A decision analysis for treatment of clinically localized prostate cancer. *J Gen Intern Med*. 1997;12:299–305.
2. Barry MJ, Fleming C, Coley CM, et al. Should medicare provide reimbursement for prostate-specific antigen testing for early detection of prostate cancer? Part IV. Estimating the risks and benefits of an early detection program. *Urology*. 1995;46:445–461.
3. Patel HR, Mirsadraee S, Emberton M. The patient's dilemma: prostate cancer treatment choices. *J Urol*. 2003;169:828–833.
4. 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.
5. Harlan L, Brawley O, Pommerenke F, et al. Geographic, age, and racial variation in the treatment of local/regional carcinoma of the prostate. *J Clin Oncol*. 1995;13:93–100.
6. Harlan LC, Potosky A, Gilliland FD, et al. Factors associated with initial therapy for clinically localized prostate cancer: prostate cancer outcomes study. *J Natl Cancer Inst*. 2001;93:1864–1871.
7. Hoffman RM, Harlan LC, Klabunde CN, et al. Racial differences in initial treatment for clinically localized prostate cancer. Results from the prostate cancer outcomes study. *J Gen Intern Med*. 2003;18:845–853.
8. Shavers VL, Brown ML. Racial and ethnic disparities in the receipt of cancer treatment. *J Natl Cancer Inst*. 2002;94:334–357.
9. Shavers VL, Brown ML, Potosky AL, et al. Race/ethnicity and the receipt of watchful waiting for the initial management of prostate cancer. *J Gen Intern Med*. 2004;19:146–155.
10. Underwood W, De Monner S, Ubel P, et al. Racial/ethnic disparities in the treatment of localized/regional prostate cancer. *J Urol*. 2004;171:1504–1507.
11. Zeliadt SB, Potosky AL, Etzioni R, et al. Racial disparity in primary and adjuvant treatment for nonmetastatic prostate cancer: SEER-Medicare trends 1991 to 1999. *Urology*. 2004;64:1171–1176.
12. Rathore SS, Krumholz HM. Differences, disparities, and biases: clarifying racial variations in health care use. *Ann Intern Med*. 2004;141:635–638.
13. Stat bite: prostate cancer incidence and mortality among white and black men, 1975–2000. *J Natl Cancer Inst.* 2003;95:1357.
14. Denberg TD, Beaty BL, Kim FJ, et al. Marriage and ethnicity predict treatment in localized prostate carcinoma. *Cancer*. 2005;103:1819–1825.
15. Klabunde CN, Potosky AL, Harlan LC, et al. Trends and black/white differences in treatment for nonmetastatic prostate cancer. *Med Care*. 1998;36:1337–1348.
16. Polednak AP, Flannery JT. Black versus white racial differences in clinical stage at diagnosis and treatment of prostatic cancer in Connecticut. *Cancer*. 1992;70:2152–2158.
17. Massachusetts Department of Public Health, Massachusetts Cancer Registry. Available at: http://www.mass.gov/dph/dphorg2.htm. Accessed November 30, 2006.
18. 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.
19. Krieger N, Chen JT, Waterman PD, et al. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am J Public Health*. 2005;95:312–323.

20. *Understanding MassHealth Members with Disabilities*. Boston, MA: Massachusetts Medicaid Policy Institute; 2004.
21. United States Census Department, 2004 American Community Survey. Available at: http://factfinder.census.gov. Accessed November 30, 2006.
22. Aschengrau A, Seage GR. *Essentials of Epidemiology in Public Health*. Sudbury, MA: Jones and Bartlett Publishers; 2003.
23. Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston, MA: Little, Brown & Co.; 1987.
24. Eastham JA, Kattan MW. Disease recurrence in black and white men undergoing radical prostatectomy for clinical stage T1-T2 prostate cancer. *J Urol*. 2000;163:143–145.
25. Grossfeld GD, Latini DM, Downs T, et al. Is ethnicity an independent predictor of prostate cancer recurrence after radical prostatectomy? *J Urol*. 2002;168:2510–2515.
26. Kim JA, Kuban DA, el-Mahdi AM, et al. Carcinoma of the prostate: race as a prognostic indicator in definitive radiation therapy. *Radiology*. 1995;194:545–549.
27. Moul JW, Connelly RR, Mooneyhan RM, et al. Racial differences in tumor volume and prostate specific antigen among radical prostatectomy patients. *J Urol*. 1999;162:394–397.
28. Moul JW, Sesterhenn IA, Connelly RR, et al. Prostate-specific antigen values at the time of prostate cancer diagnosis in African-American men. *JAMA*. 1995;274:1277–1281.
29. Powell IJ, Dey J, Dudley A, et al. Disease-free survival difference between African Americans and whites after radical prostatectomy for local prostate cancer: a multivariable analysis. *Urology*. 2002;59:907–912.
30. Underwood W III, Wei J, Rubin MA, et al. Postprostatectomy cancer-free survival of African Americans is similar to non-African Americans after adjustment for baseline cancer severity. *Urol Oncol*. 2004;22:20–24.
31. Flood AB, Wennberg JE, Nease RF Jr, et al. The importance of patient preference in the decision to screen for prostate cancer. Prostate Patient Outcomes Research Team. *J Gen Intern Med*. 1996;11:342–349.
32. Elkin EB, Cowen ME, Cahill D, et al. Preference assessment method affects decision-analytic recommendations: a prostate cancer treatment example. *Med Decis Making*. 2004;24:504–510.
33. Denberg TD, Kim FJ, Flanigan RC, et al. The influence of patient race and social vulnerability on urologist treatment recommendations in localized prostate carcinoma. *Med Care*. 2006;44:1137–1141.
34. Shavers VL, Brown M, Klabunde CN, et al. Race/ethnicity and the intensity of medical monitoring under 'watchful waiting' for prostate cancer. *Med Care*. 2004;42:239–250.

Original Paper

Understanding Uncontrolled Hypertension: Is It the Patient or the Provider?

Adam J. Rose, MD, MSc;[1,2] Dan R. Berlowitz, MD, MPH;[1,2,3] Michelle B. Orner, MPH;[1] Nancy R. Kressin, PhD[1,2]

The relative contributions of adherence and treatment intensity to blood pressure (BP) control are not well understood. The authors studied patients with uncontrolled hypertension (N=410) from 3 primary care clinics in the Veterans Affairs (VA) medical system. A questionnaire was used to assess patient adherence to therapy, and VA system pharmacy fills were used to assess the intensity of the antihypertensive regimen. At baseline, an inadequate antihypertensive regimen was implicated as the most probable reason for uncontrolled BP in a majority of patients (72%), while nonadherence could only be implicated in 13%. In multivariate longitudinal analyses, patients who had an increase in their medical treatment during the study had lower final diastolic BP levels compared with the patients who did not (–3.70 mm Hg; P<.05). While patient adherence to therapy plays a role, vigorous clinical management by the clinician is a more important contributor to BP control. (J Clin Hypertens. 2007;9:937–943) ©2007 Le Jacq

From the Center for Health Quality, Outcomes, and Economic Research (a VA Health Services Research and Development National Center of Excellence), Bedford VA Medical Center, Bedford, MA;[1] the Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA;[2] and the Department of Health Policy and Management, Boston University School of Public Health, Boston, MA[3]
Address for correspondence:
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
Manuscript received July 30, 2007; accepted August 16, 2007

www.lejacq.com ID: 7332

Despite concerted efforts to improve the treatment of hypertension, only 64% of the hypertensive patients in the United States who were treated in 2003 and 2004 had controlled blood pressure (BP) (<140/90 mm Hg).[1] There are 3 main causes of failure to control BP despite therapy: patient nonadherence, insufficient titration of therapy, and resistant disease. A large body of research has documented the contribution of patient nonadherence to poor control in many chronic conditions, including hypertension.[2–4] More recently, a growing body of literature has also documented clinician failures to escalate therapy to bring a chronic condition under control.[5–7] Hypertension is only one of many chronic conditions to be affected by this failure to appropriately titrate therapy, which Phillips and colleagues[8] have called "clinical inertia." Finally, it has long been recognized that some patients have resistant hypertension, which is defined as a BP level that remains >139/89 mm Hg despite apparently adequate adherence and therapeutic intensity.[9,10] Understanding the relative contributions of these factors to uncontrolled hypertension is important for designing effective interventions to improve hypertension control; however, no previous study has addressed this issue in a primary care population.

We undertook this prospective cohort study of a group of patients from the Veterans Affairs (VA) medical system with uncontrolled hypertension to address 2 questions. First, what proportion of patients with uncontrolled hypertension had poor adherence, inadequate management, or neither of these? Second, would patient adherence and treatment intensity predict BP control at the end of the study? By addressing these questions, we sought to address the relative impact of poor adherence and treatment intensity on BP control.

Table I. Questions Used to Assess Adherence to Therapy

1. Some people have difficulty in taking blood pressure medication as prescribed. Do you have difficulty with this?
2. How many days in the past week did you forget to take your blood pressure medication?
3. How many days in the past week did you not take your medication on purpose?
4. How many days in the past week did you add an extra pill?
5. Did you ever take less medicine because you felt you needed less?
6. Sometimes if you feel worse when you take the medicine, do you stop taking it?

METHODS

The Cohort

Our sample was drawn from a larger study of VA system patients with hypertension conducted between January 1, 2002, and April 21, 2004.[11] We identified all non-Hispanic white and non-Hispanic black patients with outpatient diagnoses of hypertension on at least 2 separate occasions in 2001 at 3 urban tertiary care VA medical centers (*International Classification of Diseases, Ninth Revision [ICD-9]* diagnosis codes: 401, 401.0, 401.1, 401.9, 405–405.11, 405.19, 405.9, 405.91, 405.99). The study was approved by the institutional review board of all participating facilities, and patients provided informed consent.

Using this "universe" of 11,731 hypertensive patients, study staff tracked these patients' primary care visits over a 14-month period and, as they presented for care, invited 1210 of them to complete a questionnaire regarding their self-care for hypertension. Of these 1210, 204 (17%) were excluded from the study: 18 because their race was other than non-Hispanic white or non-Hispanic black, 41 because of impaired cognition, 59 because they denied having hypertension, 6 because they were already enrolled in another hypertension study, and 80 for miscellaneous reasons including being too ill to participate or moving or dying before they could be enrolled in the study. Of the 1006 eligible patients, 793 (79%) completed the survey, and the remainder refused to participate. Research associates verbally administered the questionnaire to patients and recorded the responses.

Of those 793 patients, we studied only the patients whose initial BP, as recorded in the electronic medical record (EMR), was uncontrolled based on the definition in the Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC VI) guidelines (≥140/90 mm Hg).[12] Because such a small portion of our sample (1%) comprised women, we excluded them from our analyses. Thus, our final sample was composed of 410 male patients with uncontrolled BP at baseline who had completed the questionnaire.

Dependent Variable: BP

Patient BP values were our dependent variables. We used the BP level as recorded in the EMR rather than obtaining BP using a standardized research protocol; we reasoned that this was the figure that would guide clinician decisions regarding therapy. When multiple BP measurements were taken on the same day, we used the one with the lowest systolic BP.

The BP level of each patient was measured within 3 days of enrolling in the study; this constituted the initial BP. The study ended on April 21, 2004; the last BP reading obtained for each patient before this date became his outcome BP. We analyzed systolic and diastolic BP values as continuous measures of BP control. In addition, we dichotomized all BP values into 2 categories: controlled and not controlled, based on whether they were >139/89 mm Hg. If either the systolic or the diastolic BP was in excess of this goal, the BP was considered uncontrolled. While guidelines such as JNC VI and JNC 7 have called for lower BP targets in patients with diabetes and renal disease,[12,13] VA clinical guidelines continued to support a formal goal of 140/90 mm Hg for all patients throughout our study period.[14]

Independent Variable: Adherence

We measured adherence using the portion of the patient questionnaire asking about issues with adherence to therapy. The 6 questions in this portion of the instrument (Table I) were adapted from the work of Choo and associates[15] and Morisky[4] and colleagues. Such self-reported measures of adherence have been demonstrated to correlate with end points such as BP control in our cohort and others.[4,11] Among the 793 patients who completed the questionnaire, patients who endorsed at least 2 of the 6 items were significantly more likely to have uncontrolled BP at baseline (odds ratio, 1.86; $P<.001$). We took this to imply validity of this as a measure of patient adherence, which relates to BP control. Therefore, we classified patients as adherent or nonadherent with therapy based on whether they endorsed at least 2 of the 6 items.

Independent Variable: Treatment Intensity

There were 2 variables used to capture the concept of treatment intensity, one that we used at baseline

and the other that we used for the longitudinal analysis. At baseline, we divided all patients into 2 groups based on whether their regimen met our definition for adequacy. We defined adequate therapy as a regimen containing drugs from at least 3 different classes of antihypertensives, at least 1 of which had to be a loop or thiazide diuretic. Inadequate use of diuretics has been particularly identified as a common reason for uncontrolled hypertension.[10,16–19]

All 3 drugs had to be prescribed in at least a moderate dosage. Minimal dosages to satisfy this definition and the division of drugs into different classes were defined for all agents used in this study; the list in Table II was discussed among several clinicians who agreed that it had validity. This definition of an adequate regimen for difficult-to-control hypertension is based on the JNC VI guidelines,[12] which were current at the inception of this study and are echoed by the VA guidelines for the management of hypertension,[14] both now and at the time of the study.

For the longitudinal analysis, we determined for each patient whether therapy was increased between the beginning and the end of the study. The regimen was considered to have been augmented if the patient was receiving an increased number of antihypertensive medications at the end of the study than the beginning or if any of the medication doses were increased during the study.

Covariates

We adjusted for the known predictors of BP control to isolate the effects of adherence and therapy on BP. Race was based on patient self-report and was divided into 2 groups: non-Hispanic black and non-Hispanic white. Patient age was recorded in the EMR and was divided into 3 categories: 64 years or younger, 65 to 74 years and 75 years and older. Borzecki and colleagues[20] have shown that the effect of age on the management and control of hypertension is categoric rather than linear; therefore, we categorized age to capture this effect. The *ICD-9* codes for diabetes mellitus, renal disease, and coronary artery disease were taken from the EMR. Body mass index was calculated from height and weight as recorded in the EMR, and patients with a body mass index ≥30 were considered obese. We also adjusted for the frequency of clinical visits using the average number of BP values per month as a proxy measure. In all analyses, we adjusted for the clustering of outcomes by the site of care, modeling site as a random effect. Finally, in all analyses, we adjusted for the effect of baseline BP on final BP.

Table II. Antihypertensive Agents Used in Our Study, Divided by Class, With Adequate Dosages

Angiotensin-converting enzyme inhibitors
Captopril 150 mg/d
Enalapril 20 mg/d
Fosinopril 20 mg/d
Lisinopril 20 mg/d
Aldosterone antagonists
Spironolactone 25 mg/d
α-Blockers
Doxazosin 4 mg/d
Prazosin 10 mg/d
Terazosin 5 mg/d
Angiotensin receptor blockers
Candesartan 16 mg/d
Irbesartan 150 mg/d
Valsartan 160 mg/d
β-Blockers
Atenolol 50 mg/d
Carvedilol 25 mg/d
Labetalol 600 mg/d
Metoprolol 100 mg/d (either formulation)
Propranolol 80 mg/d
Calcium channel blockers (dihydropyridines)
Amlodipine 5 mg/d
Felodipine 5 mg/d
Nifedipine 60 mg/d
Calcium channel blockers (nondihydropyridines)
Diltiazem 180 mg/d
Verapamil 180 mg/d
Centrally acting vasodilators (each is a unique drug class)
Clonidine
Patch 0.2 mg
Tablets 0.6 mg/d
Hydralazine 100 mg/d
Minoxidil 20 mg/d
Diuretics (loop)
Bumetanide 1 mg/d
Furosemide 40 mg/d
Diuretics (thiazide)
Chlorthalidone 25 mg/d
Hydrochlorothiazide 25 mg/d
Metolazone 0.5 mg/d

Statistical Analyses

We began with a cross-sectional analysis of the correlates of uncontrolled hypertension at baseline. We divided our sample into 3 groups: those whose uncontrolled hypertension at baseline could be attributed to poor adherence, inadequate management, or physiologic resistance, respectively. Our first group contained the patients reporting poor

Table III. Baseline Characteristics of the Cohort[a]

Characteristic	Group 1 (n=410)	Group 2 (n=338)
Age, y		
64 or younger	173 (42)	136 (40)
65–74	131 (32)	109 (32)
75 or older	106 (26)	93 (28)
Initial systolic BP, mean (SD), mm Hg	155.0 (13.9)	155.0 (13.8)
Initial diastolic BP, mean (SD), mm Hg	80.2 (11.9)	80.6 (11.9)
<0.5 BP measurements per month	131 (32)	117 (35)
Race/ethnicity		
Non-Hispanic black	236 (58)	191 (57)
Non-Hispanic white	174 (42)	147 (43)
Obesity (body mass index ≥30)	200 (50)	152 (46)
Diabetes mellitus	189 (46)	145 (43)
Coronary artery disease	192 (47)	151 (45)
Renal disease	105 (26)	78 (23)

Values are expressed as No. (%) unless otherwise indicated.
[a]Group 1 is all patients with uncontrolled blood pressure (BP) at baseline, while group 2 is the subset of group 1 who also had inadequate therapy at baseline considering their uncontrolled BP.

adherence to therapy; if adherence is sufficiently poor, even the best management may not succeed in controlling BP. We then divided the remaining patients into those who were receiving an adequate regimen and those who were not, according to the definition discussed earlier. By process of elimination, patients whose BP was poorly controlled despite apparently adequate adherence and therapy were considered to have physiologically resistant hypertension.

We then examined the predictors of final BP using the subset of patients with uncontrolled BP at baseline whose initial therapy was also inadequate and thus presented an opportunity for intensification. We used linear regressions to measure the ability of our independent variables to predict the final systolic BP and diastolic BP levels, controlling for covariates and the site of care. We used logistic regression to measure the ability of our independent variables to predict whether the final BP would be controlled, controlling for covariates and the site of care. We performed all analyses using SAS 9.1 (SAS Institute Inc, Cary, NC).

RESULTS

There were 410 patients with baseline uncontrolled hypertension (group 1) in our study cohort (Table III). Of these 410 patients, 17% were on adequate regimens, while 67% were on <3 medications and 16% were on ≥3 medications but at inadequate doses or without the use of a diuretic. The 338 patients whose therapy was inadequate at baseline in light of their uncontrolled BP formed a second group for analysis (group 2). During the 2-year study period, 51% of the patients experienced a change in their medical regimens with an increase in medication.

The average patient was followed for 438 days, with 95% of the sample having at least 221 days of data. The median patient had 12 BP values during the study; 95% of the sample had at least 5 BP values. The average age was 66.5 years, and more than half of the patients (58%) were black. The prevalence of comorbid conditions was high; only 27% of the patient population did not have at least 1 of 3 comorbidities, namely diabetes, renal disease, and coronary artery disease. Also, 49% of the sample was obese.

Fifty-two (13%) patients endorsed ≥2 of the adherence questions and were thus labeled as poorly adherent with their therapy. Of the adherent patients, 297 (72%) were on an inadequate regimen of antihypertensive medications considering their BP. Sixty-one patients (15%) with uncontrolled hypertension did not report poor adherence and were on an adequate regimen of medications. These patients were considered to have resistant hypertension, presumably due to physiologic factors.

The mean final BP level in the entire sample was 142/75 mm Hg, a marked improvement from the initial mean level of 155/80 mm Hg. When we examined the subset of patients with inadequate therapy at baseline despite their uncontrolled hypertension (n=338), none of our independent variables significantly predicted the final systolic BP. Higher treatment intensity, however, significantly predicted a lower diastolic BP level in adjusted analyses (Table IV).

Controlled BP (<140/90 mm Hg) was achieved in a substantial proportion of the cohort (46%) by the end of the study. In bivariate analyses, BP control was achieved in patients with poor adherence 33% of the time by the end of the study, compared with 47% for the other patients (P=.04). In patients with inadequate therapy at baseline, after adjustment for covariates (Table V), this result was no longer statistically significant (odds ratio, 0.52; 95% confidence interval, 0.25–1.09; P=.08).

DISCUSSION

We compared the effects of patient adherence and clinician management of hypertension on the initial

Table IV. Predictors of Follow-Up Systolic and Diastolic BP Levels in Patients With Uncontrolled BP and Inadequate Therapy at Baseline[a]

	SBP (95% CI)	DBP (95% CI)
Poor adherence to medication	0.19 (–6.18 to 6.56)	0.18 (–3.49 to 3.85)
Any change in therapy vs none	–2.83 (–7.05 to 1.38)	–3.70 (–6.13 to –1.28)[b]
Baseline SBP or DBP (per mm Hg)	0.43 (0.27 to 0.58)[b]	0.35 (0.24 to 0.46)[b]
<0.5 BP measurements per mo	1.78 (–2.86 to 6.42)	–0.15 (–2.75 to 2.46)
Age (oldest vs youngest)	3.61 (–4.11 to 11.33)	–5.48 (-10.07 to –0.89)[b]
Age (moderate vs youngest)	–0.95 (–8.06 to 6.17)	–5.49 (–9.66 to –1.33)[b]
Body mass index ≥30	0.68 (–3.59 to 4.95)	0.78 (–1.68 to 3.24)
Diabetes mellitus	2.89 (–1.53 to 7.30)	–1.12 (–3.67 to 1.42)
Coronary artery disease	–0.93 (–5.17 to 3.32)	0.26 (–2.20 to 2.72)
Renal disease	1.50 (–3.75 to 6.75)	–1.41 (–4.42 to 1.60)
Black race	0.03 (–4.57 to 4.63)	0.50 (–2.09 to 3.09)

[a]Results from multivariate linear regression models, accounting for site of care as a random effect.[b]Results are significant at the .05 level (n=338). β-Coefficients are expressed in units of mm Hg; a β-coefficient of +1.0 indicates a 1.0-mm Hg increase in final blood pressure (BP). Abbreviations: CI, confidence interval; DBP, diastolic BP; SBP, systolic BP.

and follow-up BP of a group of patients in the VA medical system. Our baseline analysis suggests that inadequate treatment intensity, or clinical inertia, is the most common reason for uncontrolled hypertension (72%). In addition, in longitudinal analyses, increased treatment intensity was associated with a lower final diastolic BP level, while the effects of adherence were not statistically significant.

Only limited research has compared the relative contributions of patient adherence, clinical inertia, and resistant hypertension with the failure to control the BP in hypertensive patients receiving pharmacotherapy. Two case series document the experience of a single hypertension referral center.[17,21] These studies found that an inadequate medical regimen was the most common cause of uncontrolled hypertension on referral to their clinic (58%), followed by poor adherence as the second most common cause (16%); these proportions are similar to what we found. In a smaller but intriguing study, Javors and Bramble[22] reviewed the care of 30 patients with various chronic conditions to investigate the relative effects of guideline-based management and patient adherence on long-term control. In their study, BP in the majority of patients was not well controlled, and clinician failure to adhere to guideline-based management was strongly predictive of uncontrolled disease. Patient adherence was generally high in their sample and did not predict control of the chronic conditions. Like their study, ours suggests that vigorous clinical management may contribute more to the control of chronic conditions than patient adherence to therapy.

It is of interest to note that in a recent Harris Survey, >90% of hypertensive patients reported that they were receiving antihypertensive therapy

Table V. Predictors of Follow-Up BP Control in Patients With Uncontrolled BP and Inadequate Therapy at Baseline[a]

	OR	95% CI
Poor adherence to medication	0.52	0.25–1.09
Any change in therapy vs none	1.23	0.77–1.96
Initial systolic BP (per mm Hg)	0.965	0.947–0.983[b]
Fewer than 0.5 BP measurements per mo	0.99	0.60–1.64
Age (oldest vs youngest)	0.65	0.27–1.53
Age (moderate vs youngest)	1.23	0.56–2.67
Body mass index ≥30	0.82	0.52–1.32
Diabetes mellitus	0.84	0.52–1.37
Coronary artery disease	1.08	0.68–1.73
Renal disease	1.22	0.68–2.19
Black race	1.23	0.75–2.03

[a]Results of a multivariate logistic regression analysis accounting for site of care as a random effect. [b]Results are significant at the .05 level (n=338). Odds ratios (OR) >1.0 indicate a higher likelihood of controlled blood pressure (BP) at the end of the study. Abbreviation: CI, confidence interval.

but, consistent with other reports, >30% did not have their medication increased despite continued elevated BP levels.[23]

In part, our study examined clinician adherence to JNC 6 guidelines, which were current at the time of this study, regarding BP goals and what constitutes a reasonable regimen for resistant hypertension.[12] JNC guidelines have been widely disseminated for many years, but there is some evidence that clinicians may not be aware of them or may not agree with all of their recommendations. In a physician questionnaire study published in 2000, Hyman and Pavlik[24] showed that in contrast to the JNC 6 recommendation to intensify therapy

when the BP level exceeds 139/89 mm Hg, 25% of physicians would not intensify therapy for a diastolic BP level of 94 mm Hg and 33% would not intensify for a systolic BP level of 158 mm Hg. In their study, 41% of physicians were not familiar with the JNC guidelines, and such nonfamiliarity was associated with higher treatment thresholds.[24] Recent improvements in meeting BP targets in the VA system[25] may attest to increasing clinician acceptance of the 140/90 mm Hg threshold to intensify therapy.

Some strengths of our study should be noted. Ours was a multisite study including many practitioners, extending the generalizability of the results. In addition, unlike previous studies on this issue, our patients were seen in a primary care environment, the setting in which most hypertension is managed.

Our study also had several limitations, however. First, we analyzed only male patients due to the predominantly male patient population in the VA system and a high percentage of patients were black. Similarly, the VA population tends to be older and have more comorbidities than the US population, and our study is no exception.

Second, our adherence data were collected by patient report, which is a limitation although this has been shown to be a reliable measure of adherence.[4,15] There are other ways to measure patient adherence to medication, including medication possession ratios, which are derived from the frequency of pharmacy fills. Recent reports have cast doubt on the continued validity of such measures in the VA system, however, especially because many prescriptions are refilled automatically by mail and thus are not a reflection of patient adherence.[26]

Third, we used clinical BP values, rather than obtaining BP through a standardized research protocol. Since we were studying clinician behavior, however, it makes sense that we should analyze the same BP values that guided the actual decisions about whether to escalate therapy. Fourth, sample size limited our ability to detect some effects, especially with regard to systolic BP.

Fifth, regression to the mean, as well as secular trends of improving BP control at the VA, probably accounted for some of the impressive reduction in mean BP level during the study (ie, from 155/80 mm Hg to 142/75 mm Hg). The VA has incorporated BP control to <140/90 mm Hg as a performance measure during the past 5 years, and the attainment of this goal in the VA has increased during this period.[25] Since all patients would be equally affected by secular trends and regression artifact, however, this does not lessen the validity of our results.

Sixth, many of our patients had only moderately uncontrolled BP; the mean BP level was 155/80 mm Hg at study inception and 142/75 mm Hg at the end of the study. We might have found a different relationship between the effects of adherence and treatment intensity on BP control in a population with more severely uncontrolled hypertension, and our results may not apply to such patients.

Finally, our methods of analysis might have muted the effects of some predictors. For example, many studies have found that patients on more vigorous regimens actually have higher BP due to confounding by indication.[27] One way around this limitation is to compare observed with expected treatment intensity, as was done by Berlowitz and colleagues[6] We did not have the data to pursue such a strategy. Although we found that any increase in the therapy was associated with a lower final diastolic BP, we might have found a more robust effect had we used methods to limit confounding by indication.

CONCLUSIONS

We studied a group of VA system patients with hypertension to elucidate the determinants of uncontrolled hypertension. Inadequate medical regimens could be blamed for a majority of uncontrolled hypertension at baseline, followed by poor adherence by the patient as the next most common reason. In longitudinal analyses, vigorous clinical management also seemed to exert a greater effect than patient adherence on BP control. The key to better BP control may, in fact, lie with the clinician.

Disclosures: This research was supported by a grant from the Department of Veterans Affairs Health Services Research and Development Service (TRH01–038, N. Kressin, PI). Dr Berlowitz reports receiving a research grant from Bristol-Myers Squibb. Dr Rose is supported by a grant from the Veterans Administration Department of Academic Affairs, as well as a Physician Training Award (PTAPM-97–185–04) from the American Cancer Society. Dr Kressin is supported by a Research Career Scientist award from the Department of Veterans Affairs, Health Services Research & Development (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.

REFERENCES

1 Ong KL, Cheung BM, Man YB, et al. Prevalence, awareness, treatment, and control of hypertension among United States adults 1999–2004. *Hypertension.* 2007;49:69–75.
2 DiMatteo MR, Giordani PJ, Lepper HS, et al. Patient adherence and medical treatment outcomes: a meta-analysis. *Med Care.* 2002;40:794–811.
3 Haynes RB, Taylor DW, Sackett DL, et al. Can simple clinical measurements detect patient noncompliance?

Hypertension. 1980;2:757–764.
4 Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care*. 1986;24:67–74.
5 Becker DM, Raqueno JV, Yook RM, et al. Nurse-mediated cholesterol management compared with enhanced primary care in siblings of individuals with premature coronary disease. *Arch Intern Med*. 1998;158:1533–1539.
6 Berlowitz DR, Ash AS, Hickey EC, et al. Inadequate management of blood pressure in a hypertensive population. *N Engl J Med*. 1998;339:1957–1963.
7 el-Kebbi IM, Ziemer DC, Musey VC, et al. Diabetes in urban African-Americans. IX. Provider adherence to management protocols. *Diabetes Care*. 1997;20:698–703.
8 Phillips LS, Branch WT, Cook CB, et al. Clinical inertia. *Ann Intern Med*. 2001;135:825–834.
9 Kaplan NM. Resistant hypertension. *J Hypertens*. 2005;23:1441–1444.
10 Moser M, Setaro JF. Clinical practice. Resistant or difficult-to-control hypertension. *N Engl J Med*. 2006;355:385–392.
11 Kressin NR, Wang F, Long J, et al. Hypertensive patients' race, health beliefs, process of care, and medication adherence. *J Gen Intern Med*. 2007;22:768–774.
12 The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. *Arch Intern Med*. 1997;157:2413–2446.
13 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.
14 Department of Veterans Affairs. *Clinical Practice Guidelines: Hypertension*; 2006.
15 Choo PW, Rand CS, Inui TS, et al. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. *Med Care*. 1999;37:846–857.
16 Appel LJ. The verdict from ALLHAT–thiazide diuretics are the preferred initial therapy for hypertension. *JAMA*. 2002;288:3039–3042.
17 Garg JP, Elliott WJ, Folker A, et al. Resistant hypertension revisited: a comparison of two university-based cohorts. *Am J Hypertens*. 2005;18:619–626.
18 Kaplan NM. Diuretics as a basis of antihypertensive therapy. An overview. *Drugs*. 2000;59(suppl 2):21–25 [discussion 39–40].
19 Moser M. Why are physicians not prescribing diuretics more frequently in the management of hypertension? *JAMA*. 1998;279:1813–1816.
20 Borzecki AM, Glickman ME, Kader B, et al. The effect of age on hypertension control and management. *Am J Hypertens*. 2006;19:520–527.
21 Yakovlevitch M, Black HR. Resistant hypertension in a tertiary care clinic. *Arch Intern Med*. 1991;151:1786–1792.
22 Javors JR, Bramble JE. Uncontrolled chronic disease: patient non-compliance or clinical mismanagement? *Dis Manag*. 2003;6:169–178.
23 Moser M, Franklin S. Hypertension management: results of a new national survey for the hypertension education foundation: Harris interactive. *J Clin Hypertens (Greenwich)*. 2007;9:316–323.
24 Hyman DJ, Pavlik VN. Self-reported hypertension treatment practices among primary care physicians: blood pressure thresholds, drug choices, and the role of guidelines and evidence-based medicine. *Arch Intern Med*. 2000;160:2281–2286.
25 Jha AK, Perlin JB, Kizer KW, et al. Effect of the transformation of the Veterans Affairs Health Care System on the quality of care. *N Engl J Med*. 2003;348:2218–2227.
26 Rothendler J, Fincke G, Reisman J, et al. Problems in calculating medication adherence from refill intervals [Abstract]. *VA Health Services Research & Development Meeting*; 2007.
27 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.

Effectiveness of Warfarin among Patients with Cancer

Adam J. Rose, MD[1,2,5], Jeff P. Sharman, MD[3], Al Ozonoff, PhD[4], Lori E. Henault, MPH[1], and Elaine M. Hylek, MD, MPH[1]

[1]Department of Medicine, Research Unit-Section of General Internal Medicine, Boston University School of Medicine, Boston Medical Center, Boston, MA, USA; [2]Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Bedford, MA, USA; [3]Department of Medicine, Division of Oncology, Stanford University, Stanford, CA, USA; [4]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; [5]Section of General Internal Medicine, 91 E. Concord Street, Suite 200, Boston, MA 02118-2644, USA.

BACKGROUND: Among patients treated with warfarin for venous thromboembolism (VTE), cancer patients have more thrombotic and hemorrhagic events than patients without cancer. Is this also the case when cancer patients are anticoagulated for other indications?

OBJECTIVE: The objective of the study is to evaluate the effectiveness of warfarin, given for any indication, among patients with cancer in a community setting.

METHODS: We identified patients with cancer from a larger prospective cohort of 6,761 patients from 101 clinical sites in the United States, matched to controls without cancer. The proportion of time spent in the therapeutic range, international normalized ration (INR) variability, and the rate of thromboembolic and major hemorrhagic events were compared between the two groups.

RESULTS: Ninety-five patients undergoing treatment for cancer were matched to 283 patients without cancer. The cancer group spent less time in the target INR range (54 vs 66%, $P<.001$) and had more variable INR values (standard deviation around the mean INR value 1.30 vs 0.71, $P<.001$). There were more thrombotic events in the cancer group than in the control group (5 vs 0 events, $P<.001$). These analyses were repeated after excluding all of the patients anticoagulated for VTE; the results were unchanged.

CONCLUSIONS: Compared to matched controls, cancer patients receiving warfarin spend less time in the target INR range, have more variable INR values, and have more thrombotic events. These effects are not dependent on whether the patient is anticoagulated for VTE or another indication.

KEY WORDS: warfarin; anticoagulation; oncology; cancer; thrombosis.
DOI: 10.1007/s11606-007-0228-y
 2007;22:997–1002

INTRODUCTION

For many years, the mainstay of outpatient anticoagulation has been the oral vitamin K antagonist warfarin, a drug with a narrow therapeutic window. Patients with cancer may experience especially erratic control of the international normalized ratio (INR).[1] Contributing factors may include drug interactions, fluctuations in dietary vitamin K intake, therapy interruptions, hepatic dysfunction, mucositis, and diarrhea and the hypercoagulable state induced by the cancer itself.[2–10]

In response to two randomized controlled trials that convincingly showed the superiority of low molecular weight heparin (LMWH) in the treatment of VTE among cancer patients,[11–13] national expert consensus panels have recommended the use of LMWH for patients with VTE and cancer for at least the first 3 to 6 months of long-term treatment.[14,15] In the study by Lee and colleagues, patients assigned to LMWH had less recurrent VTE (HR 0.48) and decreased mortality (HR 0.50), with no significant difference between groups in the rate of bleeding.[11,12] However, while these studies provide high-quality evidence that LMWH is superior to warfarin in cancer patients with VTE, there is little evidence regarding the effectiveness of warfarin in cancer patients when it is used for other indications than the treatment of VTE.

The goal of our study was to compare INR control and clinical outcomes between patients with and without cancer, regardless of the indication for warfarin. The Anticoagulation Consortium to Improve Outcomes Nationally (ACTION) study is a large prospective cohort study designed to assess the management of warfarin in community practice within the United States. From this cohort, we identified a population of patients undergoing active treatment for a malignancy to study the effectiveness of warfarin in this setting.

METHODS

The Anticoagulation Consortium to Improve Outcomes Nationally

Study patients were identified from the 6,761 patients enrolled in a prospective cohort study to assess anticoagulation care in the United States, the ACTION cohort, which has been described elsewhere.[16] Briefly, physician practices that were registered users of CoumaCare® software (Bristol-Myers Squibb, Princeton, NJ) were invited to participate by letter and through a study website. CoumaCare® is a software program that was freely available and used by many anticoagulation management services for clinical purposes such as patient tracking, data entry, and record keeping.

All of the 101 participating sites had at least one dedicated provider managing warfarin, usually within the setting of a

Received October 3, 2006
Revised March 6, 2007
Accepted April 18, 2007
Published online May 3, 2007

community-based, physician group practice. Patients were invited to participate by letter or in person (at the time of a routine appointment). To be eligible, patients had to be 18 years of age or older and provide written informed consent. For all centers, McKessonHBOC, the data management firm, provided individual on-site training about how to recruit patients, obtain consent, and transmit data. Adverse event reporting was mandatory and study personnel were trained to carry out such reporting in accordance with federal regulatory requirements. Enrollment began in April 2000 and ended in February 2002.

Data Management

Encrypted data from each site were transmitted to the data-coordinating center weekly. Data included demographic information, indication for warfarin therapy, INR target range, medical diagnoses, INR values, warfarin dose, and patient clinical management progress notes. Missing data fields and data entry errors were flagged and resolved directly with the sites by McKessonHBOC before data were transferred to the study investigators. Any interval of 45 days or more without INR testing or any INR value >10 or <0.8 triggered a direct query from the data coordinating center. Resolution of the flag relating to the INR testing interval required validation of continued warfarin use and confirmation that the gap was not related to an adverse event. Study investigators were blinded to the identification and location of participating practices and patients.

Identification and Eligibility of Case-Patients with Cancer

All 102,728 progress notes of the 6,761 enrolled patients were reviewed for documented evidence of an active malignancy. To be eligible for inclusion, patients had to be actively engaged in treatment with chemotherapeutic or hormonal agents, radiation therapy, or palliative care. Patients whose cancer was effectively treated with surgery alone were excluded, as were patients with an observation period of less than 30 days. We also excluded patients whose target INR range encompassed a value lower than 2.0 or higher than 3.5.

The active treatment period for cancer ("window period") was determined by independent review of the patient's anticoagulation record by two investigators (AJR, EMH), each blinded to the INR data. INR data were censored at the time of cessation of warfarin therapy, transfer of medical care, death of the patient, or study site termination. For patients newly starting warfarin, the first 30 days of INR data were excluded to minimize the variability attributable to the initiation phase. Warfarin interruptions were quantified and divided into two categories: procedure or bleeding/missed dose.

Identification of Control Patients

Each case was matched to three randomly selected control patients without cancer. Controls were drawn from the same study site as the case and were matched on age within 5 years. We excluded potential controls if they had less time in the database than the corresponding case; the window period for each control was, by definition, precisely the same number of days in length for each case and its matched controls. Patients with an INR target range encompassing values below 2.0 or above 3.5 were also excluded. For two of the case-patients, only two eligible controls were available. For patients who were new to warfarin therapy, we excluded their INR data until they had been on warfarin for 30 days.

Outcomes

Primary outcomes included proportion of time spent in the therapeutic range and variability in the INR measurements during the window period. Secondary outcomes included major hemorrhage and thromboembolic events during the window period. Major hemorrhage was defined as a fatal event, an event requiring hospitalization with transfusion of at least two units of packed red blood cells or bleeding involving a critical anatomical site such as the cranium or the retroperitoneum.

Statistical Analysis

All analyses were performed using the SAS statistical system, version 9.1 (SAS Institute). To assess significance of effects when comparing categorical variables with the matched design, we used Monte Carlo permutation methods with 10,000 iterations to compute empirical *P* values. Case-control status within each "cluster" of matched observations was randomly permuted 10,000 times, with a test statistic (e.g., Pearson's chi-squared statistic) calculated upon each iteration. This was used as a reference distribution, under the null hypothesis of no association with case status, to compute the empirical *P* value. Groups were compared on continuous variables using a generalized linear model (PROC GENMOD) to account for correlation between each case and its matched controls.

We used three methods to compare INR values between the case and control groups. First, we compared the proportion of INR measurements in the following ranges: <2, 2.0–3.0, 3.1–4.0, and >4.0, using a chi-square test. Second, we used the method described by Rosendaal, et al.[17] to compare the proportion of time spent within the INR target range. The Rosendaal method uses linear interpolation to impute INR values for days between measurements. INR values are not interpolated for gaps of greater than 56 days between measurements. A linear model was used to compare the mean proportion of time spent within the INR range between the two groups during the window period, accounting for matching.

We compared the variability in INR between groups during the window period, using the method described by Fihn, et al.[18,19] This method calculates the standard deviation of INR measurements over time for each patient as a measure of variability. These data were log-transformed before comparison to accommodate the distributional requirements of the linear model. Figure 1 illustrates the concept of INR variability using the INR values of two patients from our sample; sigma is the standard deviation.

We used the Monte Carlo method described above to compare the number of thromboembolic events between the two groups during the window period. In addition, we constructed a multivariable linear regression model to investigate the determinants of INR variability among the patients in the cancer group.

Additional Analyses

We recognize that the literature has already shown that the use of warfarin is problematic in cancer patients with VTE.[5,11–13]

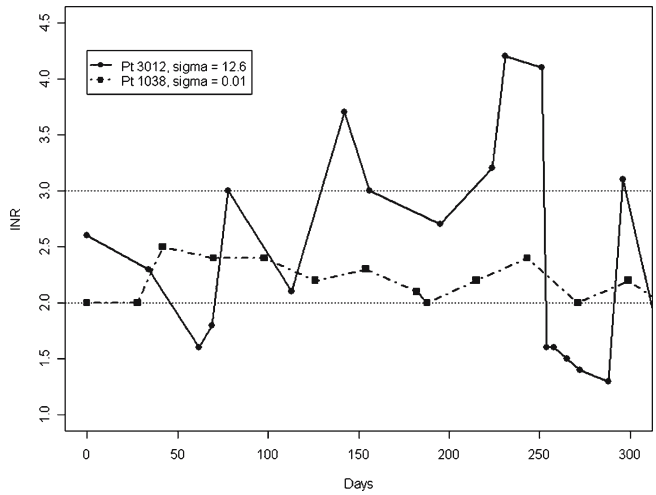


Figure 1. Graphic depiction of INR control in a patient with very low INR variability and a patient with very high INR variability.

Therefore, we were concerned that any differences between the cancer and control groups might be driven by the subset of patients anticoagulated for VTE. We addressed this concern by creating an additional, smaller dataset, composed only of the cases anticoagulated for non-VTE indications and matched controls who were also anticoagulated for non-VTE indications. We repeated our analyses using this smaller dataset; the results obtained using both datasets are presented side by side for comparison.

Study Approval and Funding

The study protocol was approved by Western Institutional Review Board® (WIRB®) of Olympia, WA and by local review boards where they existed. The funding source had no role in the collection, analysis, or interpretation of the data or in the decision to submit the study for publication.

RESULTS

Of the 6,761 patients, 127 patients were identified as being actively treated for cancer. Of these 127 patients, 95 met the inclusion criteria, and 32 were excluded (17—Surgery alone, 10—less than 30 days of data, 3–INR target range inappropriate, 2—transferred warfarin management to their oncologist). These 95 patients were matched to 283 randomly selected control patients. Of these 95 patients, 65 were anticoagulated for non-VTE indications; those 65 patients were matched to 193 control patients and were also anticoagulated for non-VTE indications to create our smaller, secondary dataset.

The baseline characteristics of the case and control groups from the larger dataset are detailed in Table 1. The two groups were similar in terms of age, gender, ethnicity, target INR range, and mean weekly warfarin dose. Although there was a statistically significant difference in the total amount of follow-up time in the overall cohort per patient (10.9 vs 13.1 months), the size of this difference was small. The median window periods in the cases and controls, corresponding to the period of active cancer therapy in the case patients, were 6.4 and 6.3 months long, respectively. The cancer patients were more likely than the control patients to be on warfarin due to a venous thrombotic event (31 vs 13%) and to have had more INR measurements per patient-month during the window period (2.41 vs 1.46).

In the cancer group, there were 0.16 interruptions of warfarin therapy for a procedure per patient-month, versus 0.04 in the controls ($P<.001$). The majority of patients with any procedure interruptions had only one such episode: 64% in the

Table 1. Baseline Characteristics in the Two Groups*

Characteristic	Cancer group (*n*=95)	Control group (*n*=283)	*P* value
Age (year)	73.5±8.95	73.3±8.71	.23
Sex (%)			.80
Male	51 (54%)	164 (58%)	
Female	44 (46%)	119 (42%)	
Race or ethnic group (%)			.95
Non-Hispanic white	87 (92%)	261 (92%)	
Non-Hispanic black	2 (2%)	6 (2%)	
Hispanic	3 (3%)	5 (2%)	
Other	3 (3%)	11 (4%)	
Target INR range (%)			.14
2.0–3.0	87 (92%)	246 (87%)	
2.5–3.5	8 (8%)	37 (13%)	
Mean time from first INR to last INR, months	10.9	13.1	<.001
Mean weekly warfarin dose, mg	30.97	30.14	.63
Mean frequency of INR measurements per 30 days	2.41	1.46	<.001
Mean interruptions of warfarin per patient-month			
Interruptions for procedures	0.16	0.04	<.001
Interruptions for bleeding or missed dose	0.07	0.05	.18
Indication for warfarin (%)			<.001
Atrial fibrillation	42 (44%)	155 (55%)	
Venous thromboembolism	29 (31%)	39 (13%)	
Prosthetic heart valve	13 (14%)	34 (12%)	
Stroke	9 (10%)	28 (10%)	
Other	2 (2%)	26 (9%)	
Incident cancer†	51 (54%)	NA	NA
Type of Cancer			
Prostate	20 (21%)	NA	NA
Breast	18 (19%)	NA	NA
Lung	8 (8%)	NA	NA
Blood-Borne (Myeloma, Leukemia, Lymphoma)	7 (7%)	NA	NA
Colon	6 (6%)	NA	NA
Bladder	3 (3%)	NA	NA
Liver	2 (2%)	NA	NA
Sarcoma	2 (2%)	NA	NA
Esophagus	1 (1%)	NA	NA
Kidney	1 (1%)	NA	NA
Melanoma	1 (1%)	NA	NA
Could not be determined by chart review	26 (27%)	NA	NA
Treatment received for cancer‡			
Surgery	24 (26%)	NA	NA
Chemotherapy	47 (49%)	NA	NA
Radiation therapy	31 (33%)	NA	NA
Hormonal therapy	21 (22%)	NA	NA
Palliative therapy	9 (10%)	NA	NA

**INR International Normalized Ratio, NA not applicable*
†Cancer diagnosis occurred after patient's enrollment in the cohort.
‡Treatment received for cancer percentages do not total 100% due to multiple therapies in some patients.

Table 2. Comparison of INR Control, Major Bleeding Events, and Thromboembolic Events

Parameters	Includes patients with VTE			Does not include patients with VTE		
	Cancer group (*n*=95)	Control group (*n*=283)	*P* value	Cancer group (*n*=65)	Control group (*n*=193)	*P* value
Frequency of INR by category (%)			<0.001			<.001
1.9 or less	406 (27%)	511 (19%)		234 (25%)	366 (19%)	
2.0–3.0	765 (51%)	1680 (61%)		481 (52%)	1153 (62%)	
3.1–4.0	214 (14%)	423 (15%)		126 (14%)	280 (15%)	
4.1 or more	128 (8%)	123 (5%)		85 (9%)	70 (4%)	
Interpolated time in INR range (Rosendaal method)						
Below target range	26% (23–30%)	19% (16–22%)	<0.001	26% (22–31%)	19% (16–21%)	.003
Within target range	54% (49–58%)	66% (64–69%)	<0.001	54% (48 –59%)	69% (66 –72%)	<.001
Above target range	20% (16–23%)	15% (12–17%)	0.02	20% (16 –25%)	12% (10 –15%)	.002
INR variability (Fihn method)						
Mean INR in each group	2.54 (2.45–2.64)	2.55 (2.49–2.60)	0.94	2.59 (2.49–2.71)	2.51 (2.45–2.56)	.13
Standard deviation in each group	1.31 (0.99–1.63)	0.71 (0.54–0.89)	<0.001	1.23 (0.95–1.52)	0.64 (0.48–0.81)	<.001
major bleeding events	1	0	0.25	0	0	NA
Thrombotic events	5	0	<0.001	4	0	.005

Ranges in parentheses are 95% confidence intervals. The 95-case group includes patients with VTE, the 65-case group does not.
NA Not applicable

cancer group and 81% among the controls. Brief interruptions of warfarin therapy for minor bleeding episodes or missed doses were similar between the two groups (0.07 vs 0.05 interruptions per patient-month, *P*=.18).

Analysis of INR Control and INR Variability

Table 2 shows a comparison of measures of INR control between groups. The mean INR values achieved in the two groups were nearly identical (2.54 and 2.55). However, the cancer group had greater INR variability (SD 1.31 vs 0.71, *P*<.001), spent proportionately less time in the therapeutic range (54 vs 66%, *P*<.001), and more time above (20 vs 15%, *P*=.02) and below (26 vs 19%, *P*<.001) the target range. These results were nearly identical when the analyses were repeated with the non-VTE dataset.

In the multivariable analysis of determinants of INR variability in the cancer group (Table 3), the receipt of chemotherapy and any interruption of warfarin therapy for a procedure were significantly associated with increased INR variability (2.0- and 1.4-fold increases, respectively). Older patients did not have especially high variability. The receipt of radiation therapy and hormonal therapy did not predict increased INR variability. Interrupting warfarin therapy due to a bleeding event, rather than for a procedure, did not predict increased INR variability. Again, these results were nearly identical when the VTE patients were excluded.

Thrombotic Events and Major Bleeding Events

Within the window period, 5 of the 95 patients (5.3%) in the cancer group sustained a thromboembolic event; 4 of these occurred among patients taking warfarin for atrial fibrillation (2 strokes and 2 pulmonary emboli) and another pulmonary embolus occurred in a patient with a prior deep venous thrombosis. For four of these patients, the event occurred during a period of warfarin interruption. No patient in the control group experienced a thromboembolism during the window period (*P*<.001 for a difference between groups). In the non-VTE dataset, there were four thrombotic events among the cases and none among the controls (*P*=.005). Three of these four events occurred during a period of warfarin interruption.

Major bleeding occurred in one case-patient with VTE (subdural bleed during bridging therapy with low molecular weight heparin). Another case-patient (non-VTE) was hospitalized with an acute gastrointestinal hemorrhage and was taken

Table 3. Multivariable Model to Predict INR Variability in the Cancer Patients

Parameters	Includes patients with VTE			Does not include patients with VTE		
	Beta coefficient	95% CI	*P* value	Beta coefficient	95% CI	*P* value
Female gender	0.057	(−0.24–0.35)	.70	0.188	(−0.14–0.51)	.25
Age≥75	0.088	(−0.21–0.39)	.57	0.209	(−0.12–0.54)	.21
Chemotherapy	**0.713**	**(0.37–1.06)**	**<.001**	**0.784**	**(−0.41–1.15)**	**<.001**
Radiation therapy	−0.102	(−0.43–0.22)	.54	−0.274	(−0.60–0.06)	.10
Hormonal therapy	0.058	(−0.32–0.44)	.77	−0.195	(−0.61–0.22)	.36
Any procedure interruptions (dichotomous)	**0.340**	**(0.03–0.65)**	**.03**	**0.363**	**(0.04–0.69)**	**.03**
Any bleeding interruptions (dichotomous)	0.105	(−0.22–0.42)	0.52	0.124	(−0.24–0.49)	.51

The R^2 for the model was .20 with the larger group and .35 with the smaller group. The effect sizes for chemotherapy and any procedure interruptions are 2.0- and 1.4-fold increases in variability in both models.
The model results on the left include the VTE patients (n=95), while those on the right exclude those patients (n=65). Boldfaced variables are significant at the .05 level.

off warfarin, but we were unable to validate the number of packed cell transfusions. No bleeding occurred among the control patients.

DISCUSSION

In this community-based cohort of patients receiving warfarin for varied indications, we found that patients with an active malignancy spent less time in their target INR ranges, had more variable INR values, received more frequent INR testing, and experienced more thrombotic events than patients without a malignancy. When we repeated our analyses after excluding patients anticoagulated for VTE, these results remained essentially unchanged. This suggests that the use of warfarin in cancer patients is challenging, regardless of the indication for anticoagulation.

While other investigational and observational studies have investigated the treatment of VTE in cancer patients,[5,11,13,20,21] to our knowledge, none have examined the effectiveness of warfarin for non-VTE indications. In our study, 69% of cancer patients were on chronic warfarin therapy for indications other than VTE, and four of the five thrombotic events in the cancer group occurred in patients who were anticoagulated for atrial fibrillation. Of the five thromboembolic events in the cancer group, four occurred during a period of warfarin interruption, which emphasizes the importance of this factor in the difficulty of managing warfarin in cancer patients. We also found an association between the receipt of chemotherapy and increased INR variability in the cancer group, which is consistent with prior observations.[2–8,22–24]

The INR control achieved among our noncancer group, 66% time-in-range, is similar to that reported for patients enrolled in randomized trials (66.4%) and anticoagulation clinics (65.5%).[25] Because our control patients were matched to the site of the case, the 54% time-in-range of the cancer patients further highlights the difficulties in warfarin management for this patient group. In a similar study, Palareti et al.[5] identified 95 patients with cancer from within a large, Italian population-based cohort study of patients receiving an oral vitamin K antagonist for treatment of VTE. The outcomes for these 95 patients were compared to the outcomes of 733 patients without cancer. The investigators found that the cancer patients had higher rates of major bleeding (5.4 vs 0.9%) and a trend toward higher rates of thrombotic complications (6.8 vs 2.5%, P=.058). In their study, cancer patients also spent a greater proportion of time above the target INR range (8.8 vs 4.5%), although this difference was not statistically significant.

Some limitations of our study should be noted. Despite the robust size of the overall cohort, we only identified 95 patients with an active cancer. Our identification of case-patients was contingent on the quality of documentation in the anticoagulation clinical notes. However, because the diagnosis and treatment of cancer would be so germane to warfarin therapy, we feel confident that the notes would have documented such information. Because signed informed consent was necessary to participate, we may have selected for a healthier group of patients. This would be expected to bias toward the null, suggesting that our findings may be an underestimate of the effect of cancer on anticoagulation.

Our study has several important strengths. The omission of the first 30 days of warfarin therapy excluded a potential source of bias due to INR variability during the initiation of therapy. The overall cohort was representative of the wide variety of patients treated with warfarin in the community. A manual review of every clinical record in the database minimized the likelihood of missed clinical events or assignment bias. Our study used a careful matching procedure to minimize between-group variability relating to site of care, age, and overall observation time on warfarin. Finally, our additional analysis, performed only among patients without VTE, excludes the possibility that the patients with VTE are driving our observed differences.

Because chemotherapy is such a potent risk factor for INR fluctuation, patients and providers need to be particularly vigilant during these periods to minimize the variability induced by exogenous factors such as known potentiating medications. Consistent dietary intake of vitamin-K rich foods may help to offset fluctuations in INR in patients prone to erratic control.[26,27] Risks and benefits associated with bridging therapy during periods of warfarin interruption need to be better defined in this patient population. Newer anticoagulant drugs with a wider therapeutic index and less potential for medication and dietary interactions will hopefully translate into improved effectiveness of anticoagulation for the prevention and treatment of thromboembolism among patients with malignancy.

In conclusion, we found that in a community setting, cancer patients receiving warfarin for diverse indications had more erratic control of their anticoagulation than matched controls without cancer. They also had a significantly increased rate of thrombotic complications. Among the cancer patients, the receipt of chemotherapy was associated with increased INR variability. When we repeated our analyses after excluding VTE patients, these findings were essentially unchanged, suggesting that the difficulty of managing warfarin in cancer patients is not limited to those with VTE. Our findings extend the results of previous studies regarding the treatment of VTE in cancer patients and suggest that warfarin may not be the optimal anticoagulant for cancer patients for any indication, especially during treatment with chemotherapy.

Acknowledgements: *This study was funded by Bristol-Myers Squibb, the makers of Coumadin® brand warfarin. Bristol-Myers Squibb had no role in the design and conduct of the study, or in the collection, analysis, and interpretation of the data, or in the preparation, review, and approval of the manuscript. Dr. Rose is supported by a grant from the Department of Veterans Affairs Office of Academic Affairs. The opinions expressed in this manuscript do not necessarily represent the views or policies of the Department of Veterans Affairs. Dr. Rose is also supported by a Physician Training Award (PTAPM-97-185-04) from the American Cancer Society.*

Conflicts of Interest: *Dr. Hylek has served as a consultant to and received research support from Bristol-Myers Squibb and AstraZeneca. None of the other authors report any potential conflicts of interest.*

Corresponding Author: *Adam J. Rose, MD; Section of General Internal Medicine, 91 E. Concord Street, Suite 200, Boston, MA 02118-2644, USA (e-mail: adamrose@bu.edu).*

REFERENCES

1. **Lee AY, Levine MN.** Venous thromboembolism and cancer: risks and outcomes. Circulation. 2003;107:I17–21.
2. **Bona RD, Sivjee KY, Hickey AD, Wallace DM, Wajcs SB.** The efficacy and safety of oral anticoagulation in patients with cancer. Thromb Haemost. 1995;74:1055–8.
3. **Gitter MJ, Jaeger TM, Petterson TM, Gersh BJ, Silverstein MD.** Bleeding and thromboembolism during anticoagulant therapy: a population-based study in Rochester, Minnesota. Mayo Clin Proc. 1995;70:725–33.
4. **Landefeld CS, Goldman L.** Major bleeding in outpatients treated with warfarin: incidence and prediction by factors known at the start of outpatient therapy. Am J Med. 1989;87:144–52.
5. **Palareti G, Legnani C, Lee A, et al.** A comparison of the safety and efficacy of oral anticoagulation for the treatment of venous thromboembolic disease in patients with or without malignancy. Thromb Haemost. 2000;84:805–10.
6. **Prandoni P, Lensing AW, Piccioli A, et al.** Recurrent venous thromboembolism and bleeding complications during anticoagulant treatment in patients with cancer and venous thrombosis. Blood. 2002;100:3484–8.
7. **Wester JP, de Valk HW, Nieuwenhuis HK, et al.** Risk factors for bleeding during treatment of acute venous thromboembolism. Thromb Haemost. 1996;76:682–8.
8. **White RH, Beyth RJ, Zhou H, Romano PS.** Major bleeding after hospitalization for deep-venous thrombosis. Am J Med. 1999;107:414–24.
9. **Hylek EM, Heiman H, Skates SJ, Sheehan MA, Singer DE.** Acetaminophen and other risk factors for excessive warfarin anticoagulation. JAMA. 1998;279:657–62.
10. **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.
11. **Lee AY, Levine MN, Baker RI, et al.** Low-molecular-weight heparin versus a coumarin for the prevention of recurrent venous thromboembolism in patients with cancer. N Engl J Med. 2003;349:146–53.
12. **Lee AY, Rickles FR, Julian JA, et al.** Randomized comparison of low molecular weight heparin and coumarin derivatives on the survival of patients with cancer and venous thromboembolism. J Clin Oncol. 2005;23:2123–9.
13. **Meyer G, Marjanovic Z, Valcke J, et al.** Comparison of low-molecular-weight heparin and warfarin for the secondary prevention of venous thromboembolism in patients with cancer: a randomized controlled study. Arch Intern Med. 2002;162:1729–35.
14. **Buller HR, Agnelli G, Hull RD, Hyers TM, Prins MH, Raskob GE.** Antithrombotic therapy for venous thromboembolic disease: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest. 2004;126:401S–4028S.
15. Practice Guidelines in Oncology-Venous Thromboembolic Disease: National Comprehensive Cancer Network; 2006.
16. **Hylek EM.** Trends in anticoagulation management across community-based practices in the United States: the Anticoagulation Consortium to Improve Outcomes Nationally (ACTION) study. J Thromb Thrombolysis. 2003;16:83–6.
17. **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.
18. **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;124:970–9.
19. **Kent DL, Vermes D, McDonell M, Henikoff J, Fihn SD.** A model for planning optimal follow-up for outpatients on warfarin anticoagulation. Warfarin Optimal Outpatient Follow-up Study Group. Med Decis Making. 1992;12:132–41.
20. **Elting LS, Escalante CP, Cooksley C, et al.** Outcomes and cost of deep venous thrombosis among patients with cancer. Arch Intern Med. 2004;164:1653–61.
21. **Hutten BA, Prins MH, Gent M, Ginsberg J, Tijssen JG, Buller HR.** Incidence of recurrent thromboembolic and bleeding complications among patients with venous thromboembolism in relation to both malignancy and achieved international normalized ratio: a retrospective analysis. J Clin Oncol. 2000;18:3078–83.
22. **Levine MN, Gent M, Hirsh J, et al.** The thrombogenic effect of anticancer drug therapy in women with stage II breast cancer. N Engl J Med. 1988;318:404–7.
23. **Shah HR, Ledbetter L, Diasio R, Saif MW.** A retrospective study of coagulation abnormalities in patients receiving concomitant capecitabine and warfarin. Clin Colorectal Cancer. 2006;5:354–8.
24. **Zangari M, Anaissie E, Barlogie B, et al.** Increased risk of deep-vein thrombosis in patients with multiple myeloma receiving thalidomide and chemotherapy. Blood. 2001;98:1614–5.
25. **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.
26. **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.
27. **Sorano GG, Biondi G, Conti M, Mameli G, Licheri D, Marongiu F.** Controlled vitamin K content diet for improving the management of poorly controlled anticoagulated patients: a clinical practice proposal. Haemostasis. 1993;23:77–82.

SUPPLEMENT

Treatment of Alcohol and Other Drug Dependence

Richard Saitz
Departments of Medicine and Epidemiology, Youth Alcohol Prevention Center, Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Boston University and Boston Medical Center, Boston, MA

KEY CONCEPTS
Addiction (e.g., alcohol or drugs) is a chronic disease.
Specific behavioral and pharmacological treatments have proven efficacy.
Addressing addiction includes intervention for risky and problem use.
Clinicians who are not addiction specialists can play important roles in addressing addiction. ***Liver Transpl 13:S59-S64, 2007.***

Unhealthy alcohol and drug use are associated with substantial morbidity, disability, mortality, and costs.[1,2] They are among the leading preventable causes of death and disability. Most cases of esophageal cancer, 60% of chronic pancreatitis, 50% of cirrhosis and chronic hepatitis, 42% of acute pancreatitis, and substantial proportions of other medical illnesses including human immunodeficiency virus (HIV) and injuries are attributable to alcohol and other drugs. More than 100,000 people in the United States die each year as a result of alcohol and illicit drugs. In the United States, substance (alcohol and drug) use cost far more than other common medical illnesses—over $400 billion a year, compared with approximately a quarter of that sum for coronary heart disease and an eighth of that sum for obstructive lung disease and asthma.

Despite these consequences and costs, and despite the fact that over a third of hospital admissions are related to unhealthy alcohol and other drug use, most such patients do not receive addiction treatment. But efficacious treatments exist. Tobacco addiction is more common and causes more deaths than alcohol or other drug use. It can be treated effectively with brief advice, individual and group counseling, and several medications (nicotine replacement, bupropion, varenicline). But this review of addiction treatment focuses on alcohol and other drug use, excluding tobacco, a topic worthy of separate focus. For alcohol and other drug disorders, behavioral and pharmacological treatments have proven efficacy. A number of these treatments have been proven effective relatively recently, have not yet been widely disseminated into practice, and can be implemented outside of addiction specialty treatment settings.[3,4]

Finally, unhealthy use of substances by people who have not yet met diagnostic criteria for a substance use disorder can also be effectively treated with brief interventions. These interventions are of particular importance to clinicians who are not addiction specialists because they are brief and can be done by nonspecialists. These clinicians also have critical roles in facilitating receipt of addiction treatment by their patients.

But before discussing specific treatments, it is important to consider the fact that addiction is a chronic disease, so that expectations of treatment outcome are appropriate.[5] Treating addiction is perhaps more like treating and managing asthma and diabetes than it is like curing a bacterial infection.

ADDICTION IS A CHRONIC DISEASE

Alcohol and drug (substance) dependence share a number of features in common with other common chronic illnesses.[6] These include the following: physiologic basis; diagnosis; definable risk factors; heritability; poor adherence to treatment; no cure; relapse common; longitudinal care required; and denial.

Substance dependence has a physiologic basis. For example, people with alcohol dependence, even during abstinence, have reduced gamma-aminobutyric acid

Abbreviations: HIV, human immunodeficiency virus; GABA, gamma-aminobutyric acid; AA, Alcoholics Anonymous.

Address reprint requests to: Richard Saitz, Departments of Medicine and Epidemiology, Youth Alcohol Prevention Center, Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Boston University and Boston Medical Center, 91 East Concord Street 200, Boston, MA 02118-2644. Telephone: 617-414-7399; Fax: 617-414-4676; E-mail: rsaitz@bu.edu

DOI 10.1002/lt.21339
Published online in Wiley InterScience (www.interscience.wiley.com).

(GABA) receptors in the brain, as measured by single positron emission tomography, than do people without alcohol dependence. Alcohol potentiates inhibitory GABA transmission in the central nervous system. These decreases in central GABA receptors may therefore be related to the uncontrollable desire to drink seen in people with dependence. Whether these changes are a cause or effect of heavy drinking is not known. Physiologic characteristics that do appear to have causal roles are enzymatic activities that lead to higher levels of serum acetaldehyde after drinking (e.g., aldehyde dehydrogenase deficiency). These higher amounts of acetaldehyde that are unpleasant (e.g., nausea) are associated with a lower risk for alcohol dependence.

Substance dependence can be reliably diagnosed. Although there is no laboratory test, structured interviews (and expert clinicians) yield diagnoses with consistency. These diagnoses have prognostic value in terms of natural history and treatment response.

Substance dependence has definable risk factors, and like other common chronic illnesses has a substantial genetic etiology—although also similarly, there is no single gene responsible for the risk. Just over 50% of alcohol dependence is related to nonshared environmental influences (e.g., peers, bad childhood experiences). The remainder is related to genetics. In adoption studies, having a parent with alcohol dependence doubles the risk. About one-third of identical twins will have alcohol dependence if their twin has it. Epidemiologic studies have identified specific genes that increase risk. For example, genes involved in dopamine synthesis, which is critical for the reward pathway where all substances of abuse have action, increase the risk. And the gene that includes the locus that codes for alcohol dehydrogenase has been identified as a risk factor for alcohol dependence.

As with other chronic medical illnesses, treatments are effective but are often not adhered to, there are no cures, and relapse is common. Similarly, substance dependence affects physical, social, and emotional well-being, and requires longitudinal health care. After alcohol-dependence treatment, 40-60% of patients are abstinent 1 year later, and an additional 15-30% have not returned to dependent drinking. In comparison, adherence to medication regimens in diabetes, hypertension, and asthma are 30-50% and are generally even worse for nonpharmacological treatments. Relapse, or recurrence of symptoms requiring additional medical care to establish remission of symptoms of these diseases, is actually similar to that seen in substance dependence (up to 70%). And these observations are true despite that fact that in practice, substance dependence treatment is often short term. Despite this, relapse rates are not 100%, as would be the case for diabetes treated only in the short term.

Last, even one of the classic characteristics of alcoholism, denial, is a common feature among patients with other chronic illnesses like diabetes and hypertension. Denial is a common response to being "accused" of (or diagnosed with) having an undesirable characteristic.

In the United States, although there are addiction medicine specialists, most addiction treatment is delivered in specialty treatment programs by substance abuse counselors. The view from outside this system is often that patients should go to "detox." Detoxification is useful when patients are beginning to cut down or abstain, and it is valuable as a first step in treatment. Physical and emotional symptoms of withdrawal can be ameliorated by using medications cross-tolerant to the substance being used (e.g., benzodiazepines for alcohol withdrawal). But detoxification alone (most often done in ambulatory settings) is not addiction treatment, in that it does not prevent relapse. For example, <20% of patients with opioid dependence are abstinent 1 year after detoxification alone. Unfortunately, most patients who undergo detoxification in the United States do not receive further addiction treatment. Addiction treatment is aimed at decreasing the chronic consequences of uncontrolled substance use, including physical and interpersonal consequences, and social, psychological, legal, and employment problems. In addition, treatment aims to reduce substance use, including lapses (usually defined as use on one occasion) and relapses (usually defined as more than one occasion of heavy use).

TREATMENT

What Is Addiction Treatment?

Alcohol treatment includes access to psychological, medical, employment, legal, and social services, sometimes removal from a drinking or otherwise harmful environment, use of mutual (self)-help groups, pharmacotherapy, and counseling by both specialists and nonspecialists.

This counseling includes brief counseling by physicians in medical settings.[7,8] Of note, as would be the case for heart disease or gastrointestinal disease, a single lecture or review article such as this one cannot even briefly mention all known efficacious addiction treatments, and certainly cannot provide detailed indications, prescribing information, and counseling ingredients. As a result, this article focuses on principles and the best proven treatments, particularly those of relevance to those who are not addiction physicians or counselors.

How Effective Is Treatment?

At 1 year, two-thirds of patients have a reduction in alcohol consequences (injury, unemployment) and consumption (by 50%).[9] One-third are abstinent or drinking moderately without consequences. Monetary benefits of alcohol and drug treatment to society outweigh costs 4-12-fold (depending on drug and treatment type). For opioid dependence, pharmacotherapy and counseling can achieve abstinence rates of 60-80%.

Principles of Treatment

The National Institute on Drug Abuse has published 13 principles of effective treatment.[10] These are:

1. No single treatment is appropriate for all individuals. Treatments should be individualized according to patient needs.
2. Treatment needs to be readily available. Patients ready for treatment can be lost if treatment is not immediately accessible.
3. Effective treatment attends to multiple needs of the individual, not just his or her drug use. These needs include addressing substance use and any associated medical, psychological, social, vocational, and legal problems.
4. As with any chronic illness, an individual's treatment and services plan must be assessed continually and modified as necessary to ensure that the plan meets the person's changing needs.
5. Remaining in treatment for an adequate period of time is critical for treatment effectiveness. The appropriate duration for an individual depends on the problems and needs but is usually at least 3 months. Treatment should include strategies to keep patients in treatment.
6. Counseling (individual and/or group) and other behavioral therapies are critical components of effective treatment for addiction. These therapies address motivation, problem solving, relationships, social functioning, and skills useful for avoiding drug use.
7. Medications are an important element of treatment for many patients, especially when combined with counseling and other behavioral therapies.
8. Addicted or drug-abusing individuals with coexisting mental disorders should have both disorders treated in an integrated way.
9. Medical detoxification is only the first stage of addiction treatment and by itself does little to change long-term drug use.
10. Treatment does not need to be voluntary to be effective. Mandated treatment can increase treatment entry and retention.
11. Possible drug use during treatment must be monitored continuously, with results presented to patients. Monitoring helps patients resist urges to use substances, and positive results can signal a need to intensify treatment.
12. Treatment programs should provide assessment for HIV–acquired immunodeficiency syndrome, hepatitis B and C, tuberculosis, and other infectious diseases, and counseling to help patients modify or change behaviors that place themselves or others at risk of infection.
13. Recovery from drug addiction can be a long-term process and frequently requires multiple episodes of treatment, or long-term treatment.

Behavioral Treatments

Behavioral treatments for addiction with proven efficacy are not simply generic counseling sessions. A number of therapies with specific content and doses, often clearly laid out in manuals, have been proven effective in well-designed studies. Motivational interviewing (or manualized 4-session motivational enhancement therapy) is a client-centered, directive method for enhancing intrinsic motivation to change by exploring and resolving ambivalence. Cognitive behavioral therapies emphasize skills to cope with situations that might precipitate substance use. Twelve-step facilitation emphasizes the disease model of addiction and encourages and facilitates full participation in 12-step groups like Alcoholics Anonymous. Contingency management (sometimes referred to as involving motivational incentives) provides rewards for treatment adherence or negative drug testing results. Marital and family therapy can also be effective treatments for addiction. In the context of medication prescription for alcohol dependence, "medical management" by patients' physicians, physician assistants, or nurses, a relatively brief form of repeated counseling similar to what medical physicians routinely do for other medications, appears to effective.[3] Medical management involves asking about medication side effects and adherence, and encouraging abstinence.

A number of other behavioral treatments have proven efficacy. These include but are not limited to the following: relapse prevention counseling, supportive-expressive psychotherapy, individualized drug counseling, behavioral therapy for adolescents, multidimensional family therapy for adolescents, multisystemic therapy, community reinforcement approach plus vouchers, voucher-based reinforcement therapy in methadone maintenance treatment, day treatment with abstinence contingencies and vouchers, and the Matrix model. The details of these and other treatments are beyond the scope of this article.

Pharmacological Treatments

Pharmacological treatments for opioid dependence include methadone maintenance, buprenorphine, and naltrexone.[11] As with other pharmacotherapies, drug use should be monitored, and patients should participate in counseling and obtain any needed social support.

Naltrexone is a synthetic opioid antagonist that blocks the euphoria associated with opioid use. To avoid precipitating withdrawal, it is given after an opiate-free period, and it is given daily or 3 times a week. It does not prevent craving, and nonadherence is common. This medication is best used in patients who are highly motivated to maintain abstinence (such as impaired physicians or people on parole) who receive counseling and careful monitoring.

Methadone is a long-acting opioid agonist taken orally.[12,13] To be effective, it must be given over a long period of time at a dose sufficient to prevent withdrawal, block effects of illicit use, and decrease craving (generally ≥60 mg once a day). Patients stabilized on doses of methadone do not experience any euphoria from taking the drug; they feel normal and can function normally (or as normal as someone not receiving psychoactive medication can feel). In the United States, methadone can only be prescribed as a treatment for

opioid dependence by physicians in licensed methadone maintenance programs or by physicians with specific permission to prescribe it as office-based opioid therapy, currently an unusual exception. In these programs, to which access is limited, patients tend to be subject to stigma and to inconvenient and even punitive rules, such as needing to present daily at a particular time for dosing. Nonetheless, as part of a comprehensive drug treatment program, methadone treatment increases survival, increases treatment retention, decreases illicit opioid use, decreases hepatitis and HIV seroconversion, decreases criminal activity, increases employment, and improves birth outcomes. As with detoxification, >80% of patients who undergo treatment will relapse to illicit opioid use in a year.

Of note, methadone is also an excellent choice for short-term prevention of withdrawal in general hospital settings. The goal of methadone treatment in those settings is to allow the treatment of the underlying medical or surgical condition. In these cases, the dose is 20-40 mg in 24 hours, usually initially provided as 15-20 mg followed by an additional 5-10 mg after 2-3 hours, until symptoms of withdrawal abate.

Buprenorphine is a partial opioid agonist provided sublingually for maintenance.[14] With the exception of pregnant women, patients should be treated with a combination tablet of buprenorphine and naloxone. Naloxone is a short-acting opioid antagonist that is not absorbed sublingually but that prevents the tablets from being abused intravenously because it would precipitate withdrawal. Buprenorphine itself can precipitate withdrawal, so induction involves waiting for marked withdrawal symptoms (from the abused opioid) to appear, after which buprenorphine can be initiated at a low dose, then titrated higher to the usual maintenance doses of 8-16 mg once daily. Buprenorphine is quite safe, in part because the opioid effect has a ceiling: after a certain dose, no additional effect accrues. Although experience with buprenorphine is more limited than that with methadone, clinical trials have found that buprenorphine increases abstinence from illicit opioids, increases retention in treatment, and decreases opioid craving and even mortality. Some patients maintained on methadone can be switched to buprenorphine. Those less likely to succeed on buprenorphine are those requiring higher doses of methadone or requiring more structured treatment programs.

Pharmacological treatments for alcohol dependence include disulfiram, acamprosate, naltrexone, and long-acting injectable naltrexone. These medications reduce heavy drinking and increase abstinence.

Disulfiram, an inhibitor of aldehyde dehydrogenase, results in increased levels of acetaldehyde and an unpleasant reaction after consumption of ethanol. The usual oral daily dose is 500 mg. In one of the largest studies of this medication, disulfiram was no better than placebo in achieving abstinence. But it is not clear whether a placebo-controlled trial is the best way to test a drug whose efficacy depends largely on the patient knowing that they may experience a very unpleasant reaction. Of note, in post hoc analyses, the drug was more effective in those who were adherent to it. In at least 5 controlled studies, disulfiram was associated with marked increases in abstinence when administration of the drug was supervised by a concerned other. Disulfiram has numerous contraindications, and the risk-benefit ratio for people at risk of complications should they experience the ethanol reaction (e.g., those with esophageal varices) needs to be considered. High doses can lead to an idiosyncratic fulminant hepatitis and neuropathy.

Acamprosate increases continuous abstinence at 1 year by 8%, from 15% to 23%, and increases abstinent days by 27 days.[15] The mechanism of action is unclear, but the drug appears to work by affecting the glutamate system. The usual oral dose is 666 mg 3 times daily. The main side effect is diarrhea, which subsides with continued use, and the medication needs adjustment for renal insufficiency (and is contraindicated in patients with renal failure).

Naltrexone decreases relapse to heavy drinking by approximately 11% (decrease in absolute risk from 48% to 37%).[15] Naltrexone is a long-acting opioid antagonist that is absorbed when taken orally. The usual dose is 50 mg daily. The medication blocks endogenous opioids, thus decreasing the reinforcing pleasurable effects of drinking. The main side effects, nausea and dizziness, subside with continued use. Naltrexone cannot be given to patients with opioid dependence or a need for opioids. In the event of an acute need for opioids to treat pain, naltrexone should be discontinued, and in the short term, very high doses of opioids will be required under close monitoring. Monthly injections of naltrexone (380 mg intramuscularly), as established by a placebo-controlled trial, also decreases heavy drinking, and may address the problems with adherence that occur with oral pharmacotherapies that need to be provided daily or more frequently.[4] The manufacturer has a program that coordinates product delivery for all patients being treated and provides information regarding appropriate storage and administration.

Most studies of pharmacotherapies have been in patients who have completed detoxification first, but naltrexone appears to be efficacious even in patients who have a short duration of abstinence before beginning treatment. Of course, abstinence must be achieved before beginning disulfiram to avoid the disulfiram-ethanol reaction. During naltrexone or disulfiram treatment, liver enzymes should be monitored periodically. Although the drugs can cause increases in liver enzymes, most studies of alcohol dependence pharmacotherapy find decreases or no difference in levels in treated patients compared with control patients. All alcohol-dependence pharmacotherapies are category C; they should only be prescribed during pregnancy if risks will clearly outweigh benefits. Combinations of naltrexone and acamprosate do not appear to offer clear additional efficacy compared with either drug alone.

For dependence on drugs other than alcohol, tobacco, and opioids, there are no well-established pharmacotherapies, although many are under investigation.

CO-OCCURRING MENTAL HEALTH CONDITIONS

Diagnosis of co-occurring mental health problems can be challenging in people with addictions because there is substantial overlap in symptoms, and sometimes the temporal relations are difficult to sort out. Nonetheless, patients with co-occurring mental health conditions should have the conditions treated, regardless of whether the condition preceded the addiction or not. Psychiatric illness can interfere with adherence to and participation in addiction treatment, and it can trigger relapse. Psychiatric treatment can decrease these effects. For example, for those with anxiety disorders, buspirone can decrease heavy alcohol consumption. Fluoxetine is similarly effective in those with alcohol dependence and major depression.

ADDRESSING ADDICTION

What Is Risky and Problem Use?

Excessive use of alcohol (e.g., >14 standard drinks per week or 4 drinks per occasion by men, 7 and 3 for women and the elderly) and use of drugs by people who do not meet criteria for substance dependence is more common than addiction. Problem use describes people who are using substances (or drinking heavily) and experiencing consequences of that use but who do not meet dependence criteria. People with risky use have not yet experienced consequences but are at risk (e.g., excessive alcohol use, any illicit drug use).

How Can Risky and Problem Use Be Identified?

Risky and problem use should be identified because brief intervention has efficacy (proven for alcohol, some evidence for efficacy for drug use), and because the prevalence is higher than that of dependence. Brief intervention can prevent future use and can likely decrease consequences. Screening tests can identify risky and problem use, and all adults should be screened. A single question can identify risky alcohol use: "How many times in the past year have you had 5 (4 for women) or more drinks in a day?" (a positive test is one or more times). Although screening tests are less well validated for drug use, several questionnaires have been developed, and a consensus panel recommended the following single question as a screening test: "Have you ever used street drugs more than 5 times in your life?" Longer screening questionnaires include the Alcohol Use Disorders Identification Test, a 2-item conjoint (drug and alcohol) screening test, the World Health Organization Alcohol Smoking and Substance Involvement Screening Test. Laboratory tests are generally insensitive and nonspecific when used for screening for unhealthy alcohol use. Laboratory testing for drug use is similarly not very useful for screening patients because of short half-lives and the need to test for many drugs of abuse.

Brief Intervention

"Brief intervention" generally refers to 10-15 minutes of counseling, with feedback about use, advice, and goal setting, and follow-up contact with a clinician. The advice should be appropriate to the patient's readiness to change. Randomized trials in diverse clinical settings have found that brief interventions can reduce risky drinking amounts by 11%. Decreased heroin and cocaine use may result from brief intervention.[16] Although not consistently found in controlled studies, brief intervention may also decrease consequences of alcohol and other drug use, including serum gamma-glutamyltransferase levels, hospitalization, and death.

Most People With Addictions Do Not Receive Treatment

Most people with addictions do not receive effective treatment. Most are not identified in medical settings. Even when patients with addictions are identified in medical settings, they often do not receive any efficacious interventions. And even when patients enter addiction care via detoxification, the initial steps are most often not followed by efficacious addiction treatment. For example, 80% of people with opioid addiction do not receive medication-assisted treatment.

Screening and Brief Intervention

Clinicians who are not addiction specialists can play critical roles in improving identification and management of patients with addictions. Universal screening can identify patients with addictions. Brief interventions can reduce substance use and/or recommend linkage to addictions care. Follow-up and support of specialty addiction treatment plans can contribute to relapse prevention efforts.[1]

Referral to Mutual Help Groups

Mutual or self-help groups such as Alcoholics Anonymous (AA) provide social support and an alcohol- and drug-free social network. Observational studies suggest that AA can increase abstinence for people with alcohol dependence who participate. Clinicians can refer patients to these groups, and they can help by suggesting that patients try groups until they find one they are comfortable with, and by asking about meeting attendance and participation.[1]

Pharmacotherapy for Alcohol Dependence

Clinicians who are not addiction specialists can become familiar with prescribing pharmacotherapies for alcohol dependence, and can then do so while having addiction specialists manage the many other interventions needed for successful addiction treatment (e.g., counseling, assistance with employment or housing).

Buprenorphine

Clinicians who are not addiction specialists can also become familiar with prescribing buprenorphine for opioid

dependence and prescribe this treatment. In the United States, a waiver is required from the Drug Enforcement Administration.[17,18] Obtaining the waiver requires certification in addiction medicine or psychiatry, or evidence of training (at least 8 hours) specifically acceptable for this purpose. Then, as with pharmacotherapy for alcohol dependence, the clinician can prescribe, knowing that other aspects of addiction treatment are addressed either in their practice or by addiction specialty clinicians or programs. Prescribing this effective opioid treatment in a medical office represents a dramatic contrast with the requirement to attend a program to receive methadone. Buprenorphine has the potential to be more accessible to patients than methadone because it can be prescribed in a doctor's office. Access is currently limited in the United States by the numbers of physicians prescribing this treatment and by a regulatory limit of 30 patients per physician, which can be increased to 100 after the first year.

Integrating Care

Clinicians who are not addiction specialists can help patients receive the care that they need over time. Patients with addictions often require coordinated, integrated care for addiction, mental health, and medical problems. These services are often delivered in different places by different clinicians, in short-term programs, leading to fragmented, uncoordinated care, and increasing the risk of errors and suboptimal care. Attention to the whole clinical picture and monitoring of all of the care received (e.g., case management) can improve care. Systems that integrate medical, addiction, and psychiatric care have the potential to improve patient outcomes.

CONCLUSIONS

To manage addictions appropriately, one must recognize that alcohol and other drug dependence are chronic illnesses. As such, the focus should be on long-term management with the goal of relapse prevention. Continued treatment is associated with better outcome than short-term treatment. Behavioral treatments with proven efficacy include motivational enhancement and cognitive behavioral therapy, contingency management, less intense and less specialized medication management, and brief counseling interventions that have efficacy specifically for nondependent unhealthy alcohol use, and perhaps for initiating referral for care. Pharmacological treatments, usually in the context of counseling, also have proven efficacy. The best proven of these include acamprosate, naltrexone (oral and long-acting injectable), and disulfiram for alcohol dependence, and methadone and buprenorphine for opioid dependence. To avoid addressing only the most severely affected patients, attention needs to be directed to the larger proportion of patients with risky use to whom the largest number of health problems can be attributed. Care for addictions should include attention to psychiatric comorbidity. Clinicians who are not addiction specialists can play very important roles in addressing addiction by identifying patients with risky, problem, or dependent use; by providing brief counseling and prescribing efficacious medications; by addressing common comorbidities; and by referring to specialists when needed.

BIBLIOGRAPHY

1. Friedmann PD, Saitz R, Samet JH. Management of adults recovering from alcohol or other drug problems. JAMA 1998;279:1227-31.
2. Saitz R. Unhealthy alcohol use. N Engl J Med 2005;352; 596-607.
3. Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, et al, for the COMBINE Study Research Group. Effect of combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE Study: a randomized controlled trial. JAMA 2006;295: 2003-2017.
4. Garbutt JC, Kranzler HR, O'Malley SS, Gastfriend DR, Pettinati HM, Silverman BL, et al, for the Vivitrex Study Group. Efficacy and tolerability of long-acting injectable naltrexone for alcohol dependence: a randomized controlled trial. JAMA 2005;293:1617-1625.
5. McLellan AT. Have we evaluated addiction treatment correctly? Implications from a continuing care perspective. Addiction 2002;88:106-109.
6. McLellan AT, O'Brien CP, Lewis DL, Kleber HD. Drug addiction as a chronic medical illness: implications for treatment, insurance and evaluation. JAMA 2000;284:1689-1695.
7. Kaner EF, Beyer F, Dickinson HO, Pienaar E, Campbell F, Schlesinger C, Heather N, et al. Effectiveness of brief alcohol interventions in primary care populations [review]. Cochrane Database Syst Rev 2007;(2):CD004148.
8. US Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. Ann Intern Med 2004;140:554-556.
9. Miller WR, Walters ST, Bennett ME. How effective is alcoholism treatment in the United States? J Stud Alcohol 2001;62:211-220.
10. National Institute on Drug Abuse. National Institutes of Health. Principles of Addiction Treatment: A Research-Based Guide. NIH Publication 99-4180. Bethesda, MD: National Institutes of Health; 1999.
11. O'Connor PG, Fiellin DA. Pharmacologic treatment of heroin-dependent patients. Ann Intern Med 2000;133:40-54.
12. Ball JC, Ross A. The Effectiveness of Methadone Maintenance Treatment. New York, NY: Springer-Verlag; 1991.
13. Dole VP, Nyswander M. A medical treatment for diacetylmorphine (heroin) addiction: a clinical trial with methadone hydrochloride. JAMA 1965;193:80-84.
14. Fiellin DA, O'Connor PG. Office-based treatment of opioid-dependent patients. N Engl J Med 2002;347:817-823.
15. Carmen B, Angeles M, Ana M, Maria AJ. Efficacy and safety of naltrexone and acamprosate in the treatment of alcohol dependence: a systematic review. Addiction 2004;99:811-828.
16. Bernstein J, Bernstein E, Tassiopoulos K, Heeren T, Levenson S, Hingson R. Brief motivational intervention at a clinic visit reduces cocaine and heroin use. Drug Alcohol Depend 2005;77:49-59.
17. Fiellin DA, O'Connor PG. New federal initiatives to enhance the medical treatment of opioid dependence. Ann Intern Med 2002;137:688-692.
18. Fiellin DA et al. Consensus statement on office-based treatment of opioid dependence using buprenorphine. J Substance Abuse Treat 2004;27:153-159.-

Journal of Urban Health: Bulletin of the New York Academy of Medicine, Vol. 84, No. 2
doi:10.1007/s11524-006-9149-z

Risk of Mortality during Four Years after Substance Detoxification in Urban Adults

Richard Saitz, Jessie Gaeta, Debbie M. Cheng, Jessica M. Richardson, Mary Jo Larson, and Jeffrey H. Samet

ABSTRACT *The objective of this analysis was to assess the mortality rate and risk factors in adults, with substance dependence, who are not receiving primary medical care (PC). Date and cause of death were identified using the National Death Index data and death certificates for 470 adults without PC over a period of almost 4 years after detailed clinical assessment after detoxification. Factors associated with risk of mortality were determined using stepwise Cox proportional hazards models. Subjects were 76% male, 47% homeless, and 47% with chronic medical illness; 40% reported alcohol, 27% heroin, and 33% cocaine as substance of choice. Median age was 35. During a period of up to 4 years, 27 (6%) subjects died. Median age at death was 39. Causes included: poisoning by any substance (40.9% of deaths), trauma (13%), cardiovascular disease (13.6%), and exposure to cold (9.1%). The age adjusted mortality rate was 4.4 times that of the general population in the same city. Among these individuals without PC in a detoxification unit, risk factors associated with death were the following: drug of choice [heroin: hazard ratio (HR) 6.9 (95% confidence interval (CI) 1.6–31.1]; alcohol: HR 3.7 (95% CI 0.79–16.9) compared to cocaine); past suicide attempt (HR 2.1, 95% CI 0.96–4.5); persistent homelessness (HR 2.4, 95% CI 1.1–5.3); and history of any chronic medical illness (HR 2.1, 95% CI 0.93–4.7). Receipt of primary care was not significantly associated with death (HR 0.85, 95% CI 0.34–2.1). Risk of mortality is high in patients with addictions and risk factors identifiable when these patients seek help from the health care system (i.e., for detoxification) may help identify those at highest risk for whom interventions could be targeted.*

KEYWORDS *Alcoholism and addictive behavior, Drug abuse, Substance abuse, Primary care, Mortality*

INTRODUCTION

Addictive disorders are common in the United States at 3.8% for alcohol and 0.6% for drug dependence.[1] Many more people use alcohol and other drugs at levels that place them at risk for consequences.[2] A sizeable number of deaths in the United

Saitz, Cheng, Richardson, and Samet are with the Clinical Addiction Research and Education Unit, Department of Medicine, Boston Medical Center, Boston, MA, USA; Saitz, Gaeta, Cheng, Richardson, and Samet are with the Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston, MA, USA; Saitz and Richardson are with the Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA; Gaeta is with the Boston Health Care for the Homeless Program, Boston, MA, USA; Cheng is with the Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; Larson is with the New England Research Institutes, Watertown, MA, USA; Samet is with the Department of Social & Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA.

Correspondence: Richard Saitz, MD, MPH, FASAM, FACP, 91 East Concord Street, Suite 200, Boston, MA 02118, USA. (E-mail: rsaitz@bu.edu)

States are caused by alcohol (85,000) and illicit drugs (17,000) each year (4.2% of all deaths).[3] Alcohol and drug dependence increase the risk of death substantially.[4–8]

As in the general population, death in adults with substance dependence is more common in men than in women.[9–11] Similarly, older age is also associated with mortality in this population.[10,12,13] But in younger persons, there may also be high risk of death, particularly among blacks and urban residents.[9,14] Finally, racial and ethnic minorities with substance dependence are at higher risk of death than majority populations with these illnesses.[9,15,16]

Much of the mortality research including people with substance dependence has focused on injection drug users, primarily of heroin.[15,17,18] This is likely because of the high prevalence of overdose, among other serious medical consequences, and the concomitant risk of death.[4] More than two-thirds of illicit drug users in an Australian study reported nonfatal drug overdose in their lifetime,[17] and a substantial proportion of the excess mortality in drug users is because of drug overdose.[4]

In addition to direct pharmacologic consequences, social factors may contribute to excess mortality. For example, homelessness is associated with death among substance users.[19,20] More than half of all deaths in homeless persons in Atlanta were because of substance use.[21] The well-known common psychiatric comorbidities of homelessness are not the only explanation for increased mortality in this population. In fact, Hibbs et al. found that homeless persons with substance dependence and no other mental health problems were four times more likely to die than those with mental health problems and no substance dependence. Homeless persons with mental health problems and no substance dependence were three times more likely to die than the general population.[20]

As might be expected, numerous studies have identified medical illness as a predictor of death in people with substance dependence.[13,22,23] HIV, chronic liver disease, cardiovascular, infectious, digestive, respiratory, endocrine, metabolic, and hematological disorders contribute to these deaths.[19,23]

A better understanding of predictors of death among substance-dependent persons may provide information that could help decrease the risk of mortality. Furthermore, studying those without primary medical care (common among people with substance dependence) may select for people, at particularly high risk for death, who are amenable to preventive interventions. Among people with substance dependence, barriers to receipt of care may eclipse need, leading to inadequate care and preventable illnesses. It may be particularly useful to understand these predictors among such persons without primary medical care who are willing to seek help from the health care system (i.e., via detoxification), a point at which intervention has the potential to prevent further morbidity and mortality. The main objectives of this study were to (1) describe the mortality rate and causes of death in alcohol and drug-dependent adults (without primary medical care) after detoxification and (2) identify clinical predictors of mortality that can be recognized at the time of detoxification.

METHODS

Design

The Health Evaluation and Linkage to Primary care (HELP) study was a randomized controlled trial (RCT) testing the effectiveness of a multidisciplinary health assessment and referral to link alcohol and other drug-dependent individuals

to primary care. A detailed description of the HELP study RCT was previously reported.[24] After providing written informed consent, eligible subjects were enrolled in the RCT. In the current study, we considered risk of mortality in the prospective follow-up of this cohort. As part of the RCT, subjects provided information regarding substance use, demographic, social, and health status and were sought for follow-up interviews every 6 months for 2 years. The Institutional Review Board at Boston University Medical Center approved the study. Additional privacy protection was secured by the issuance of a Certificate of Confidentiality by the U.S. Department of Health and Human Services.

Subjects

Enrollment occurred between February 1, 1997 and April 1, 1999. Eligible subjects were adult inpatients at a single urban residential detoxification unit, who spoke Spanish or English, reported alcohol, heroin, or cocaine as their first or second drug of choice, and resided in proximity to the primary care clinic to which they would be referred or were homeless. Patients with primary care physicians whom they had seen within the prior 2 years; significant dementia; plans to leave the Boston area that would prevent research participation; failure to provide contact information for tracking purposes; or pregnancy were excluded. Of 642 eligible subjects, 470 (73%) consented to participate and were enrolled.

Data Collection Procedures

Trained research associates administered standard interviews to subjects at study entry (baseline) and follow-up interviews. The Spanish interview instrument used the standardized Spanish versions of scales when available. The remainder of the Spanish questionnaire was translated from the English version, back translated, checked for accuracy, and then corrected.

Independent Variables

Independent variables of interest were all assessed by interview at study entry. They included: age, gender, health literacy [Rapid Estimate of Adult Literacy in Medicine (REALM) score[25]], first language [English (yes/no)], level of education [high school graduate (yes/no)], suicide attempt (ever), drug or alcohol overdose requiring emergency department attention (ever), drug of choice (alcohol, heroin, or cocaine), addiction severity [Addiction Severity Index (ASI) alcohol and drug scales[26]], alcohol or drug problems [Inventory of Drug Use Consequences (InDUC-2R)[27]], physical and mental health-related quality of life [the physical and mental component summary scores (PCS and MCS, respectively) derived from the Short Form Health Survey (SF-36)[28,29]], depressive symptoms [Center for Epidemiologic Studies Depression (CES-D) scale score[30]], physician diagnosed chronic medical conditions,[31] tobacco use in the past 30 days before study entry, and the following in the past 6 months before study entry: health insurance (receipt of Medicaid, Medicare, health insurance from a job or family member's job, or any other health benefits plan that paid for medical care expenses), injection drug use, past physical or sexual abuse,[32] buying or selling of sex, any homelessness (one or more nights in a shelter or on the street), and persistent homelessness (>21 days in a shelter or on the street). Receipt of primary care (yes/no) since the last research interview was assessed by a series of detailed questions and was entered as a time-varying covariate.[33]

TABLE 1 Baseline characteristics of residential detoxification patients ($N = 470$) and associated risk for death in bivariable analyses

Characteristic	Frequency	Unadjusted hazard ratio (95% CI)
Median age (Range)	35 (18–60)	1.03 (0.98–1.1)
Male (%)	76	2.5 (0.76–8.4)
Race (%)		
Black	46	1.0
White	37	2.6 (0.98–6.96)
Hispanic	11	4.8 (1.6–14.96)**
Other	6	4.1 (1.0–16.3)*
Uninsured (%)	60	1.3 (0.58–2.9)
Any homelessness (past 6 months) (%)	47	1.4 (0.68–3.1)
Persistent homelessness (≥21 nights past 6 months) (%)	24	2.7 (1.3–5.7)*
Suicide attempt (lifetime) (%)	22	2.6 (1.2–5.5)*
Drug overdose (lifetime) (%)	31	2.5 (1.2–5.4)*
Tobacco (past 30 days) (%)	86	0.63 (0.24–1.7)
Chronic medical illness[a] (%)	47	2.3 (1.0–5.2)*
Drug of choice (%)		
Cocaine	33	1.0
Alcohol	40	5.7 (1.3–25.4)*
Heroin	27	7.7 (1.7–34.4)**
First language (not English) (%)	11	2.0 (0.75–5.3)
Education (high school graduate) (%)	69	1.5 (0.62–3.8)
Injection drug use (past 6 months) (%)	22	2.1 (0.95–4.5)
Physical or sexual abuse (past 6 months) (%)	37	0.91 (0.41–2.1)
Buying or selling sex (past 6 months) (%)	28	0.32 (0.10–1.1)
Health literacy (REALM[b]) median (range)	61 (0–66)	1.0 (0.97–1.0)
Alcohol severity (ASI[c]) median (range)	0.51 (0.00–1.03)	2.3 (0.72–7.2)
Drug severity (ASI[c]) median (range)	0.29 (0.00–0.58)	0.27 (0.02–3.1)
Physical health (PCS[d]) median (range)	48.89 (14.07–74.81)	0.97 (0.93–0.998)*
Mental health (MCS[e]) median (range)	28.56 (6.76–62.18)	0.98 (0.95–1.0)
Depressive symptoms (CES-D[f]) median (range)	34 (1–60)	1.0 (0.99–1.1)

[a]Of the subjects, 72% reported undergoing HIV testing; of those, 2.6% reported receipt of positive results.
[b]Rapid Estimate of Adult Literacy in Medicine
[c]Addiction Severity Index
[d]SF-36 Physical Component Summary
[e]SF-36 Mental Component Summary
[f]Center for Epidemiologic Studies-Depression Scale
*$P < 0.05$
**$P < 0.01$

Outcome

The primary outcome was time between study entry and death. Deaths and dates of death were identified using either the National Death Index (NDI) (dates of death for NDI come from death certificates),[34,35] which was provided under an agreement between Boston Medical Center and the Department of Health and Human Services, or reports from family or friends of study subjects. At the end of the study, we searched for all subjects in the NDI during the calendar years 1997–2001. The NDI matches subjects to death records based on 12 identifying criteria (day of birth,

month of birth, year of birth, social security number, first, middle, and last name, sex, race, marital status, state of birth, and state of residence). An algorithm considering the correspondence and weight of each identifying criterion is then used to classify the quality of the matches. We accepted only those matches that fit NDI-recommended quality guidelines. NDI characterized deaths by cause, which was identified by International Classification of Diseases (ICD) ninth edition codes from death certificates.

The survival time was defined as the duration of time between date of study entry and date of death (for those who died). The observation time for those who did not die was defined as the time between date of study entry and December 31, 2001, the end of the NDI search period. The Kaplan–Meier estimator was used to determine the survival probability across the mortality data collection period.

City of Boston mortality rate data for years 1997–2001 were provided by the Boston Public Health Commission. We used this data as a descriptive comparison for the rate observed in study subjects.

Analysis

All analyses used SAS/STAT software, Version 8.2.[36] All study subjects were eligible for analysis. Descriptive statistics were used to characterize the study subjects. Age-adjusted mortality rates were estimated for both the study sample and the general population of Boston. The age adjustment was performed using the method of direct standardization, and the United States year 2000 population was used as the standard for the age adjustment. For the purpose of comparison, the mortality rates for the current study sample and the city of Boston (calculated for the years 1997–2001) were annualized to represent a 1-year mortality rate.

Given the small number of events, it was not feasible to fit multivariable Cox regression models containing many predictors. Thus, we used an iterative process to select variables for entry into a stepwise model. Preliminary unadjusted Cox proportional hazards models were used to assess the bivariable associations between each of the 22 independent variables and time until death. Factors that were significant at an alpha level of 0.05 were selected as candidate variables for a stepwise procedure, with the exception of demographic characteristics (age, gender, and race), which were considered potential confounders, and thus, were not included in the stepwise model. The stepwise entry and removal criteria were set at a P value of 0.15. Correlations between independent variables were obtained, and no pair of variables had a correlation greater than 0.40. To assess potential confounding because of static demographic characteristics, we fit an additional model that included age, gender, and race with the statistically significant variables identified by the stepwise procedure. Reported P values are two-tailed, and a P value of less than 0.05 was considered statistically significant.

RESULTS

Subject Characteristics ($N = 470$)

Subject characteristics appear in Table 1. Of the 470 subjects, 2 died before their 6-month interview, and 400 of 468 (85%) completed at least one follow-up research interview; of these, 253 (63%) reported receipt of at least 1 primary care visit.

TABLE 2 Risk factors for death in adults with addictions in multivariable analysis*

Outcome	Hazard ratio (95% CI)	*P* value
Drug of choice		
Heroin vs. cocaine	6.9 (1.6–31.1)	0.01
Alcohol vs. cocaine	3.7 (0.78–16.9)	0.09
Persistent homelessness ($\geq$ 21 nights in past 6 months)	2.4 (1.1–5.3)	0.03
Suicide attempt (ever)	2.1 (0.96–4.5)	0.06
Chronic medical illness	2.1 (0.93–4.7)	0.07

*Analysis based on a stepwise Cox regression model (entry and removal criteria was set at a *P* value of 0.15) that considered the variables listed in the table and drug overdose ever

Mortality and Causes of Death

The 470 subjects were observed for a median (range) of 46 (1–56) months and 1,728 person-years; of the 470 subjects, 27 (6%) died between 1997 and 2001. The probability of death by 12, 24, and 36 months was 1.7, 3.4, and 5.1%, respectively. The median age at death was 39. Twenty-four (89%) deaths were identified using NDI data, and the remainder was identified in reports from families or friends. Cause of death data was available for 22 (82%) of the deaths. Of these 22 deaths, the most common cause of death was poisoning (41%). Of the nine deaths caused by poisoning, seven were attributed to narcotics, one to analgesics, and one was unspecified. Trauma (one each of homicide, motor vehicle crash, and pedestrian motor vehicle crash) and cardiovascular disease were each causes for three (14%) of the deaths. Exposure to cold and alcohol abuse were each the cause of two (9%) deaths. Diabetes, malignant neoplasm of lung, and intracerebral hemorrhage each caused one (5%) death.

The age adjusted mortality rate for the study cohort was estimated as 1,608 per 100,000 people, 4.4 times that of the general population in the City of Boston (368 per 100,000 people).

Predictors of Risk of Death

In unadjusted bivariable analyses of static demographic characteristics, race was associated with higher risk of mortality. The risk of death was higher for those who described themselves as Hispanic [hazard ratio (HR) 4.8, 95% CI 1.6–14.96] and other race (HR 4.1, 95% CI 1.0–16.3) compared to blacks. Other race included those who were not white, black, or Hispanic. Age and gender were not associated with risk of mortality. In unadjusted bivariable analyses of nondemographic characteristics, alcohol and heroin as drugs of choice (compared with cocaine), persistent homelessness, suicide attempt (ever), drug or alcohol overdose requiring emergency medical attention (ever), and chronic medical illness were each associated with a higher risk of death (Table 1). Better physical health-related quality of life was associated with a lower risk of death. Receipt of primary care was not significantly associated with death (HR 0.85, 95% CI 0.34–2.1). The stepwise procedure resulted in a model that included the following nondemographic risk factors: drug of choice (heroin, alcohol, or cocaine), persistent homelessness, suicide attempt, and chronic medical illness (Table 2). In this multivariable model, heroin (HR 6.9 compared to cocaine, 95% CI 1.6–31.1) and persistent homelessness (HR 2.4, 95% CI 1.1–5.3) were significantly associated with an increased risk of death. To assess confounding by demographic characteristics, a subsequent

regression model was fit that included drug of choice, persistent homelessness (the two significant factors from the stepwise model), and the potential confounders, age, gender, and race. In this adjusted model, heroin as drug of choice remained significantly associated with a higher risk of death (HR 5.4 compared to cocaine, 95% CI 1.1–26.5, $P = 0.04$), alcohol as drug of choice was not significantly associated with risk of death (HR 3.0 compared to cocaine, 95% CI 0.6–14.9, $P = 0.18$), and the effect of persistent homelessness on risk of death became attenuated and of borderline statistical significance. (HR 2.1, 95% CI 0.97–4.8, $P = 0.06$). Age, gender, and race were not associated with risk of mortality ($P > 0.15$) in this adjusted model. The probability of death by the end of the follow-up period was 7.1, 1.3, and 17.5% for the substances of choice, alcohol, cocaine, and heroin, respectively; 5.9% for no persistent homelessness and 14.4% for persistent homelessness; 5.7% for no suicide attempt and 16.8% for suicide attempt; and 7.4% for no chronic medical illness and 8.1% for chronic medical illness.

DISCUSSION

Despite the young age of this cohort of men and women with substance dependence, most of whom were not using injection drugs, and who underwent residential detoxification, the short-term mortality rate was high and notably higher than that in the surrounding community. Causes of death could all be related to drug or alcohol dependence. Risk factors identifiable by detailed evaluation at the time of detoxification characterized individuals at even higher risk for death, including heroin as the drug of choice and persistent homelessness. These findings are particularly notable because this cohort had not received primary medical care, and they were at a potential point of entry into the general health system, where they might have risks assessed and addressed. Furthermore, alcohol was the drug of choice for most subjects, thus, adding to the mortality research literature primarily focused on injection drug users. Finally, although there was limited ability to study the impact of health services on mortality, our findings regarding receipt of primary care suggest that such preventive services should be further studied to assess for possible protective effects.

These results expand but are consistent with literature that has reported an increased risk of mortality for people with substance dependence.[3–8] While mortality rates cannot be directly compared across studies without age or sex adjustment, our results appear to be consistent with other literature reports. A prospective study of mortality among drug users after treatment showed an equally high mortality rate (1.2%), which was six times higher than that for a general, age-matched population.[37] A prospective study of mortality after alcohol and drug abuse treatment showed a higher mortality of 2.38% per year, three times that expected in male veterans.[38] Moos et al. reported that mortality in treated alcohol-dependent patients was two to five times higher than age- and sex-matched comparison groups.[23] The annual mortality rate for people with opioid abuse in a 12-year follow-up study was 1.4%,[4] and among subjects with a history of injection drug use in Scotland, the average annual mortality rate was 2.3% over a 21-year period.[7] These rates are all similar to the 1.7% per year rate in our study and our finding of increased mortality compared with the general population.

Prior studies of substance users have reported causes of death similar to those found in this study. Poisoning, the most common cause of death reported here,

increased in the US by 56% between 1991 and 2001.[39] While poisoning accounted for 40.9% of deaths in our study, it was the cause of 22% of the deaths in a 33-year study of heroin users.[40] A longitudinal intervention study of mortality after drug treatment found the same three leading causes of death: poisoning, trauma, and medical illness.[37] Heroin as primary drug of choice is a major focus in the addiction mortality literature because of the high risk of overdose and infectious disease associated with its use.[4,15,17,18,40,41] Homelessness is another predictor of risk of death for substance-dependent people.[19–21] Of note, any homelessness was not a significant risk factor in our analyses. Only the more severe "persistent" homelessness was a significant predictor. Our findings also confirm that suicide attempts,[42–44] chronic medical illness[13,22,23] and alcohol as drug of choice[22,45,46] are predictors of mortality in substance-dependent individuals. In this study, neither mental illness nor health literacy was associated with mortality. This observation is not what would be expected based on prior research,[47,48] and may be because of competing mortality risks in the sample. In addition, we did not assess knowledge of overdose risk or prevention knowledge, which could have been associated with death.

This study relied on rich detailed clinical data obtained by self-report using standardized instruments. Outcomes were ascertained using the well-described National Death Index[49–51] and supplemented by reports of family and friends provided during active follow-up of the cohort. Nonetheless, several important limitations should be considered. First, it is possible that the death index and the supplemental family reports missed deaths, and the number of deaths in the study was relatively small. As a result, analyses could not consider large numbers of variables simultaneously, and we could not explore interactions or understand possible effects of smaller groups (e.g., individual racial and ethnic groups). Similarly, we may not have been able to identify all predictors of mortality as a result of limited power (e.g., injection drug use, drug overdose, alcohol addiction severity). The absence of a detectable effect of primary care is notable, although one should avoid drawing a conclusion of no mortality benefit. Aside from power limitations and an inability to be confident that we captured all primary care exposures because of limitations in follow-up, we were unable to examine primary care in adjusted analyses over a long-enough period of time to see the effects of high quality preventive and chronic care interventions. However, we were able to identify significant predictors of risk of mortality while adjusting for basic demographic characteristics. Finally, the observational nature of the study limited our ability to draw firm conclusions about causality. This limitation is somewhat tempered by the prospective nature of the study.

Substance dependence is common.[1] Although tobacco use is clearly a substantial contributor to long-term mortality,[52] in this population of post-detoxification individuals with substance dependence (most of whom smoked cigarettes), relatively short-term mortality is so substantial that it merits consideration for intervention. At least some causes of death in these individuals may be preventable. We have identified characteristics that can be used to identify those at increased risk of mortality after detoxification. Although our findings are consistent with the literature in this area, we add to that literature in a substantial way by reporting on this prospective cohort and by considering a range of clinical characteristics assessed first hand using standardized tools, by including a range of drugs of choice, and by focusing on a common characteristic in substance-dependent adults—lack of regular primary medical care. Given that people who

enter detoxification often have no medical or addiction follow-up care,[53–57] this is an opportune moment during which people at particularly high risk can be identified and potential interventions can be developed. As the risk for death among detoxification patients overall is elevated, even among those without specific risk factors, attention should be given as to how best to address these individuals' health needs. These findings support arguments for better integrating detoxification and other addictions, psychiatric, and general health care.

ACKNOWLEDGEMENTS

We gratefully acknowledge the assistance of Michael Winter, MPH, in the preparation of this manuscript. Primary grant support for this study came from the National Institute on Drug Abuse (R01-10019) and the National Institute on Alcohol Abuse and Alcoholism (R01-10870). This research was conducted in part in the General Clinical Research Center at Boston University School of Medicine (M01-RR00533). Preliminary results were presented at the annual meeting of the Society General Internal Medicine, May 2004 in Chicago, IL. We gratefully acknowledge the staff of the HELP study for their help in the conduct of this research.

REFERENCES

1. Grant BF, Stinson FS, Dawson DA, et al. 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(8):807–816.
2. Saitz R. Clinical practice. Unhealthy alcohol use. *N Engl J Med*. 2005;352(6):596–607.
3. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA*. 2004;291(10):1238–1245.
4. Joe GW, Simpson DD. Mortality rates among opioid addicts in a longitudinal study. *Am J Public Health*. 1987;77(3):347–348.
5. Barr HL, Antes D, Ottenberg DJ, Rosen A. Mortality of treated alcoholics and drug addicts: the benefits of abstinence. *J Stud Alcohol*. 1984; 45(5):440–452.
6. Marshall EJ, Edwards G, Taylor C. Mortality in men with drinking problems: a 20-year follow-up. *Addiction*. 1994;89(10):1293–1298.
7. Copeland L, Budd J, Robertson JR, Elton RA. Changing patterns in causes of death in a cohort of injecting drug users, 1980–2001. *Arch Intern Med*. 2004;164(11):1214–1220.
8. Hulse GK, English DR, Milne E, Holman CD. The quantification of mortality resulting from the regular use of illicit opiates. *Addiction*. 1999;94(2):221–229.
9. Shai D. Mortality associated with drug misuse among blacks in New York City, 1979–1981. *Int J Addict*. 1992;27(12):1433–1443.
10. Bird SM, Hutchinson SJ, Goldberg DJ. Drug-related deaths by region, sex, and age group per 100 injecting drug users in Scotland, 2000-01. *Lancet*. 2003;362(9388):941–944.
11. Feuerlein W, Kufner H, Flohrschutz T. Mortality in alcoholic patients given inpatient treatment. *Addiction*. 1994;89(7):841–849.
12. Mackenzie A, Allen RP, Funderburk FR. Mortality and illness in male alcoholics: an 8-year follow-up. *Int J Addict*. 1986;21(8):865–882.
13. Finney JW, Moos RH. The long-term course of treated alcoholism: II. Predictors and correlates of 10-year functioning and mortality. *J Stud Alcohol*. 1992;53(2):142–153.
14. Oyefeso A, Ghodse H, Clancy C, Corkery J, Goldfinch R. Drug abuse-related mortality: a study of teenage addicts over a 20-year period. *Soc Psychiatry Psychiatr Epidemiol*. 1999;34(8):437–441.

15. Galea S, Ahern J, Tardiff K, et al. Racial/ethnic disparities in overdose mortality trends in New York City, 1990–1998. *J Urban Health*. 2003; 80(2):201–211.
16. Substance Abuse and Mental Health Services Administration (SAMHSA). Office of Applied Studies. U.S. Department of Health and Human Services. 2003 National Survey on Drug Use and Health: Results. Available at http://www.oas.samhsa.gov. Accessed December 1, 2006
17. Darke S, Zador D. Fatal heroin 'overdose': a review. *Addiction*. 1996;91(12):1765–1772.
18. Oppenheimer E, Tobutt C, Taylor C, Andrew T. Death and survival in a cohort of heroin addicts from London clinics: a 22-year follow-up study. *Addiction*. 1994;89(10):1299–1308.
19. Hwang SW, Lebow JM, Bierer MF, O'Connell JJ, Orav EJ, Brennan TA. Risk factors for death in homeless adults in Boston. *Arch Intern Med*. 1998;158(13):1454–1460.
20. Hibbs JR, Benner L, Klugman L, et al. Mortality in a cohort of homeless adults in Philadelphia. *N Engl J Med*. 1994;331(5):304–309.
21. Hanzlick R, Lazarchick J. Health care history and utilization for Atlantans who died homeless. *J Med Assoc Georgia*. 1989;78(4):205–208.
22. Lewis CE, Smith E, Kercher C, Spitznagel E. Predictors of mortality in alcoholic men: a 20-year follow-up study. *Alcohol Clin Exp Res*. 1995;19(4):984–991.
23. Moos RH, Brennan PL, Mertens JR. Mortality rates and predictors of mortality among late-middle-aged and older substance abuse patients. *Alcohol Clin Exp Res*. 1994; 18(1):187–195.
24. Samet JH, Larson MJ, Horton NJ, Doyle K, Winter M, Saitz R. Linking alcohol- and drug-dependent adults to primary medical care: a randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. *Addiction*. 2003;98(4): 509–516.
25. Davis TC, Jackson RH, George RB, et al. Reading ability in patients in substance misuse treatment centers. *Int J Addict*. 1993;28(6):571–582.
26. McLellan AT, Kushner H, Metzger D, et al. The fifth edition of the Addiction Severity Index. *J Subst Abuse Treat*. 1992;9:199–213.
27. Tonigan JS, Miller WR. The Inventory of Drug Use Consequences (InDUC): test–retest stability and sensitivity to detect change. *Psychol Addict Behav*. 2002;16:165–168.
28. Ware JE. *SF-36 Health Survey. Manual and Interpretation Guide*. Boston, MA: The Health Institute; 1993.
29. Ware JE. *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. Boston, Massachusetts: The Health Institute; 1994.
30. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1:385–401.
31. De Alba I, Samet JH, Saitz R. Burden of medical illness in drug and alcohol dependent persons without primary care. *Am J Addict*. 2004;13:33–45.
32. Liebschutz J, Savetsky JB, Saitz R, Horton NJ, Lloyd-Travaglini C, Samet JH. The relationship between sexual and physical abuse and substance abuse consequences. *J Subst Abuse Treat*. 2002;22(3):121–128.
33. 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(1):70–78.
34. Bilgrad R. *National Death Index User's Manual*. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention; 1995.
35. Bilgrad R. *National Death Index Plus: Coded Causes of Death Supplement to the National Death Index User's Manual*. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention; 1997.
36. SAS/STAT Software, Version 8.2 of the SAS System for Windows, Copyright © 1999-2001. Cary, NC: SAS Institute Inc.; 2005
37. Gossop M, Stewart D, Treacy S, Marsden J. A prospective study of mortality among drug misusers during a 4-year period after seeking treatment. *Addiction*. 2002;97(1):39–47.

38. Johnson JE, Finney JW, Moos RH. Predictors of 5-year mortality following inpatient/residential group treatment for substance use disorders. *Addict Behav*. 2005;30(7):1300–1316.
39. Department of Health and Human Services. Centers for Disease Control and Prevention. Unintentional and undetermined poisoning deaths—11 states, 1990–2001. *MMWR*. 2004;53:233–238.
40. Hser YI, Hoffman V, Grella CE, Anglin MD. A 33-year follow-up of narcotics addicts. *Arch Gen Psychiatry*. 2001;58(5):503–508.
41. Sanchez-Carbonell X, Vilaregut A. A 10-year follow-up study on the health status of heroin addicts based on official registers. *Addiction*. 2001;96(12):1777–1786.
42. Beautrais AL. Subsequent mortality in medically serious suicide attempts: a 5 year follow-up. *Aust N Z J Psychiatry*. 2003;37(5):595–599.
43. Suokas J, Suominen K, Isometsa E, Ostamo A, Lonnqvist J. Long-term risk factors for suicide mortality after attempted suicide—findings of a 14-year follow-up study. *Acta Psychiatr Scand*. 2001;104(2):117–121.
44. Ostamo A, Lonnqvist J. Excess mortality of suicide attempters. *Soc Psychiatry Psychiatr Epidemiol*. 2001;36(1):29–35.
45. Vaillant GE. A long-term follow-up of male alcohol abuse. *Arch Gen Psychiatry*. 1996;53(3):243–249.
46. Kristenson H, Osterling A, Nilsson JA, Lindgarde F. Prevention of alcohol-related deaths in middle-aged heavy drinkers. *Alcohol Clin Exp Res*. 2002;26(4):478–484.
47. Dembling BP, Chen DT, Vachon L. Life expectancy and causes of death in a population treated for serious mental illness. *Psychiatr Serv*. 1999;50(8):1036–1042.
48. 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(12):1228–1239.
49. Doody MM, Hayes HM, Bilgrad R. Comparability of national death index plus and standard procedures for determining causes of death in epidemiologic studies. *Ann Epidemiol*. 2001;11(1):46–50.
50. Lash TL, Silliman RA. A comparison of the National Death Index and Social Security Administration databases to ascertain vital status. *Epidemiology*. 2001;12(2):259–261.
51. Sesso HD, Paffenbarger RS, Lee IM. Comparison of National Death Index and World Wide Web death searches. *Am J Epidemiol*. 2000;152(2):107–111.
52. Hurt RD, Offord KP, Croghan IT, et al. Mortality following inpatient addictions treatment. Role of tobacco use in a community-based cohort. *JAMA*. 1996;275(14):1097–1103.
53. Samet JH, Saitz R, Larson MJ. A case for enhanced linkage of substance abusers to primary medical care. *Subst Abuse*. 1996;17:181–192.
54. 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.
55. Weisner C, McCarty D, Schmidt L. New directions in alcohol and drug treatment under managed care. *Am J Manag Care*. 1999;5:SP57–SP69.
56. D'Aunno T. *Linking Substance-Abuse Treatment and Primary Health Care. Treating Drug Abusers Effectively*. Cambridge, MA: Blackwell; 1997:311–331.
57. Laine C, Newschaffer C, Zhang D, Rothman J, Hauck WW, Turner BJ. Models of care in New York State Medicaid substance abuse clinics. Range of services and linkages to medical care. *J Subst Abuse*. 2000;12(3):271–285.

Annals of Internal Medicine | ARTICLE

Brief Intervention for Medical Inpatients with Unhealthy Alcohol Use

A Randomized, Controlled Trial

Richard Saitz, MD, MPH; Tibor P. Palfai, PhD; Debbie M. Cheng, ScD; Nicholas J. Horton, ScD; Naomi Freedner, MPH; Kim Dukes, PhD; Kevin L. Kraemer, MD, MSc; Mark S. Roberts, MD, MPP; Rosanne T. Guerriero, MPH; and Jeffrey H. Samet, MD, MA, MPH

Background: The efficacy of brief intervention in reducing alcohol consumption is well established for selected outpatients but not for medical inpatients.

Objective: To determine whether brief intervention improves alcohol outcomes in medical inpatients who were identified by screening as having unhealthy alcohol use.

Design: Randomized, controlled trial.

Setting: Medical service of an urban hospital.

Patients: 341 medical inpatients who were drinking risky amounts of alcohol (defined for eligibility as >14 drinks/wk or ≥5 drinks/occasion for men and >11 drinks/wk or ≥4 drinks/occasion for women and persons ≥66 y); 77% had alcohol dependence as determined by the Composite International Diagnostic Interview Alcohol Module.

Intervention: A 30-minute session of motivational counseling given by trained counselors during a patient's hospitalization (n = 172) versus usual care (n = 169).

Measurements: Self-reported primary outcomes were receipt of alcohol assistance (for example, alcohol disorders specialty treatment) by 3 months in dependent drinkers and change in the mean number of drinks per day from enrollment to 12 months in all patients.

Results: The intervention was not significantly associated with receipt of alcohol assistance by 3 months among alcohol-dependent patients (adjusted proportions receiving assistance, 49% for the intervention group and 44% for the control group; intervention–control difference, 5% [95% CI, −8% to 19%]) or with drinks per day at 12 months among all patients (adjusted mean decreases, 1.5 for patients who received the intervention and 3.1 for patients who received usual care; adjusted mean group difference, −1.5 [CI, −3.7 to 0.6]). There was no significant interaction between the intervention and alcohol dependence in statistical models predicting drinks per day (P = 0.24).

Limitations: Baseline imbalances existed between randomized groups. Patients who received usual care were assessed and advised that they could discuss their drinking with their physicians.

Conclusions: Brief intervention is insufficient for linking medical inpatients with treatment for alcohol dependence and for changing alcohol consumption. Medical inpatients with unhealthy alcohol use require more extensive, tailored alcohol interventions.

Ann Intern Med. 2007;146:167-176. www.annals.org
For author affiliations, see end of text.
ClinicalTrials.gov Identifier: NCT00183105.

Professional organizations recommend that clinicians screen their patients for unhealthy alcohol use (that is, the spectrum from drinking risky amounts to dependence) and conduct a brief intervention when indicated (1, 2). Despite this recommendation and the existence of brief, valid screening tools (3–5), patients with unhealthy alcohol use often are not identified and do not receive timely care.

Although widely recommended, brief intervention has proven efficacy in decreasing alcohol consumption and related consequences only in unhealthy drinkers without alcohol dependence and in outpatient settings (6). Its efficacy among other populations (for example, persons with alcohol dependence) and in inpatient settings remains unclear (7).

Evidence suggests, however, that medical inpatients—a group with a high prevalence of alcohol-related problems—may benefit from brief intervention. Some studies have demonstrated the efficacy of brief intervention in settings similar to medical services in which alcohol-related problems are common and their related consequences are severe (8, 9). Further, brief interventions are well suited to medical services. Patients who otherwise might not seek care are accessible and have time for an intervention. Persons admitted because of an alcohol-related problem may recognize the link between drinking and hospitalization, thus providing a "teachable moment" (10). Also, busy staff might implement a brief intervention because of its brevity and flexibility.

The unmet need for alcohol screening and intervention and opportunities for implementation underscore the importance of determining the efficacy of brief intervention in medical inpatients with unhealthy alcohol use. In addition, evaluating its effectiveness and practicality in real-world settings is critical to help clinicians make informed decisions when treating their patients (11). Therefore, we conducted a randomized, controlled trial to exam-

See also:

Print
Editors' Notes 168
Editorial comment. 223
Summary for Patients. I-22

Web-Only
Appendix
Conversion of figure and tables into slides

Context

Brief interventions reduce alcohol use in outpatients who drink unhealthy amounts but are not alcohol-dependent. Their effect in medical inpatients is unknown.

Contribution

The authors screened all adult medical inpatients at an urban teaching hospital and randomly assigned 341 risky drinkers to a 30-minute motivational counseling intervention followed gy treatment planning or to usual care. By 3 months, the same proportion of patients from both groups had received alcohol assistance, and both groups had reduced their drinking to the same degree.

Cautions

Three quarters of the participants met Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, criteria for alcohol dependence.

Implications

In this well-done study, brief intervention did not affect alcohol-related outcomes in persons who drank unhealthy amounts.

—The Editors

ine whether screening followed by brief intervention would improve alcohol-related outcomes in "typical" medical inpatients (that is, a racially diverse group with a range of unhealthy alcohol use, comorbid conditions, and readiness to change). We hypothesized that screening and brief intervention would lead to the following: receipt of alcohol assistance (for example, specialty treatment) among persons with alcohol dependence and, among all persons decreased alcohol consumption, alcohol-related problems, and health care utilization and improved readiness to change and health-related quality of life.

METHODS

Patients

As previously described, we recruited patients from the inpatient medical service of a large, urban teaching hospital (12). Trained research associates approached all patients who were age 18 years or older and whose physicians did not decline patient contact. Patients fluent in English or Spanish who gave verbal consent were asked to complete a screening interview to determine eligibility: currently (past month) drinking risky amounts (defined for eligibility as >14 standard drinks/wk or ≥5 drinks/occasion for men and >11 drinks/wk or ≥4 drinks/occasion for women and persons ≥66 years); 2 contacts to assist with follow-up; no plans to move from the area in the next year; and a Mini-Mental State Examination score of 21 or greater (13, 14).

Research associates assessed demographic characteristics and administered the Alcohol Use Disorders Identification Test (AUDIT) (15) by interview. To better characterize current alcohol use, they assessed the average numbers of drinking-days per week and drinks consumed on a typical day, and the maximum number of drinks consumed per occasion (16, 17). For the first 7 months of the study, research associates asked these additional questions only to patients with an AUDIT score of 8 or greater (a recommended cutoff for screening) (18). For the remaining 22 months, research associates asked the additional questions to anyone who drank in the past 12 months to maximize identification of drinkers of risky amounts. Lastly, the research associates asked all patients who were drinking risky amounts to describe their readiness to change by using a visual analog scale ranging from 0 to 10 (19).

Enrolled patients provided written informed consent and were compensated for each completed interview. The institutional review board at Boston University Medical Center approved this study. We secured additional privacy protection with a certificate of confidentiality from the National Institute on Alcohol Abuse and Alcoholism.

Assessment at Enrollment

Research associates interviewed patients before randomization to assess the characteristics shown in **Table 1**. One author reviewed the medical records to determine medical diagnoses (29). Diagnoses of alcohol use disorders were based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (30), and were determined with the Composite International Diagnostic Interview (CIDI) Alcohol Module (31, 32).

Randomization and Intervention

An off-site data management group generated assignments to control and intervention groups by using a permuted block (size 8) randomization procedure stratified by AUDIT score (<12 vs. ≥12) and provided us the assignments in sealed opaque envelopes. We used the AUDIT score to stratify because we could not score the CIDI before randomization. After each baseline assessment, research associates opened an envelope and informed the patient of his or her assignment.

Patients in the control group received usual care (that is, they were told the screening results and that they could discuss their drinking with their physicians). Specialists were available by referral. Systematic alcohol screening and brief intervention were not routine at this hospital.

We assigned patients in the intervention group to a 30-minute session of brief motivational counseling (19, 33) conducted by counseling and clinical psychology doctoral students whom we trained and supervised. Sessions were audiotaped and included feedback, an open discussion, and construction of a change plan (**Appendix**, available at www.annals.org).

Outcomes and Measurements

The first primary outcome was self-reported receipt of alcohol assistance in the past 3 months by patients with

CIDI-determined alcohol dependence. This outcome was measured at the 3-month follow-up visit with a standardized interview based on the Treatment Services Review (34) and Form 90 (35). Assistance included residential treatment, outpatient treatment (for example, specialty counseling or therapy), medications, employee assistance programs, or mutual-help groups (for example, Alcoholics Anonymous).

The other primary outcome was the change in the number of mean drinks per day in the past 30 days from enrollment to 12 months among all patients. We determined consumption with the Timeline Follow-back method (36).

Five secondary consumption outcomes (past 30 days) included changes from enrollment to 12 months in the numbers of heavy drinking episodes (≥5 drinks/occasion for men and ≥4 drinks/occasion for women and for persons ≥66 y) and days abstinent; and the proportions of patients drinking risky amounts (>14 drinks/wk or ≥5 drinks/occasion for men and >7 drinks/wk or ≥4 drinks/occasion for women and persons ≥66 y) (37), having 1 or more heavy drinking episodes, and abstaining for all 30 days.

Other secondary outcomes included the changes at 12 months in readiness to change (Taking Steps scale on the Stages of Change Readiness and Treatment Eagerness Scale) (38), alcohol problems (total score on the Short Inventory of Problems) (39), physical and mental health-related quality of life (Physical and Mental Component Summary scale scores on the Short-Form Health Survey) (40), and emergency department visits and days of medical hospitalization (both determined by a standardized interview based on the Treatment Services Review and Form 90) (34, 35).

Follow-up Procedures

Research associates conducted follow-up visits, which included reassessment of most domains covered at enrollment, usually in person and at 3 and 12 months (10% and 13%, respectively, by telephone; similar by randomized group). They performed alcohol breath tests at in-person follow-up visits (41).

Although they were involved in the randomization assignment, research associates were not involved in the intervention. Further, 64% of patients at 3-month follow-up and 85% of patients at 12-month follow-up were interviewed by a different research associate than at baseline.

Statistical Analysis

We analyzed all patients in the groups to which they were randomly assigned. Reported *P* values are 2-tailed and are considered statistically significant if they were less than 0.05. We analyzed data with SAS/STAT software, versions 8.2 and 9.1.3 (SAS Institute, Inc., Cary, North Carolina).

To describe the study sample and to compare groups, we used the chi-square test, Fisher exact test, 2-sample *t* test, and Wilcoxon rank-sum test, as appropriate. For the primary analyses, we used logistic and linear regression models to analyze dichotomous and continuous outcomes, respectively. A priori, we planned to assess for confounding of clinically important imbalances at baseline. Regression models adjusted for sex, alcohol assistance in the 3 months before enrollment, family history of alcoholism, any drug use, alcohol problem score, alcohol-attributable medical diagnoses, and mean number of drinks per day (for dichotomous consumption outcomes only). For the change in mean number of drinks per day at 12 months, we assessed possible effect modification by CIDI-determined dependence status. We included an interaction term in regression models and planned a stratified analysis if a significant interaction ($P \leq 0.100$) was identified. Post hoc, we tested the association between patient–interventionist sex concordance and the primary outcomes in unadjusted logistic and linear regression models (for dichotomous and continuous outcomes, respectively) only in patients in the intervention group.

We also conducted analyses using longitudinal mixed-effects models (linear for continuous outcomes and nonlinear for dichotomous outcomes), which included patients who may have completed 1 follow-up visit but not the other follow-up visit. Thus, we did not exclude patients with missing data at a single time point. We fit a random intercept model with an unstructured variance–covariance matrix.

We initially planned to enroll and randomly assign 500 persons, assuming that 4% of dependent patients in the control group would receive assistance, that all patients in the control group consumed an average of 2.2 drinks per day, and that withdrawal rates would be 10% and 20% at 3 and 12 months, respectively (42, 43). We estimated that 250 persons per group would allow us to detect an absolute increase in assistance of 12% with 87% power and a decrease of 0.6 drinks per day with 83% power. We repeated power calculations (but did not use a formal reassessment method) after enrolling 300 persons because the observed values of the outcomes and their SDs in the control group differed from those initially assumed, and withdrawal rates at 12 months were lower than anticipated. For the recalculations, we assumed that 40% of dependent patients in the control group would receive assistance, that these patients would consume an average of 6.3 drinks per day, and that the withdrawal rates would be 20% and 25% at 3 and 12 months, respectively. We determined that enrolling 175 persons per group would allow us to detect an absolute increase in assistance of 19% and a decrease in drinking of 2.9 drinks per day, both with 80% power. We decided to end further enrollment for the trial when we reached the new target.

Role of the Funding Sources

The study was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA RO1 12617) and a General Clinical Research center grant from the National

Table 1. **Characteristics at Enrollment of All Study Patients and of the Subgroup with Alcohol Dependence***

Variable	Overall		Patients with Alcohol Dependence	
	Control Group (*n* = 169)	Intervention Group (*n* = 172)	Control Group (*n* = 129)	Intervention Group (*n* = 132)
Demographic characteristics				
Women, *n (%)*	59 (35)	40 (23)	45 (35)	31 (23)
Mean age (SD), *y*	44 (11)	45 (11)	44 (10)	44 (10)
Race/ethnicity, *n (%)*				
Black	80 (47)	75 (44)	64 (50)	60 (45)
White	66 (39)	67 (39)	46 (36)	49 (37)
Hispanic	13 (8)	17 (10)	11 (9)	13 (10)
Unemployed during the past 3 mo, *n (%)*	104 (62)	112 (65)	83 (64)	91 (69)
Homeless ≥1 night during the past 3 mo, *n (%)*	39 (23)	47 (27)	34 (26)	44 (33)
Medical diagnoses				
Principal diagnosis most common at current admission, *n (%)*†				
Rule out myocardial infarction	30 (18)	31 (18)	26 (20)	22 (17)
Asthma, bronchitis, and COPD	21 (12)	15 (9)	10 (8)	12 (9)
Pancreatitis	13 (8)	20 (12)	13 (10)	19 (14)
Cellulitis	14 (8)	8 (5)	9 (7)	6 (5)
Diabetes	5 (3)	9 (5)	5 (4)	5 (4)
Alcohol-attributable diagnosis‡	20 (12)	31 (18)	17 (13)	28 (21)
Current admission for any alcohol-attributable diagnosis‡	66 (39)	90 (52)	57 (44)	80 (61)
Median lifetime comorbidity score (Q1–Q3)§	1 (0–2)	1 (0–2)	1 (0–2)	1 (0–2)
DSM-IV alcohol diagnoses during the past year, *n (%)*†				
Alcohol abuse	8 (5)	7 (4)	–	–
Alcohol dependence	129 (76)	132 (77)	–	–
No diagnosis	32 (19)	33 (19)	–	–
Alcohol consumption during the past 30 days†				
Median drinks/day (Q1–Q3)	3 (1–8)	4 (1–9)	5 (2–12)	5 (2–10)
Median drinks/drinking-day (Q1–Q3)	8 (5–14)	10 (6–15)	9 (6–16)	12 (7–16)
Median maximum drinks/occasion (Q1–Q3)	14 (8–24)	14 (9–24)	17 (12–24)	18 (12–24)
Alcohol-related characteristics				
AUDIT score ≥12, *n (%)*	114 (67)	116 (67)	100 (78)	105 (80)
Median score on readiness to change Taking Steps scale (Q1–Q3)†	30 (24–34)	29 (24–34)	31 (26–34)	31 (27–35)
Family history of alcoholism, *n (%)*‖	136 (82)	148 (88)	105 (83)	119 (93)
Median score for alcohol problems during the past 3 months (Q1–Q3)†	12 (3–24)	14 (4–29)	16 (9–27)	21 (9–34)
Drug use during the past 30 days, *n* (%)				
Cigarettes¶	129 (76)	128 (74)	105 (81)	102 (77)
Heroin or cocaine**	51 (30)	37 (22)	43 (33)	33 (25)
Any drug use††	106 (63)	89 (52)	86 (67)	75 (57)
History of psychiatric disorders or violence, *n (%)*				
Current panic disorder ‡‡	25 (15)	31 (18)	24 (19)	31 (23)
Current generalized anxiety disorder‡‡	126 (75)	121 (71)	109 (85)	104 (79)
Current substantial depressive symptoms§§	121 (72)	122 (71)	101 (79)	110 (83)
Current substantial PTSD symptoms‖‖	61 (36)	78 (45)	54 (42)	75 (57)
Any lifetime interpersonal violence (e.g., physical or sexual) ¶¶	124 (73)	115 (67)	102 (79)	96 (73)
Mean health-related quality-of-life score (SD)†				
Physical score	38 (9)	38 (9)	38 (9)	38 (9)
Mental score	40 (12)	40 (14)	38 (11)	37 (13)
Health care utilization during the past 3 months				
Alcohol assistance, *n (%)*†	34 (20)	52 (30)	33 (26)	52 (40)
Any alcohol treatment services, *n (%)***	45 (27)	60 (35)	42 (33)	60 (46)
Any psychiatric treatment, *n (%)*	39 (23)	44 (26)	34 (26)	42 (32)
Psychiatric hospitalization, *n (%)*	6 (4)	6 (4)	6 (5)	6 (5)
Medical hospitalization, *n (%)*	49 (29)	49 (28)	39 (30)	45 (34)
Median days hospitalized (Q1–Q3)	0 (0–1)	0 (0–1)	0 (0–2)	0 (0–2)

Table 1—**Continued**

Variable	Overall		Patients with Alcohol Dependence	
	Control Group (*n* = 169)	Intervention Group (*n* = 172)	Control Group (*n* = 129)	Intervention Group (*n* = 132)
Emergency department use, *n (%)*	79 (47)	76 (44)	65 (50)	67 (51)
Median emergency department visits (Q1–Q3), *n*	0 (0–2)	0 (0–1)	1 (0–2)	1 (0–2)

* AUDIT = Alcohol Use Disorders Identification Test; COPD = chronic obstructive pulmonary disease; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; PTSD = posttraumatic stress disorder; Q1 = quartile 1 (or 25th percentile); Q3 = quartile 3 (or 75th percentile).
† See the Methods section for a description of how this characteristic was measured.
‡ 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 amnestic 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) (29).
§ Determined by a validated questionnaire (20).
‖ Determined by the Family History-Research Diagnostic Criteria (21).
¶ Based on a response of "yes, every day in the past 30 days" to the question: "Do you currently smoke?"(22).
** Determined by the Addiction Severity Index (ASI) (23).
†† Determined by the ASI and includes use of heroin; methadone; other opiates or analgesics; barbiturates; sedatives, hypnotics, or tranquilizers; cocaine; amphetamines; marijuana or cannabis; or hallucinogens.
‡‡ Determined by the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (24).
§§ A score of ≥16 on the Center for Epidemiologic Studies Depression scale (25, 26).
‖‖ A score of ≥44 on the Post Traumatic Stress Disorder Checklist (27).
¶¶ Determined by adapted items from the Traumatic Life Events Questionnaire, revised (28).
*** Includes alcohol assistance, except for medications, plus hospitalization for detoxification (any type); participation in any detoxification program; or halfway-house services.

Center for Research Resources (M01 RR00533). The funding agencies did not contribute to study design; data collection, analysis, or interpretation; or the decision to submit the manuscript for publication.

RESULTS

Research associates approached 7824 persons (10 273 admissions) (**Figure**). Of 5813 patients who were screened, 986 (17%) reported drinking risky amounts of alcohol in the past month; 341 of these patients enrolled in the study. According to screening data, these persons were significantly more likely to be black (45% vs. 31%) and to drink greater amounts (median, 24 vs. 18 drinks/wk) than were the 183 eligible patients who declined enrollment.

Of patients who enrolled in the study, 172 were randomly assigned to the intervention group and 169 were randomly assigned to the usual care group. Six patients in the intervention group left the hospital before receiving the intervention. Over 12 months, 11 patients died and 90% ($n = 308$) of all enrolled persons completed at least 1 follow-up visit. According to baseline data, persons who completed any follow-up were significantly more likely to be unemployed (66% vs. 42%), have visited an emergency department in the past 3 months (48% vs. 18%), and have substantial symptoms of posttraumatic stress disorder (a score of ≥44 on the Post Traumatic Stress Disorder Checklist [43% vs. 21%]) (27) than were those who were lost to follow-up.

The randomized groups had similar characteristics at enrollment, except for sex, alcohol-attributable medical diagnosis, receipt of alcohol assistance, and drug use (**Table 1**). More than three fourths of patients had current alcohol dependence. At 12 months, 37% of 140 patients in the intervention group recalled discussing their drinking with a counselor; 28% of 145 patients in the control group reported such a discussion.

Alcohol Assistance (Dependent Patients Only)

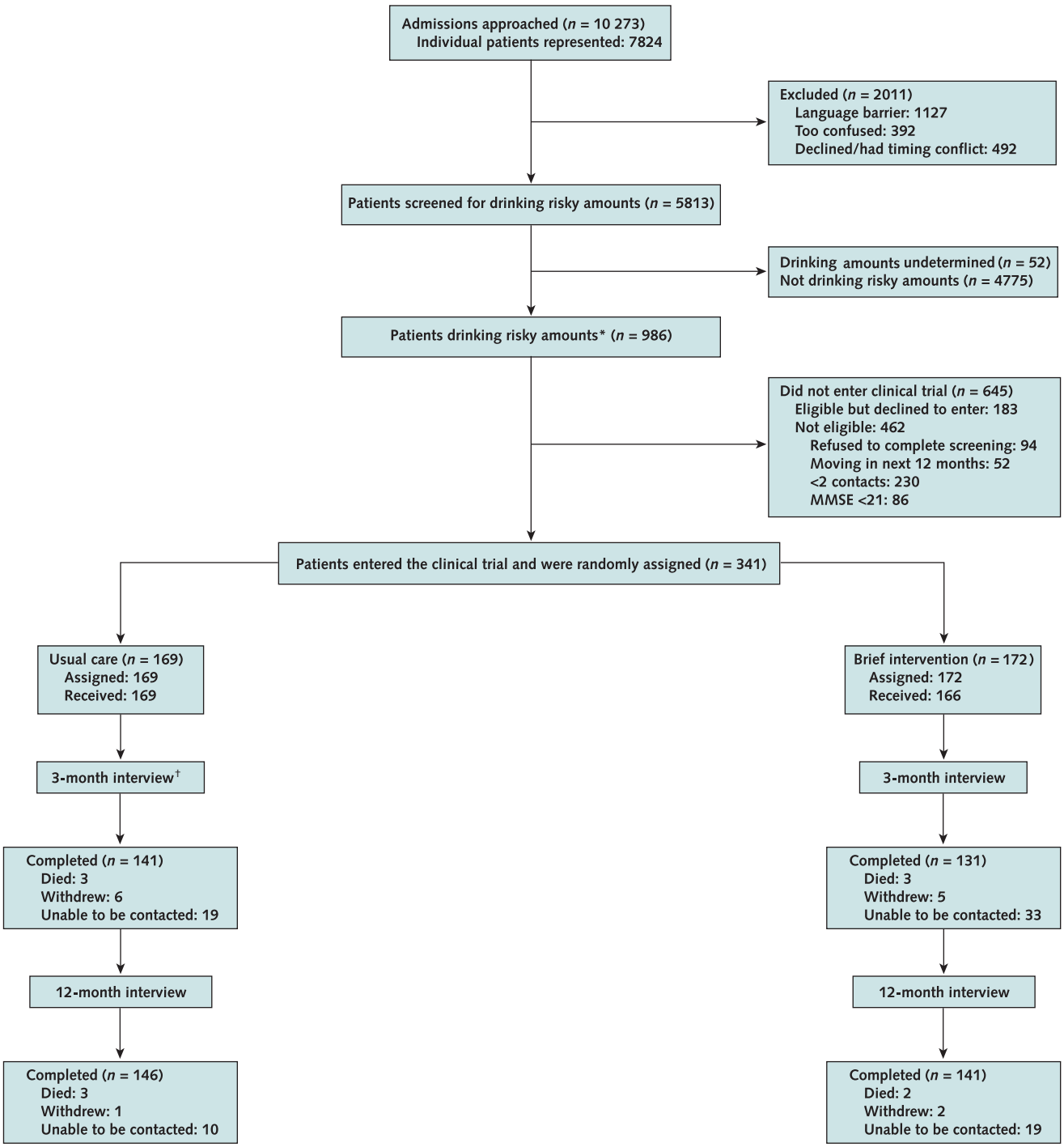
In adjusted analyses among dependent patients ($n = 204$), 49% of persons in the intervention group and 44% of those in the control group received alcohol assistance by 3 months (adjusted odds ratio, 1.2 [CI, 0.6 to 2.5] and adjusted intervention–control difference, 5% [CI, −8% to 19%]; $P = 0.55$) (**Table 2**). Results were also nonsignificant (adjusted odds ratio, 1.0 [CI, 0.4 to 2.6]; $P = 0.93$) in adjusted, nonlinear mixed-effects models, as was the interaction between the intervention and both follow-up points ($P = 0.34$). In adjusted analyses among patients with AUDIT scores of 12 or greater, 48% in the intervention group and 43% in the control group received alcohol assistance by 3 months ($n = 183$) (adjusted odds ratio, 1.3 [CI, 0.6 to 2.7] and intervention–control difference, 6% [CI, −9% to 20%]) ($P = 0.55$).

The types of assistance received did not significantly differ between groups (**Appendix**, available at www.annals.org). Subject–interventionist sex concordance did not significantly affect results (unadjusted odds ratio for same-sex vs. opposite-sex interventionist, 2.1 [CI, 0.9 to 5.0]) ($P = 0.099$).

Changes in Alcohol Consumption

The number of drinks per day decreased in both groups at 12 months, although there was no significant difference between groups in adjusted analyses (**Table 3**). There was no significant interaction between the intervention and alcohol dependence ($P = 0.24$). In adjusted, linear mixed-effects models, the impact of brief intervention (adjusted mean group difference, −1.0 drink/d [CI, −3.0

Figure. **Study flow diagram.**



Participants who dropped out at 3 months were permanently lost to follow-up. Participants who could not be contacted at 3 months may have been contacted at 12 months. MMSE=Mini-Mental State Examination. *During the first 7 months of the study, 22% of the 5813 total screened participants were screened; 19% of those screened during this time were drinking risky amounts. During the remainder of the study (when we changed screening criteria and screened 78% of the 5813), 17% were drinking risky amounts. †Analyses for assistance at 3 months included only participants with alcohol dependence; at 3 months, 112 persons in the control group and 98 persons in the intervention group had alcohol dependence. All other analyses included all randomly assigned participants with available data.

Table 2. **Receipt of Alcohol Assistance by 3 Months in Patients with Alcohol Dependence***

Analysis†	Numbers and Proportions		Odds Ratio (95% CI) (in Intervention Group)	Intervention–Control Difference (95% CI), *Percentage Points*	*P* Value
	Control Group	Intervention Group			
Unadjusted, % *(n/n)*‡	39 (44/112)	52 (50/97)	1.6 (0.9 to 2.8)	12 (−1 to 26)	0.08
Adjusted, %§	44	49	1.2 (0.6 to 2.5)‖	5 (−8 to 19)	0.55

* In adjusted analyses among patients with Alcohol Use Disorders Identification Test scores ≥12, 48% in the intervention group and 43% in the control group linked with alcohol assistance by 3 months ($n = 183$); adjusted odds ratio, 1.3 (95% CI, 0.6 to 2.7); intervention–control difference, 6 (CI, −9 to 20) percentage points ($P = 0.55$).
† There were 209 patients in the unadjusted analysis and 204 (110 control, 94 intervention) patients in the adjusted analysis.
‡ A total of 210 persons with alcohol dependence were interviewed at 3 months; however, 1 did not answer questions about alcohol assistance.
§ Adjusted for sex, alcohol assistance at 3 months before enrollment, family history of alcoholism, any drug use, alcohol problem score, and alcohol-attributable medical diagnoses.
‖ Results in this table are based on a complete-case analysis. In mixed-effects models, the adjusted odds ratio was 1.0 (CI, 0.4 to 2.6) ($P = 0.93$).

to 1.0], favoring patients in the control group) ($P = 0.33$) and the interaction between the intervention and both follow-up points ($P = 0.29$), were nonsignificant. Sex concordance did not significantly affect results (unadjusted mean difference for same-sex vs. opposite-sex interventionist, 1.1 drinks/d [CI, −2.0 to 4.1]; $P = 0.49$).

The intervention did not significantly improve the other consumption outcomes (**Table 3**). Patients in both groups decreased heavy drinking episodes and increased the number of days abstinent. There was a significant adjusted mean group difference in days abstinent favoring the patients in the control group.

Readiness to Change, Alcohol Problems, Quality of Life, and Health Care Utilization

In adjusted analyses at 12 months, the intervention did not significantly affect readiness to change, alcohol problems, physical or mental health-related quality of life, emergency department visits, or days hospitalized (data not shown).

Discussion

Screening for unhealthy alcohol use and providing brief intervention did not lead to receipt of assistance for alcohol-dependent medical inpatients and had no effect on consumption or nonconsumption outcomes among all patients. The 95% CIs for the intervention versus control differences did not consistently include important intervention effects.

Studies of the efficacy of brief intervention for unhealthy alcohol use in hospitalized patients have produced mixed results. Some studies of medical inpatients support its efficacy, whereas others do not (41, 44–49). A systematic review of controlled studies found an association between brief intervention and decreased alcohol-related problems but not alcohol consumption for inpatients on medical and other hospital services (7). Studies of inpatients on nonmedical services, however, have more consistently demonstrated the efficacy of brief intervention for decreasing alcohol consumption (9, 42, 50–52) and in-

Table 3. **Alcohol Consumption Outcomes at 12 Months in Patients with Unhealthy Alcohol Use**

Consumption Measures during the Past 30 Days	Mean Unadjusted Changes from Enrollment ± SE		Mean Adjusted Group Differences* (95% CI) ($n = 280$)	*P* Value
	Control Group ($n = 146$)	Intervention Group ($n = 141$)		
Decrease in drinks/d, *n*†	2.6 ± 0.8	1.8 ± 0.7	−1.5 (−3.7 to 0.6)‡	0.169
Decrease in heavy drinking episodes, *n*§	3.8 ± 0.9	3.4 ± 1.0	−1.7 (−4.4 to 0.9)	0.193
Increase in days abstinent, *n*	4.2 ± 1.0	2.5 ± 1.0	−2.9 (−5.7 to −0.1)	0.042
	Numbers and Unadjusted Proportions		Adjusted Odds Ratio‖ (95% CI)	
Drinking risky amounts, *n (%)*¶	93 (64)	87 (62)	1.1 (0.7 to 1.9)	0.69
Heavy drinking episodes, *n (%)*§	91 (62)	87 (62)	1.2 (0.7 to 2.0)	0.55
Abstinence, *n (%)*	40 (27)	42 (30)	0.9 (0.5 to 1.6)	0.78

* Adjusted for sex, alcohol assistance in the 3 months before enrollment, family history of alcoholism, any drug use, alcohol problem score, and alcohol-attributable medical diagnoses.
† Adjusted mean decreases in drinks per day: 3.1 ± 0.8 for patients in the control group and 1.5 ± 0.8 for patients in the intervention group.
‡ Results in this table are based on a complete-case analysis. In adjusted, linear mixed-effects models, the adjusted mean group difference for drinks per day was −1.0 drink per day (95% CI, −3.0 to 1.0; $P = 0.33$).
§ ≥5 drinks per occasion for men or ≥4 drinks per occasion for women and persons ≥ 66 years.
‖ Control group is the reference; adjusted for mean drinks per day at enrollment, sex, alcohol assistance in the 3 months before enrollment, family history of alcoholism, any drug use, alcohol problem score, and alcohol-attributable medical diagnoses.
¶ >14 standard drinks per week or ≥5 drinks per occasion for men; >7 drinks per week or ≥4 drinks per occasion for women and persons ≥ 66 years.

creasing entry into alcohol treatment entry and use of mutual-help groups (42, 50) (although some studies, including those done in emergency departments and trauma centers, have shown nonsignificant main effects) (9, 53).

Various factors might explain why the intervention did not have the anticipated effects. First, medical illnesses, hospitalization, or a research assessment that could have motivated patients contemplating change may have contributed to decreased alcohol consumption in both randomized groups. Second, we shaped the intervention to ensure its feasibility and replicability; a more extensive intervention might produce greater effects but would probably be harder to disseminate. Third, although it is uncertain, patients in the intervention group might have fared better if a medical clinician and/or a clinician they knew had provided the intervention. Fourth, the inadequacy of brief intervention in dependent drinkers may not be surprising because its efficacy for addressing dependence has been shown only in persons who seek treatment (54), unlike our patients.

Our study has several notable strengths. Unlike some studies, ours was randomized and used validated screening and diagnostic tools. We concealed randomization and analyzed patients in the groups to which they were randomized regardless of receipt of intervention. We conducted screening and brief intervention that could be reproduced in clinical practice (for example, no strict exclusion criteria, such as psychiatric illness or other drug use; pragmatic staff training; strategies that staff with a range of expertise could easily administer). We also examined the entire spectrum of unhealthy alcohol use (including dependence) in a diverse sample. This is especially important because screening in clinical practice identifies the entire spectrum and requires clinicians to address nondependent and dependent drinking.

Several methodological limitations should be considered when interpreting these results. Baseline imbalances existed between groups despite randomization; however, we adjusted for these imbalances. The complete-case analysis may be biased because we achieved 90%, not 100%, follow-up. Still, results from mixed-effects models were similar. Although all analyses compared patients in the groups to which they were randomly assigned, the alcohol assistance analysis was conducted only in the subgroup of patients for whom it was relevant—those with dependence (identified by the CIDI and not the AUDIT, which had been the stratification variable). Nevertheless, an analysis of patients with AUDIT scores of 12 or greater yielded similar results.

Research associates and patients could not be blinded; regardless, patients were generally interviewed at follow-up by a research associate whom they had not met; research associates would most likely have forgotten group assignment; and during follow-up, patients often could not correctly remember the group to which they had been randomized. Primary outcomes were self-reported but were assessed by trained research associates using validated, standardized procedures (55). Further, biological measures are not better than self-report for consumption outcomes, and administrative data cannot capture the range of assistance received, including mutual help.

We changed screening criteria early in the study, but the proportions of patients identified as drinking risky amounts were similar with the original criteria and the changed criteria (19% and 17%, respectively) (12). Screening may not have been sensitive for nondependent unhealthy alcohol use. Still, a recent systematic review of hospital studies also found that 17% of inpatients had positive results on screening for unhealthy alcohol use (56). Patients in the control group received slightly more than usual care (assessment and notification, as part of informed consent, that their unhealthy drinking made them eligible and that they could discuss their drinking with their clinicians). An assessment effect, however, would probably be minimal in these primarily non–treatment-seeking patients (54, 57). Further, many studies supporting the efficacy of brief intervention include a control group that provided informed consent and had substantial assessments. Lastly, generalizability may be limited to patients seen in many large, urban academic hospitals.

According to results from our study and others, brief intervention—a currently recommended practice—is inadequate for medical inpatients with unhealthy alcohol use (primarily dependence) to link with assistance for alcohol dependence and to reduce consumption and problems. In fact, our finding that patients in the intervention group had a smaller increase in abstinent days is consistent with the findings of a previous report of the possible harms of brief intervention (58). Although our study does not exclude the possibility that medical inpatients with nondependent unhealthy alcohol use (a small group) might respond to brief intervention, evidence for this benefit is unclear (7). Because of the need for effective brief approaches, future research should identify strategies that strengthen currently available interventions, such as multiple or booster sessions and components tailored to patients' needs. New interventions should involve staff with expertise in health behavior change and include all known efficacious therapies (for example, naltrexone or acamprosate for dependence). Further, studies of new efforts should use methods that separate intervention effects from the effects of research participation.

The high prevalence of unhealthy alcohol use among medical inpatients has important implications for acute and long-term patient care. Brief interventions alone may not be the solution. Screening hospitalized patients for unhealthy alcohol use has some value (for example, to identify drug interactions and risk for withdrawal). Brief intervention in hospitals, if supplemented with discussions with a clinician over time, also may serve as a seed that later helps catalyze change. But as a matter of policy, efforts focused on screening and brief intervention in hospitals should be

directed elsewhere, possibly toward more intensive interventions, follow-up care, or on subgroups of patients who are more likely to benefit. Additional research must identify the most effective interventions to address unhealthy alcohol use, particularly dependence, among hospitalized patients.

From Boston Medical Center and Boston University, Boston, Massachusetts; Smith College, Northampton, Massachusetts; DM-STAT, Malden, Massachusetts; and University of Pittsburgh, Pittsburgh, Pennsylvania.

Acknowledgments: The authors thank the staff and patients of the medical inpatient service and CARE Unit research associates at Boston Medical Center; the staff and house staff of the Boston University Internal Medicine Residency Training Program; and Karen Sullivan, Nicole Tibbetts, Alison Pedley, and other data management staff at DM-STAT, Malden, Massachusetts.

Grant Support: This study was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA RO1 12617) and a General Clinical Research Center grant from the National Center for Research Resources (M01 RR00533).

Potential Financial Conflicts of Interest: *Honoraria:* R. Saitz (Fusion Medical Education). All authors have received grant support from the National Institute on Alcohol Abuse and Alcoholism.

Requests for Single Reprints: Richard Saitz, MD, MPH, Boston Medical Center, 91 East Concord Street, Suite 200, Boston, MA 02118; e-mail, rsaitz@bu.edu.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Institute of Medicine.** Broadening the Base of Treatment for Alcohol Problems: Report of a Study by a Committee of the Institute of Medicine, Division of Mental Health and Behavioral Medicine. Washington, DC: National Academy Pr; 1990.
2. **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. [PMID: 15068984]
3. **Buchsbaum DG, Buchanan RG, Poses RM, Schnoll SH, Lawton MJ.** Physician detection of drinking problems in patients attending a general medicine practice. J Gen Intern Med. 1992;7:517-21. [PMID: 1403208]
4. **Saitz R.** Clinical practice. Unhealthy alcohol use. N Engl J Med. 2005;352:596-607. [PMID: 15703424]
5. **Maisto SA, Saitz R.** Alcohol use disorders: screening and diagnosis. Am J Addict. 2003;12 Suppl 1:S12-25. [PMID: 14972777]
6. **Wilk AI, Jensen NM, Havighurst TC.** Meta-analysis of randomized control trials addressing brief interventions in heavy alcohol drinkers. J Gen Intern Med. 1997;12:274-83. [PMID: 9159696]
7. **Emmen MJ, Schippers GM, Bleijenberg G, Wollersheim H.** Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: systematic review. BMJ. 2004;328:318. [PMID: 14729657]
8. **Longabaugh R, Woolard RE, Nirenberg TD, Minugh AP, Becker B, Clifford PR, et al.** Evaluating the effects of a brief motivational intervention for injured drinkers in the emergency department. J Stud Alcohol. 2001;62:806-16. [PMID: 11838918]
9. **Gentilello LM, Rivara FP, Donovan DM, Jurkovich GJ, Daranciang E, Dunn CW, et al.** Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. Ann Surg. 1999;230:473-80; discussion 480-3. [PMID: 10522717]
10. **Mitka M.** "Teachable moments" provide a means for physicians to lower alcohol abuse. JAMA. 1998;279:1767-8. [PMID: 9628694]
11. **Tunis SR, Stryer DB, Clancy CM.** Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA. 2003;290:1624-32. [PMID: 14506122]
12. **Saitz R, Freedner N, Palfai TP, Horton NJ, Samet JH.** The severity of unhealthy alcohol use in hospitalized medical patients. The spectrum is narrow. J Gen Intern Med. 2006;21:381-5. [PMID: 16686818]
13. **Smith KL, Horton NJ, Saitz R, Samet JH.** The use of the mini-mental state examination in recruitment for substance abuse research studies. Drug Alcohol Depend. 2006;82:231-7. [PMID: 16256278]
14. **Folstein MF, Folstein SE, McHugh PR.** "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189-98. [PMID: 1202204]
15. **Reinert DF, Allen JP.** The Alcohol Use Disorders Identification Test (AUDIT): a review of recent research. Alcohol Clin Exp Res. 2002;26:272-9. [PMID: 11964568]
16. **Friedmann PD, Saitz R, Gogineni A, Zhang JX, Stein MD.** Validation of the screening strategy in the NIAAA "Physicians' Guide to Helping Patients with Alcohol Problems". J Stud Alcohol. 2001;62:234-8. [PMID: 11332444]
17. **National Institute on Alcohol Abuse and Alcoholism.** Helping Patients with Alcohol Problems. A Health Practitioner's Guide. Bethesda, MD: National Institutes of Health; 2003.
18. **Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M.** Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption—II. Addiction. 1993;88:791-804. [PMID: 8329970]
19. **Miller WR, Rollnick S.** Motivational Interviewing: Preparing People for Change. 2nd ed. New York: Guilford Pr; 2002.
20. **Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW.** Can comorbidity be measured by questionnaire rather than medical record review? Med Care. 1996;34:73-84. [PMID: 8551813]
21. **Andreasen NC, Endicott J, Spitzer RL, Winokur G.** The family history method using diagnostic criteria. Reliability and validity. Arch Gen Psychiatry. 1977;34:1229-35. [PMID: 911222]
22. **Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S.** The validity of self-reported smoking: a review and meta-analysis. Am J Public Health. 1994;84:1086-93. [PMID: 8017530]
23. **McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, et al.** The Fifth Edition of the Addiction Severity Index. J Subst Abuse Treat. 1992;9:199-213. [PMID: 1334156]
24. **Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV 3rd, Hahn SR, et al.** Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. JAMA. 1994;272:1749-56. [PMID: 7966923]
25. **Boyd JH, Weissman MM, Thompson WD, Myers JK.** Screening for depression in a community sample. Understanding the discrepancies between depression symptom and diagnostic scales. Arch Gen Psychiatry. 1982;39:1195-200. [PMID: 7125849]
26. **Radloff LS.** The CES-D scale: a self-report depression scale for research in the general population. Applied Psychological Measurement. 1977;1:385-401.
27. **Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA.** Psychometric properties of the PTSD Checklist (PCL). Behav Res Ther. 1996;34:669-73. [PMID: 8870294]
28. **Kubany ES, Haynes SN, Leisen MB, Owens JA, Kaplan AS, Watson SB, et al.** Development and preliminary validation of a brief broad-spectrum measure of trauma exposure: the Traumatic Life Events Questionnaire. Psychol Assess. 2000;12:210-24. [PMID: 10887767]
29. **Adams WL, Yuan Z, Barboriak JJ, Rimm AA.** Alcohol-related hospitalizations of elderly people. Prevalence and geographic variation in the United States. JAMA. 1993;270:1222-5. [PMID: 8355385]
30. **American Psychiatric Association.** Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994.
31. **Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al.** The Composite International Diagnostic Interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry. 1988;45:1069-77. [PMID: 2848472]
32. **World Health Organization.** Composite International Diagnostic Interview (CIDI) (Core Version 2.0). Geneva, Switzerland: World Health Organization; 1996.
33. **Miller WR, Rollnick S.** Motivational Interviewing: Preparing People to

Change Addictive Behavior. New York: Guilford Press; 1991.
34. **McLellan AT, Alterman AI, Cacciola J, Metzger D, O'Brien CP.** A new measure of substance abuse treatment. Initial studies of the treatment services review. J Nerv Ment Dis. 1992;180:101-10. [PMID: 1737971]
35. **Miller WR.** Form 90. A Structured Assessment Interview for Drinking and Related Behaviors. Test Manual. Project Match Monograph Series, 5. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 1996.
36. **Sobell LC, Sobell MB.** Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: Litten RZ, Allen JP, eds. Measuring Alcohol Consumption: Psychosocial and Biochemical Methods. Totowa, NJ: Humana Pr, Inc.; 1992:41-69.-
37. **National Institute on Alcohol Abuse and Alcoholism.** Helping Patients Who Drink Too Much: A Clinician's Guide. Bethesda, MD: National Institutes of Health; 2005.
38. **Miller W, Tonigan J.** Assessing drinkers' motivation for change: the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). Journal of Psychological Addictive Behaviors. 1996;10:81-9.
39. **Miller WR, Tonigan J, Longabaugh R.** The Drinker Inventory of Consequences (DrInC). An Instrument for Assessing Adverse Consequences of Alcohol Abuse. Test Manual. Project MATCH Monograph Series 4. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 1995.
40. **Ware JE, Kosinkski M, Keller SD.** SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. 3rd ed. Lincoln, RI and Boston, MA: QualityMetric Inc. and the Health Assessment Lab 1998.
41. **Welte JW, Perry P, Longabaugh R, Clifford PR.** An outcome evaluation of a hospital-based early intervention program. Addiction. 1998;93:573-81. [PMID: 9684395]
42. **Elvy GA, Wells JE, Baird KA.** Attempted referral as intervention for problem drinking in the general hospital. Br J Addict. 1988;83:83-9. [PMID: 3345386]
43. **Fleming MF, Barry KL, Manwell LB, Johnson K, London R.** Brief physician advice for problem alcohol drinkers. A randomized controlled trial in community-based primary care practices. JAMA. 1997;277:1039-45. [PMID: 9091691]
44. **Heather N, Rollnick S, Bell A, Richmond R.** Effects of brief counselling among male heavy drinkers identified on general hospital wards. Drug Alcohol Rev. 1996;15:29-38. [PMID: 16203349]
45. **Chick J, Lloyd G, Crombie E.** Counselling problem drinkers in medical wards: a controlled study. Br Med J (Clin Res Ed). 1985;290:965-7. [PMID: 2858246]
46. **Watson HE.** A study of minimal interventions for problem drinkers in acute care settings. Int J Nurs Stud. 1999;36:425-34. [PMID: 10519687]
47. **Rowland N, Maynard AK.** Standardized alcohol education: a hit or miss affair? Health Promot Int. 1993;8:5-12.
48. **Kuchipudi V, Hobein K, Flickinger A, Iber FL.** Failure of a 2-hour motivational intervention to alter recurrent drinking behavior in alcoholics with gastrointestinal disease. J Stud Alcohol. 1990;51:356-60. [PMID: 2359309]
49. **Dunn CW, Ries R.** Linking substance abuse services with general medical care: integrated, brief interventions with hospitalized patients. Am J Drug Alcohol Abuse. 1997;23:1-13. [PMID: 9048144]
50. **Blondell RD, Looney SW, Northington AP, Lasch ME, Rhodes SB, McDaniels RL.** Can recovering alcoholics help hospitalized patients with alcohol problems? J Fam Pract. 2001;50:447. [PMID: 11350721]
51. **Hulse GK, Tait RJ.** Six-month outcomes associated with a brief alcohol intervention for adult in-patients with psychiatric disorders. Drug Alcohol Rev. 2002;21:105-12. [PMID: 12188988]
52. **Antti-Poika I, Karaharju E, Roine R, Salaspuro M.** Intervention of heavy drinking—a prospective and controlled study of 438 consecutive injured male patients. Alcohol Alcohol. 1988;23:115-21. [PMID: 3390235]
53. **Monti PM, Colby SM, Barnett NP, Spirito A, Rohsenow DJ, Myers M, et al.** Brief intervention for harm reduction with alcohol-positive older adolescents in a hospital emergency department. J Consult Clin Psychol. 1999;67:989-94. [PMID: 10596521]
54. **Moyer A, Finney JW, Swearingen CE, Vergun P.** Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. Addiction. 2002;97:279-92. [PMID: 11964101]
55. **Del Boca FK, Noll JA.** Truth or consequences: the validity of self-report data in health services research on addictions. Addiction. 2000;95 Suppl 3:S347-60. [PMID: 11132362]
56. **Roche AM, Freeman T, Skinner N.** From data to evidence, to action: findings from a systematic review of hospital screening studies for high risk alcohol consumption. Drug Alcohol Depend. 2006;83:1-14. [PMID: 16310323]
57. **Palfai TP, Fairholme C, Zisserson R, Saitz R.** Effects of normative feedback and motivation assessment on college student hazardous drinking [Abstract]. Alcohol Clin Exp Res. 2006;30:210A.
58. **Beich A, Thorsen T, Rollnick S.** Screening in brief intervention trials targeting excessive drinkers in general practice: systematic review and meta-analysis. BMJ. 2003;327:536-42. [PMID: 12958114]

Current Author Addresses: Drs. Saitz, Cheng, and Samet and Ms. Guerriero: Boston Medical Center, 91 East Concord Street, Suite 200, Boston, MA 02118.
Dr. Palfai: Psychology Department, Boston University, 648 Beacon Street, Boston, MA 02215.
Dr. Horton: Department of Mathematics, Smith College, 44 College Lane, Northampton, MA 01063.
Ms. Freedner: ORC Macro, 126 College Street, Burlington, VT 05401.
Dr. Dukes: DM-STAT, Inc., 1 Salem Street, Suite 300, Malden, MA 02148.
Dr. Kraemer: University of Pittsburgh School of Medicine, 230 McKee Place, Suite 600, Pittsburgh, PA 15213
Dr. Roberts: University of Pittsburgh School of Medicine, 200 Meyran Street, Suite 200, Pittsburgh, PA 15213.

Author Contributions: Conception and design: R. Saitz, T.P. Palfai, D.M. Cheng, K. Dukes, M.S. Roberts, J.H. Samet.
Analysis and interpretation of the data: R. Saitz, T.P. Palfai, D.M. Cheng, N.J. Horton, N. Freedner, K. Dukes, K.L. Kraemer, M.S. Roberts, J.H. Samet.
Drafting of the article: R. Saitz, T.P. Palfai, D.M. Cheng, N.J. Horton, N. Freedner, K.L. Kraemer, R.T. Guerriero.
Critical revision of the article for important intellectual content: R. Saitz, T.P. Palfai, D.M. Cheng, N. Freedner, K.L. Kraemer, M.S. Roberts, R.T. Guerriero, J.H. Samet.
Final approval of the article: R. Saitz, T.P. Palfai, D.M. Cheng, N. Freedner, N.J. Horton, K. Dukes, K.L. Kraemer, M.S. Roberts, R.T. Guerriero, J.H. Samet.
Provision of study materials or patients: R. Saitz.
Statistical expertise: D.M. Cheng, N.J. Horton, K. Dukes.
Obtaining of funding: R. Saitz.
Administrative, technical, or logistic support: R. Saitz, N. Freedner, R.T. Guerriero.
Collection and assembly of data: R. Saitz, N. Freedner, K. Dukes.

59. **Miller WR, Zweben A, DiClemente CC, Rychtarik RG.** Motivational Enhancement Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals with Alcohol Abuse and Dependence. Project MATCH Monograph Series, 2. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 1995.

Appendix: Description of the Intervention

We assigned the patients in the intervention group to a 30-minute session of brief motivational counseling based on the principles of motivational interviewing (19, 32) and strategies of motivational enhancement therapy (59). Interventionists had 1 to 6 years of clinical experience and received 2.5 days of skills-based training by motivational interviewing trainers. Interventionists practiced with one another and with 1 to 3 inpatients until they demonstrated the required skills to the trainers. A licensed clinical psychologist provided weekly group supervision that included a review of approximately 30% of patient interventions (that had been audiotaped) to prevent drift.

Through an empathic, respectful, and collaborative interviewing style, interventionists aimed to increase patients' awareness about the risks and problems of drinking, to help patients recognize the discrepancy between their current alcohol use and their values and goals, and to enhance patients' self-efficacy for change. Counseling sessions, which began only after interventionists gained the explicit permission of patients to discuss alcohol use, included feedback, an open discussion (lasting approximately 20 minutes), and construction of a change plan. Patients received individualized feedback based on data from their interviews (current alcohol consumption, Alcohol Use Disorders Identification Test [AUDIT] scores, readiness to change, and alcohol problems) and their clinical records (alcohol-related laboratory results and medical diagnoses). After being told how their drinking compared with sex-specific national norms and about their risks for harm (59), patients were invited to ask questions or to comment on the interventionist's feedback. Interventionists responded and followed with open-ended questions and reflective listening to convey empathy and to help the patients become more aware of how alcohol was influencing their lives. Finally, interventionists and patients generated a drinking change plan, which included anticipating obstacles and ranged from agreeing to consider changes in drinking to taking specific actions. When a patient granted permission, interventionists shared the contents of the session with his or her inpatient physician team and/or social workers. (We do not know how often this occurred). Within a week of the intervention, we mailed patients a copy of the change plans and a personalized letter from the interventionist that summarized the discussion and supported the patient's change efforts.

Postintervention Survey

To determine whether the intervention was consistent with the principles of motivational interviewing, we asked the patients to complete a survey immediately after the intervention, which they returned in a sealed envelope to the interventionist. (We informed the patients that the interventionists would not view the survey responses.) The survey included 12 questions that assessed the empathy of interventionists and whether they helped increase self-efficacy, instilled a sense of personal responsibility for change, reviewed the pros and cons of drinking, and facilitated change planning.

Of the 161 patients who completed the postintervention survey, most agreed that they felt understood (88%) or cared for (87%) by the interventionist. They agreed that the interventionist helped them believe that they could change their drinking (85%), see the pros and cons of their drinking (79%), recognize that changing drinking was their personal responsibility (83%), and generate a plan for their drinking (73%).

Recall of the Intervention

At the 12-month follow-up, we asked the patients in the intervention and control groups the following question to assess their recall of receiving counseling at study entry: "As you may recall, everyone in this study was interviewed in the hospital for about an hour and asked questions about their physical and mental health and about their drinking. After this, some people also met with a counselor to discuss their drinking. Did you meet with a counselor in the hospital to discuss your drinking?" Thirty-seven percent of 140 patients in the intervention group recalled discussing their drinking with a counselor; 28% of 145 patients in the control group reported such a discussion.

Types of Alcohol Assistance Received

The types of alcohol assistance received did not significantly differ between the intervention and control groups: In unadjusted analyses of alcohol-dependent patients, 5% of patients in the intervention group and 6% of patients in the control group received residential treatment; 8% and 6%, respectively, received outpatient care; and 46% and 36%, respectively, participated in a mutual-help group. Only 3 patients participated in an employee-assistance program, and 1 patient received naltrexone for drinking.

Alcohol & Alcoholism Vol. 42, No. 1, pp. 28–36, 2007
Advance Access publication 26 November 2006
doi:10.1093/alcalc/agl092

SCREENING AND BRIEF INTERVENTION ONLINE FOR COLLEGE STUDENTS: THE iHEALTH STUDY

RICHARD SAITZ[1,2,3]*, TIBOR P. PALFAI[2,8], NAOMI FREEDNER[1,2], MICHAEL R. WINTER[2,4], ALEXANDRA MACDONALD[2,8], JOHN LU[2,4], AL OZONOFF[2,5], DAVID L. ROSENBLOOM[2,6,7] and WILLIAM DEJONG[2,7]

[1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, [2]Youth Alcohol Prevention Center [3]Department of Epidemiology, [4]Data Coordinating Center, [5]Department of Biostatistics, [6]Join Together Program, [7]Department of Social and Behavioral Sciences, Boston University School of Public Health and [8]Department of Psychology, College of Arts and Sciences, Boston University, Boston, MA USA

(*Received* 17 *August* 2006; *first review notified* 6 *September* 2006; *in revised form* 18 *October* 2006; *accepted* 19 *October* 2006; *advance access publication* 26 *November* 2006)

Abstract — Aims: To test the feasibility of online alcohol screening and brief intervention (BI) by comparing (i) two approaches to inviting all students to be screened, and (ii) a minimal versus a more extensive BI. **Methods:** Freshmen students at one university were randomized to receive one of two types of email invitations to an online anonymous: (i) general health assessment, or (ii) alcohol-specific assessment. All were linked to the same alcohol screening survey. Those with unhealthy alcohol use (AUDIT ⩾8) were randomly assigned to minimal or more extensive online alcohol BI. **Results:** In both invitation groups (4008 students), 55% of students completed the online screening. Overall, 37% of men and 26% of women had unhealthy alcohol use. Compared to minimal BI, more extensive BI was associated with intention to seek help among men and with a greater increase in readiness to change among women. One month after BI, 75% of students completed another assessment, 33% of women and 15% of men with unhealthy alcohol use at baseline no longer had unhealthy alcohol use. There were no significant differences on drinking measures by BI randomization group. **Conclusions:** Over half of an entire freshman class of college students were reached by email and completed alcohol screening and brief intervention. Even an alcohol-specific invitation did not deter students. Although brief interventions that differed had some gender specific effects on readiness to change and intention, in general, unhealthy alcohol use decreased after brief intervention. Web screening and brief intervention show promise for addressing unhealthy alcohol use by college students.

INTRODUCTION

Unhealthy alcohol use, ranging from risky drinking through abuse and dependence, is common among college students (Hingson *et al*., 2002; Wechsler *et al*., 2002). Almost half (44%) of college students report one or more heavy drinking episodes in a 2 week period, while 37% meet criteria for alcohol abuse or dependence (Knight *et al*., 2002). Freshmen students in particular are at higher risk for heavy drinking and alcohol-related consequences, such as poisoning, injuries, and assaults (Pope, *et al*., 1990; Turrisi *et al*., 2000; Gruenewald *et al*., 2003).

Alcohol screening tests to identify unhealthy use have been validated in college samples (Fleming *et al*., 1991; Clements, 1998; Aertgeerts *et al*., 2000; Kokotailo *et al*., 2004), and brief interventions can reduce drinking and its consequences (Marlatt *et al*., 1998; Baer *et al*., 2001; Larimer *et al*., 2002), but these strategies have generally not been tested using campus-wide approaches that reach large numbers of students. Fully 93% of American college students regularly use the Internet (Harris Interactive, 2002), and ∼25% of Internet users aged 15–24 years have looked up information about drug or alcohol problems (The Kaiser Family Foundation, 2001). Several websites that assess drinking and provide normative feedback or brief intervention have been developed (Cunningham *et al*., 2000; Koski-Janne and Cunningham, 2001; Cloud *et al*., 2001), and several sites have been designed for college students specifically (Walters, 2000; Walters *et al*., 2000; Schroeder, 2001; Chiauzzi *et al*., 2005; Moore *et al*., 2005; Walters *et al*., 2005b). These sites vary in their theoretical underpinnings (e.g. health education versus brief intervention principles), length of the program, time to complete the program, and cost. Most of these sites have not yet been evaluated to determine the relationship between these various features and efficacy.

Thus, although the web may be a reasonable medium for achieving universal screening and increasing participation in brief intervention by college students, the best approaches to implementation remain unknown. For example, some studies have taken the approach of embedding an alcohol screening test within a general health assessment (Fleming *et al*., 1997). Whether such an approach is necessary and more efficacious than an approach focused on alcohol screening and brief intervention remains unknown; these approaches have not yet been compared in randomized trials.

Therefore, this study aimed to compare two different recruitment strategies for universal screening and to pilot test two levels of online brief intervention for college students with unhealthy alcohol use. We expected to find that electronic mail recruitment and web-based brief intervention for those with unhealthy alcohol use would be a feasible strategy for reaching freshman college students. We hypothesized that (i) an email invitation to a general-health screening would produce a greater student response than an invitation to an alcohol-specific screening; and (ii) a more extensive online

*Author to whom correspondence should be addressed at: Boston Medical Center, 91 East Concord Street, Suite No. 200, Boston, MA 02118, USA. Tel.: +1 617 414 7399 Fax: +1 617 414 4676; E-mail: rsaitz@bu.edu

brief intervention would be more effective than a minimal brief intervention.

METHODS

The Institutional Review and Privacy Boards of Boston University Medical Center approved this study with a waiver of consent elements and of documentation and HIPAA authorization.

Eligibility and enrollment

Eligible subjects were all freshman at a large urban private university in the Northeast who were ⩾18 years and had a valid university email address.

All eligible students received an introductory email from the Dean of Students on October 1, 2004. This email encouraged students to participate, explaining that those who completed the assessment survey would be eligible for a drawing to win one of twenty $50 Amazon.com gift certificates. The email also assured students that the study was anonymous and voluntary, meaning that university officials would not know whether a student participated and would not have access to individual survey responses.

The actual invitation to participate in the study, as well as a hypertext link to access the assessment survey was sent on October 4, 2004 by the university Information Technology Department, from the iHealth@bu.edu email address. The deadline for completing the assessment was October 15. Up to two email reminders were sent to non-respondents.

In order to access the survey, students created a study identification code (ID) that contained no unique identifiers. To create the ID, students were prompted to enter (i) the first letter of their first name, (ii) the last letter of their last name, (iii) the last 2 digits of their university identification number, and (iv) the first and last letters of town/city where they last attended school.

Study design

There were two randomizations in this study. First, freshman class email addresses were randomized (simple randomization) to receive either a general health screening invitation or an alcohol-specific invitation to participate in the study. Both the invitation emails contained a link to the same web-based alcohol assessment.

Second, immediately following this assessment, all students received an online individualized 'minimal' brief intervention (BI). Students with unhealthy alcohol use—that is, those with an Alcohol Use Disorders Identification Test (AUDIT) score of ⩾8—were subsequently randomized to receive either additional, 'more extensive' BI or no additional intervention. This second randomization was executed with stratification on AUDIT score (8–11 versus ⩾12) and invitation group (general health versus alcohol-specific).

In all cases, the brief intervention was followed by further assessment and an invitation to receive a follow-up survey one month later. Students were told that those completing this later survey would be eligible for a drawing to win one of ten $100 Amazon.com gift certificates or a an Apple iPod™. Students were then provided with a list of web links for online resources related to all of the health behaviors asked about in the initial assessment.

One month later, students who enrolled for follow-up were sent email invitations to complete a final assessment, again described as either a general health or alcohol assessment, consistent with the initial invitation randomization. Students received the same prompts to facilitate re-entry of their ID. If the ID entered at follow-up did not match a baseline ID, the ID was still accepted and the subject was permitted to complete the assessment. Non-responders were sent up to two reminder emails. Follow-up surveys were completed before students left the university for Thanksgiving break.

Assessments

We incorporated validity checks into the web assessment survey and required that all items on a page be completed in order to continue. Demographics (gender, ethnicity, and race) and subjects' past month health behaviors in the following areas were assessed: aerobic activity (Godin and Shepherd, 1985), sleep habits (Wolfson *et al*., 2003), smoking status, stress (Cohen *et al*., 1983), and alcohol use.

All subjects were screened for unhealthy alcohol use by means of the AUDIT (Reinert *et al*., 2002). The AUDIT is a brief instrument and has been validated in at least four studies in colleges (Fleming *et al*., 1991; Clements, 1998; Aertgeerts *et al*., 2000; Kokotailo *et al*., 2004); it is similarly accurate when done by computer or paper and pencil.(Chan-Pensley, 1999) Several AUDIT response options were modified to allow more detailed responses with regard to consumption quantity and frequency. Three additional questions were included to determine past-month alcohol consumption (number of drinks per typical week, maximum number of drinks on one occasion, and number of heavy drinking episodes) (National Institute on Alcohol Abuse and Alcoholism, 2003). An AUDIT score of ⩾8 was used as the cut-off for unhealthy alcohol use (Kokotailo *et al*., 2004). Subjects were then asked a single item regarding their readiness to change drinking [a 0–10 visual analogue scale included the following descriptions: 0 indicated 'No thought of changing'; 2 indicated 'Think I need to consider changing someday'; 5 indicated 'Think I should change, but not quite ready'; 8 indicated 'Starting to think about how to change my drinking patterns'; 10 indicated 'Taking action to change (e.g. cutting down)'] (Miller *et al*., 2002).

Students with unhealthy alcohol use were asked 13 items from the Young Adult Alcohol Problem Screening Test (YAAPST) (Hurlbut *et al*., 1992). The items were those that (i) were not redundant with the AUDIT, and (ii) had <50% prevalence but were not rare, as indicated by surveys completed in 2002 and 2003 by 547 freshmen students in introductory psychology courses at the same university.

Following the brief intervention there was a single item assessing help-seeking intention, three drinking age of onset questions (Wechsler; National Institute on Alcohol Abuse and Alcoholism, 2002), and an item about family history (National Institute on Alcohol Abuse and Alcoholism, 2002).

Readiness to change drinking was re-assessed post-BI. A single item assessed subjects' preference for obtaining personalized information about exercise, stress, and alcohol (e.g. web versus other). The follow-up assessment included reassessment of the domains covered at baseline, with the exception of characteristics that would not change (e.g. race).

Brief intervention (minimal versus more extensive)

The intervention was based on elements of BASICS (Dimeff *et al.*, 1999), motivational interviewing (Miller *et al.*, 2002), self-change approaches (Sobell and Sobell, 1993), and feedback about social norms (Baer *et al.*, 1991; Agostinelli *et al.*, 1995; Perkins *et al.*, 1999). Normative comparisons were based on data from the previously mentioned introductory psychology students. Minimal BI consisted of three web screens. The first two showed gender-specific graphic comparison to local norms for (i) number of drinks per typical week in the past month, and (ii) number of heavy drinking episodes in the past month (defined as ⩾5 drinks on an occasion for men, ⩾4 for women). The third web screen provided information about drinking guidelines, dependence symptoms, pregnancy, legal drinking age, coexisting medical conditions, and medication use.

Students in the more extensive BI group received three additional web screens: (i) highest blood alcohol level obtained in the past month and a chart describing the effects of alcohol on cognition and behavior at different levels, (ii) a graphic profile of consequences of drinking reported in the past year, with normative information about the percent of freshmen of the same gender experiencing these consequences (see Fig. 1), and (iii) the amount of money spent per week and per year on alcohol, the number of alcohol calories consumed per month (also reported as the equivalent number of sticks of butter), and the amount of time required on a treadmill to burn these calories and maintain current weight. After the feedback screens, all students were shown specific drinking guidelines and risk levels, a description of alcohol dependence symptoms, referral information, and a page of alcohol assistance web links. All students received feedback regarding sleep, stress, physical activity, and (for smokers) smoking.

Outcomes

The primary outcome measure was the proportion of students completing the entire initial assessment. Outcomes immediately post-intervention included readiness to change and intention to seek help (for those with unhealthy alcohol use). Additional outcomes included the proportions of students who both enrolled for and completed the 1-month follow-up; the prevalence of unhealthy alcohol use at follow-up; and changes in drinking outcomes from initial assessment to follow-up (drinks per week, heavy drinking episodes, and maximum number of drinks on a single occasion).

Statistical analysis

All analyses were carried out using SAS/STAT software version 8.2.(1999) Initial analyses consisted of descriptive statistics (means, standard deviations, medians, interquartile ranges, and proportions). Comparisons were performed with two-sample *t*-tests for continuous variables and chi square tests for categorical variables. Reported p-values are two-tailed, with a *P*-value <0.05 considered statistically significant. Among students with identical IDs, changes from initial assessment to follow-up) were compared with an analysis of covariance to test for intervention and time effects. Changes over time were compared using paired *t*-tests in analyses stratified on gender. In a sensitivity analysis we repeated these analyses after adding data from subjects with gender-concordant IDs that differed by one character between initial and follow-up assessment.

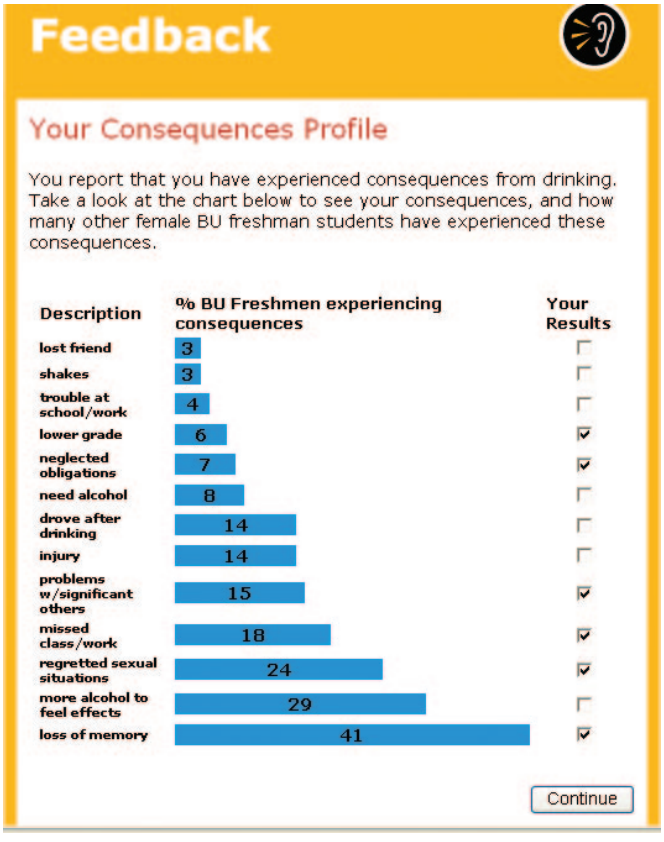


Fig. 1. Example of brief intervention web screen.

Sample size and statistical power

The pilot trial was designed to have 90% power, using a two-tailed alpha of 0.05, to detect a difference in response rates of 50 versus 55% between the two invitation approaches. For alcohol consumption measurement outcomes, the trial was designed to have 80% power to detect an effect size of 0.30 using a two-sample *t*-test with a two-tailed alpha at the 0.05 level. For the categorical outcome intention to seek help, there was 80% power for detecting an increase from 10 to 22%.

RESULTS

Subject characteristics and completion by invitation type

Invitation emails were sent to 4008 freshman students (2004 in each of the two invitation conditions; Fig. 2). There were only six unsuccessful transmissions in each of the two invitation groups. Overall, 2,194 (54.7%) students completed the initial assessment; this included 13 assessments completed by university students who had not been invited (12 followed the general-health invitation link and one followed the alcohol-specific link [they could not be removed from the analytic sample because there was no link maintained between the data and any identifiers). There were no statistically significant differences in baseline drinking characteristics, other health behaviors, or stress by invitation group (Table 1). There was no difference in completion rate by invitation group: 54.8% in the alcohol-specific invitation group versus 54.6% in the general health invitation group. A total of 95% of

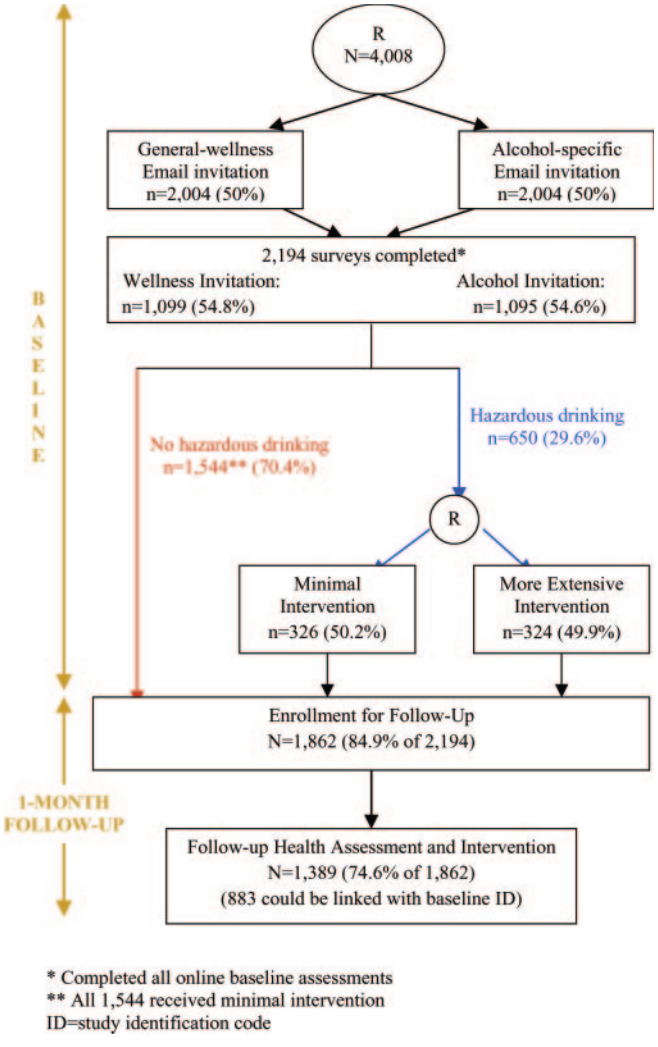


Fig. 2. Participant flow through study.

students who logged on to the assessment website completed the entire initial assessment.

Gender and race/ethnicity were similar to that of the freshman class, which was 40% male, 12% Asian, 62% White, and 9% Other/Hispanic (Table 1). One quarter of respondents were abstinent (males 27% and females 25%), and half had at least one heavy drinking episode in the past 30 days (males 53% and females 51%). Many (41%) stated they would use the web and 6% the phone to obtain personalized information about drinking.

Unhealthy alcohol use: subject characteristics, randomization, and immediate outcomes

About one-third of respondents (n = 650) reported unhealthy alcohol use (AUDIT ⩾8) (males 37% and females 26%). Of those with unhealthy alcohol use, 41% stated that they would use the web and 6% the telephone to obtain personalized information regarding drinking. Among the 650 subjects with unhealthy alcohol use, there were no differences in demographics or drinking characteristics between the two brief intervention groups (Table 2). Subjects with unhealthy alcohol use reported earlier initiation of drinking. By age 14, those with unhealthy alcohol use were more likely than those without, to have had a first drink (20.0 versus 11.2%, $P < 0.0001$), to have had a drink at least once a week (1.9 versus 0.53%, $P < 0.007$), and to have gotten drunk (9.7 versus 3.0%, $P < 0.0001$). Unhealthy drinkers were also significantly more likely to report a family history of alcohol problems (42.5 versus 31.8%, $P < 0.0001$).

Immediately post-intervention during the initial assessment, 8.9% of subjects with unhealthy alcohol use indicated that they intended to seek assistance in cutting down or stopping drinking. Among men, those in the more extensive BI group were significantly more likely than those in the minimal BI group to report intentions to seek help (13 versus 5%, $P = 0.016$)(Table 3). Immediately post-intervention, compared

Table 1. Characteristics of 2194 college freshmen completing online health assessment, by gender and invitation randomization group

	Gender		Invitation randomization group	
Characteristics	Male (n = 797)	Female (n = 1397)	General-health invitation (n = 1099)	Alcohol-specific invitation* (n = 1095)
Female (%)	—	100	62	65
Hispanic (%)	6	7	6	6
Race (%)				
White	71	76	74	73
Asian	20	14	15	17
Other	9	11	10	10
Exercise (days/week) (Mean, SD)	2.5 (2.0)	2.2 (1.9)	2.2 (1.9)	2.3 (2.0)
Exercise (minutes/day) (Mean, SD)	41.9 (40.4)	34.7 (33.9)	36.6 (36.0)	38.0 (37.2)
Sleep (hours/night) (Mean, SD)	6.6 (1.2)	6.5 (1.2)	6.5 (1.2)	6.5 (1.2)
Stress** (past 30 days) (%)	33	45	42	40
Current smoker*** (%)	9	10	9	10
Alcohol use characteristics:				
AUDIT ⩾8 (%)	37	26	29	30
Abstinent (%)	27	25	26	27
Heavy drinking episodes[a] (past 30 days) (%)	53	51	52	52

[a]⩾4 drinks for women and ⩾5 drinks for men, on one occasion.
*There were no statistically significant differences by invitation randomization group.
**Positive response to either of the following questions: In the past 30 days, have you (i) felt that you were unable to control the important things in your life, or (ii) felt difficulties were piling up so high that you could not overcome them?'
***Positive response to the question: Do you currently smoke cigarettes?'

Table 2. Characteristics of 650 drinkers with unhealthy alcohol use by gender and intervention randomization group

Characteristics	Gender		Intervention randomization group*	
	Male (n = 292)	Female (n = 358)	Minimal intervention (n = 326)	More extensive intervention (n = 324)
Female (%)	—	100	54	56
Hispanic (%)	6	5	4	6
Race (%)				
White	79	83	81	81
Asian	12	10	12	10
Other	9	7	7	9
AUDIT ⩾12	46	40	43	43
No. of drinks/week (past 30 days)	22.2 (32.4)	12.4 (8.4)	17.1 (22.4)	16.5 (23.7)
No. of most drinks per occasion (past 30 days)	12.6 (7.2)	8.4 (3.9)	10.3 (5.8)	10.3 (6.2)
No. of heavy drinking episodes (past 30 days)	6.2 (3.9)	5.9 (3.8)	6.3 (3.8)	5.8 (3.9)
No. of alcohol consequences[a](past year)	3.1 (2.2)	3.4 (2.1)	3.4 (2.3)	3.1 (2.0)
Readiness to change drinking**	2.0 (2.4)	2.1 (2.6)	2.3 (2.6)	1.9 (2.5)
Family history of alcoholism*** (%)	38	46	44	41
Age had first drink of alcohol[b] [mean (SD)]	14.6 (2.3)	14.8 (2.0)	14.8 (2.0)	14.6 (2.2)
Age first drank ⩾1 time per week[c] [mean (SD)]	16.0 (4.0)	15.2 (5.0)	15.6 (4.6)	15.5 (4.7)
Age first drunk[d] [mean (SD)]	15.3 (2.3)	15.4 (1.9)	15.3 (2.2)	15.4 (1.9)

[a]Represents the sum of the number of consequences endorsed by the subject.
[b]Response to the question: About how old were you when you had your first drink of alcohol, other than a few sips?'
[c]Response to the question: About how old were you when you first started drinking at least once a week?'
[d]Response to the question: About how old were you when you first got drunk?'
*There were no statistically significant differences by intervention randomization group.
**Response to the instruction: Each rung on this ladder represents where various people are in their thinking about changing their drinking. Fill in the number (on a scale from 0 to 10 with 0 indicating no thought of changing and 10 indicating taking action to change) that best indicates where you are now.'
***Positive response to the question: In your judgment, have any of the following people (mother/father/brothers/sisters, grandparents/aunts/uncles) been alcoholics or problem drinkers at any time in their lives?'

Table 3. Help-seeking and changes in readiness assessed immediately after completion of initial assessment and intervention for 650 subjects with unhealthy alcohol use*

	Minimal intervention (n = 326)	More extensive intervention (n = 324)	P-value for intervention group comparison
Change in readiness to change drinking ** (Mean,SD)	0.2 (1.2)	0.5 (1.5)	**0.01**
Male	0.1 (1.0)	0.3 (1.1)	0.21
Female	0.3 (1.3)	0.7 (1.6)	**0.03**
Intention to seek help*** (%)	8	10	0.26
Male	5	13	**0.02**
Female	10	8	0.54

*Unhealthy alcohol use includes the spectrum from risky drinking amounts through alcohol dependence, and is defined here as an AUDIT score ⩾8
**Response to the instruction: Each rung on this ladder represents where various people are in their thinking about changing their drinking. Fill in the number (on a scale from 0 to 10 with 0 indicating no thought of changing and 10 indicating taking action to change) that best indicates where you are now.'
***Positive response to the question: Do you intend to seek any assistance in cutting down or stopping drinking?'
Results in bold type indicate statistical significance.

to those in the minimal BI group, subjects with unhealthy alcohol use in the more extensive BI group had a significantly greater increase in readiness to change drinking (mean score change +0.5 (SD 1.5) versus +0.2 (SD 1.2), $P = 0.01$), an effect that was statistically significant for women but not men.

Follow-up: all subjects

Of the 2194 students who completed the initial assessment, 1874 (85.4%) agreed to complete the follow-up assessment. There were no differences in agreement to follow-up by gender (males 85.2% and females 85.5%) or by invitation randomization group (alcohol group 85.1%, general health group 85.7%). However, subjects with unhealthy alcohol use were less likely than other students to agree to follow-up (82.0 versus 86.9%, $P = 0.003$). Among students with unhealthy alcohol use at initial assessment, there was no difference between the two brief intervention groups in agreement to follow-up (minimal BI 81.6%, more extensive BI 82.4%).

Of the 1874 subjects who agreed to complete follow-up, 12 completed the initial assessment after the deadline and therefore were not eligible to complete the follow-up. Of the subjects who agreed and were eligible to complete follow-up, 74.6% (1389 out of 1862) completed the follow-up assessment (Fig. 2). At follow-up, 26.2% of subjects had unhealthy alcohol use (compared with 29.6% at baseline). Among men, a similar proportion had unhealthy alcohol use at baseline (36.6%) and follow-up (37.3%), but a smaller proportion of women at follow-up (20.3%) than at baseline (25.6%) had unhealthy alcohol use.

Follow-up and 1-month outcomes: subjects with unhealthy alcohol use

At follow-up, 47.4% of subjects (883/1862) provided a study ID identical to that at initial assessment, allowing data to be linked. There was no difference in the proportion of linked subjects between those with and without unhealthy alcohol use (44.5 and 48.6%, respectively, $P = 0.113$). An additional 19.1% (355 out of 1862) had study IDs at initial assessment and follow-up that did not differ by more than one character

Table 4. Outcomes of brief intervention for students with unhealthy alcohol use*

	Overall (n = 235)	Statistics for time effects	Minimal intervention (n = 109)	More extensive intervention (n = 126)	P-value for intervention group differences
AUDIT ⩾8 (absolute % change)		**95% CI**			
Men	–15	**8–21**	–17	–13	0.616
Women	–33	**25–41**	–33	–32	0.948
No. of drinks/week [mean (SD)]		***P*-value**			
Men	–1.22 (25.03)	0.621	+0.12 (8.83)	–2.15 (31.78)	0.783
Women	–1.68 (6.75)	**0.005**	–2.37 (6.10)	–0.97 (7.34)	0.914
Most no. of drinks per occasion [mean (SD)]					
Men	–0.43 (5.37)	0.421	–0.52 (4.51)	–0.36 (5.92)	0.546
Women	–0.48 (3.30)	0.099	–0.67 (3.53)	–0.28 (3.06)	0.807
No. of heavy drinking episodes [mean (SD)]					
Men	–0.10 (3.12)	0.753	–0.33 (2.94)	+0.07 (3.26)	0.332
Women	–0.83 (2.97)	**0.002**	–1.03 (2.91)	–0.63 (3.03)	0.948
Readiness to change drinking ** [mean (SD)]					
Men	+0.32 (2.42)	0.182	+0.02 (2.01)	+0.52 (2.67)	0.876
Women	+0.17 (2.53)	0.450	–0.07 (2.72)	+0.42 (2.30)	0.527

*Table is limited to the 235 subjects with identical study IDs at initial and follow-up assessments. All outcomes are expressed as changes from initial assessment to follow-up.
**Response to the instruction: Each rung on this ladder represents where various people are in their thinking about changing their drinking. Fill in the number (on a scale from 0 to 10 with 0 indicating no thought of changing and 10 indicating taking action to change) that best indicates where you are now.'
Results in bold type indicate statistical significance.

(and were of the same gender). There was no difference in proportion of IDs that were off by one character between subjects with and without unhealthy alcohol use (16.9 and 19.9%, respectively, $P = 0.127$).

We examined changes over time in the 235 subjects with unhealthy alcohol who had linked initial and follow-up assessments (Table 4). There were no statistically significant differences on drinking measures by brief intervention randomization group. Overall, unhealthy alcohol use decreased among men by 15% [95% confidence interval (CI) 8–21] and among women by 33% (95% CI, 25–41). Compared to the initial assessment, female students drank less per week (mean change –1.7 drinks (SD 6.8), $P = 0.005$) and had fewer heavy drinking episodes (mean change –0.8 episodes (SD 3.0), $P = 0.002$) at follow-up. Among women, a decrease in maximum drinks per occasion (mean change –0.5 drinks (SD 3.3), $P = 0.098$) was not significant. Among males, there were no changes in consumption.

We repeated the analyses of brief intervention outcomes including subjects who had study IDs that were identical to and that differed by one character from initial assessment. Again, there were no statistically significant differences on drinking measures by BI randomization group. Changes from initial assessment to follow-up were similar to those seen in primary analyses except that for women the decrease in maximum drinks per occasion became significant (mean change –0.72 drinks (SD 3.14), $P = 0.0026$).

DISCUSSION

In this study, a brief, web-based alcohol assessment and brief intervention reached a large proportion of students (more than half) despite participation not being required by school administrators, despite not providing compensation to all respondents, and despite a lack of any personal contact. One-third of these students had unhealthy alcohol use. We employed two approaches to inviting students to participate (general-health versus alcohol-specific email invitations). Despite concerns that students might be deterred from completing an 'alcohol' assessment, response rates for subjects were comparable in both groups. Similar proportions of students completed the intervention regardless of whether they had to complete a minimal or more extensive intervention. The more extensive brief intervention may have had greater efficacy increasing readiness to change drinking (particularly in women) and intention to seek help (in men), than minimal brief intervention. These two gender-specific results should be considered hypothesis generating. And most participating students completed a one-month follow-up assessment during which fewer students (more so for women) reported unhealthy alcohol use, regardless of brief intervention (BI) group, which should also be considered a preliminary finding.

Our overall findings resemble benefits identified by others of electronic (web, personal computer) brief interventions in colleges and other settings (Hester *et al*., 1997; Dimeff *et al*., 2000; Miller, 2001; Neighbors *et al*., 2004; Kypri *et al*., 2004; Chiauzzi *et al*., 2005). Computers have been used to assist physicians with providing interventions to patients (Hester *et al*., 1997; Dimeff *et al*., 2000). Personal computer-based BIs delivered by inviting college students to present to a location on campus also have promise. Six-weeks after a BI, subjects decreased number of drinks and heavy drinking episodes (Kypri *et al*., 2004). Though these decreases in consumption no longer reached significance at 6-months, reductions in alcohol consequences remained significant. In another study, a computerized BI resulted in reductions in consumption at 3 and 6 months (Neighbors *et al*., 2004). In a third study, when compared to a no-treatment group, a computer-based BI yielded alcohol consumption and consequence decreases similar to that of an in-person Alcohol Skills Training Program (ASTP) (Miller, 2001).

There are few studies of the use of the web to deliver BI to students wherever they may be. In one such study, Chiauzzi *et al*. (2005) found that although web-BI did not appear to have an overall effect, it was beneficial for certain

student subgroups and outcomes. For example, persistent heavy drinkers (as compared with less persistent heavy drinkers) experienced a more rapid decrease in average and peak consumption as well as in a composite score. Women in the intervention group reported a significantly greater reduction in alcohol-related problems. Precontemplation-stage drinkers in the intervention group reduced average consumption at a faster rate than those in the control group. At least one screening and BI website appears to have efficacy in several small controlled trials, with decreased alcohol consumption and alcohol-related problems in college freshmen (Henry *et al*., 2004; Steiner *et al*., 2005; Walters *et al*., 2005a). We are unable to compare the magnitude of change we observed with that of these studies, since effect sizes for those other studies have not yet been reported.

Our response rate of 54% is equivalent to that reported in Marlatt *et al*'s study, which recruited college subjects via mailed questionnaire (1998). Although today's postal mail studies are unlikely to duplicate such a rate, web-based research holds promise for maintaining or exceeding these rates. Our follow-up rate, at 74.6%, was similar to rates in other college web-based studies (72% at 4 weeks (Walters *et al*., 2005a), 80% at 6 weeks (Kypri *et al*., 2004), and 80% at 3 months (Chiauzzi *et al*., 2005)). Students with unhealthy alcohol use were less likely to be followed, but only by a small margin.

Although the optimal length and components of in-person BI are not known, very brief, single contacts are less efficacious than in-person, multi-contact BI's. (U.S. Preventive Services Task Force, 2004) Also, most efficacious BI's have some or all of the following components: personalized normative feedback, advice, and goal-setting.(U.S.Preventive Services Task Force, 2004) Adapting known efficacious in-person BI's to the web does not guarantee their continued effectiveness or brevity. The web-BI's with apparent efficacy consist of relatively brief, 10-min assessments, (Chiauzzi *et al*., 2005) with common components including personalized normative feedback and advice (Walters *et al*., 2005a). In this pilot study, we compared among three web brief intervention screens (minimal BI) and six web-BI screens (more-extensive BI); the additional screens provided feedback on highest blood alcohol level, consequences, costs and calories. In general, minimal and more extensive intervention were similarly effective. Whether this is because both forms of the intervention were relatively minimal, because the assessment (also minimal) had effects on drinking or whether intervention length is simply not important, cannot be determined from our data. Clearly an even more extensive intervention including specific advice and change strategies might have been more effective though extending the length might also limit public health impact (by limiting the number of students who participate). We also cannot determine whether specific components are essential. Therefore, further work is needed to identify optimal components and length (Del Boca *et al*., 2004).

Several limitations should be considered when interpreting the results of this study. First, there were invitation emails that were not successfully sent, and responses from students who had not been invited to participate. However, there were very few students in these groups. Second, we did not administer extensive alcohol assessments and therefore cannot characterize alcohol diagnoses or consumption details. In general, short, validated instruments were employed to maximize student participation. The readiness measure (examined as one of the outcomes), has not been well validated though it does seem to have face validity. Third, conclusions drawn from the follow-up assessment are limited by the study design. The design maintained anonymity in order to encourage enrollment, but this made measuring changes over time possible only for the sub-group of subjects who generated identical study IDs at initial assessment and follow-up. However, when we included subjects in analyses with IDs that did not differ by more than one character, results were similar. Nonetheless, the validity of analyses of drinking changes at follow-up may be threatened by this inability to link over half of the individual responses. However, our choice to implement an anonymous study was purposeful in that we wished to maximize the possibility that a student would complete the screening and assessment, thus replicating the way colleges would most likely use alcohol screening and BI if they intended maximum participation. If students had not had their data recorded anonymously and/or if they had to complete written informed consent or be presented with extensive confidentiality disclaimers (as required by the interpretation of human subjects regulations by some for such studies when done confidentially rather than anonymously) the level of participation would likely have been dramatically lower. We do recognize however, that web interventions could also be used by colleges for identifying students in need of assistance. A fourth potential limitation is that although the reductions in drinking in the linked sample are consistent with an effect of brief intervention, we cannot draw conclusions about the true efficacy of this approach without a no-intervention comparison group. And little should be made of the decrease in unhealthy use in the follow-up sample since these do not represent linked subjects. We did ensure that the period from initial assessment to follow-up did not include any breaks, major holidays, or final exams, when drinking patterns tend to show the greatest fluctuation (Del Boca *et al*., 2004). However, this design cannot substitute for an untreated control group. And the 1-month follow-up was short and we did not assess actual help-seeking over time, only intention to seek help; future studies should address long-term effects. Finally, the difference between the minimal and more extensive BI conditions was 3 web screens. This difference may not have been large enough to produce differences by BI group, potentially underestimating the effect of a more extensive BI. Furthermore, we cannot attribute differences in results to the difference in BI content versus length.

Despite these limitations, we were able to reach a large proportion of freshman students who completed alcohol screening, brief intervention, and one-month follow-up, and the study design also allowed comparison of two levels of intensity of brief intervention. Results also suggest that more extensive BI may have more efficacy for improving intention to seek help and readiness to change, but that even minimal BI may have decreased unhealthy alcohol use (though a secular trend cannot be excluded).

A number of questions remain regarding web-based screening and intervention for unhealthy alcohol use among college students. These include questions about widespread effectiveness (particularly for impact on alcohol consequences), necessary and sufficient components of web-based brief

interventions, duration, frequency and intensity of interventions, effects in important subgroups (e.g. gender), and comparisons of costs and effectiveness with other approaches. Nonetheless, universal alcohol screening and brief intervention online is feasible, the approach can reach many college students regardless of whether the message focuses on alcohol or general health, the feedback intervention can vary in length, and it may decrease unhealthy alcohol use. Web pages such as those used for this study could be available free of charge without restriction and could be implemented at colleges very inexpensively. With further study, the method could show promise for decreasing costs and consequences associated with drinking in a group of people with a high prevalence of heavy drinking and related consequences.

Acknowledgements — Preliminary results of this analysis were presented at the annual meeting of the Research Society on Alcoholism, June 25–29, 2005, Santa Barbara, CA. This study was supported by two grants from the National Institute on Alcohol Abuse and Alcoholism (P60 AA013759 and R01 AA12617). The authors of this manuscript reported no conflicts of interest. Funding to pay the Open Access publication charges for this article was provided by Boston University School of Public Health.

REFERENCES

Aertgeerts, B., Buntinx, F., Bande-Knops, J. *et al*. (2000) The value of CAGE, CUGE, and AUDIT in screening for alcohol abuse and dependence among college freshmen. *Alcoholism*: *Clinical and Experimental Research* **24**, 53–57.

Agostinelli, G., Brown, J. M. and Miller, W. R. (1995) Effects of normative feedback on consumption among heavy drinking college students. Journal of Drug Education **25**, 31–40.

Baer, J. S., Kivlahan, D. R., Blume, A. W. *et al*. (2001) Brief intervention for heavy-drinking college students: 4-year follow-up and natural history. *American Journal of Public Health* **91**, 1310–1316.

Baer, J. S., Stacy, A. and Larimer, M. (1991) Biases in the perception of drinking norms among college students. *Journal of Studies on Alcohol* **52**, 580–586.

Chan-Pensley, E. (1999) Alcohol-Use Disorders Identification Test: a comparison between paper and pencil and computerized versions. *Alcohol and Alcoholism* **34**, 882–885.

Chiauzzi, E., Green, T. C., Lord, S. *et al*. (2005) My student body: a high-risk drinking prevention web site for college students. *Journal of American College Health* **53**, 263–274.

Clements, R. (1998) A critical evaluation of several alcohol screening instruments using the CIDI-SAM as a criterion measure. *Alcoholism: Clinical and Experimental Research* **22**, 985–993.

Cloud, R. N. and Peacock, P. L. (2001) Internet screening and interventions for problem drinking: Results from the www.carebetter.com pilot study. *Alcoholism Treatment Quarterly* **19**, 23–44.

Cohen, S., Kamarck, T. and Mermelstein, R. (1983) A global measure of perceived stress. *Journal of Health and Social Behavior* **24**, 385–396.

Cunningham, J. A., Humphreys, K. and Koski-Jannes, A. (2000) Providing personalized assessment feedback for problem drinking on the Internet: a pilot project. *Journal of Studies on Alcohol* **61**, 794–798.

Del Boca, F. K., Darkes, J., Greenbaum, P. E. *et al*. (2004) Up close and personal: temporal variability in the drinking of individual college students during their first year. *Journal of Consulting and Clinical Psychology* **72**, 155–164.

Dimeff, L. A., Baer, J. S., Kivlahan, D. R. *et al*. (1999) *Brief Alcohol Screening and Intervention for College Students: a Harm Reduction Approach*, Guilford Press, New York.

Dimeff, L. A. and McNeely, M. (2000) Computer-enhanced primary care practitioner advice for high-risk college drinkers in a student primary health-care setting. *Cognitive and Behavioral Practice* **7**, 82–100.

Fleming, M. F., Barry, K. L. and MacDonald, R. (1991) The alcohol use disorders identification test (AUDIT) in a college sample. *International Journal of the Addictions* **26**, 1173–1185.

Fleming, M. F., Barry, K. L., Manwell, L. B. *et al*. (1997) Brief physician advice for problem alcohol drinkers. A randomized controlled trial in community-based primary care practices. *Journal of the American Medical Association* **277**, 1039–1045.

Godin, G. and Shephard, R. J. (1985) A simple method to assess exercise behavior in the community. *Canadian Journal of Applied Sports Sciences* **10**, 141–146.

Gruenewald, P. J., Johnson, F. W., Light, J. M. *et al*. (2003) Drinking to extremes: theoretical and empirical analyses of peak drinking levels among college students. *Journal of Studies on Alcohol* **64**, 817–824.

Harris Interactive. (2002) College Students Spend $200 Billion Per Year. Accessed at http://www.harrisinteractive.com/news/allnewsbydate.asp?NewsID=480 on March 6, 2006.

Henry, S., Lange, J. E. and Wilson, L. (2004) Evaluation of E-Chug integrated into two classroom-based alcohol interventions. Abstract presented at the U.S. Department of Education 18th Annual National Meeting on Alcohol, Other Drug Abuse and Violence Prevention in Higher Education, Washington, DC, October 16–19, 2004.

Hester, R. K. and Delaney, H. D. (1997) Behavioral Self-Control Program for Windows: results of a controlled clinical trial. Journal of Consulting and Clinical Psychology **65**, 686–693.

Hingson, R. W., Heeren, T., Zakocs, R. C. *et al*. (2002) Magnitude of alcohol-related mortality and morbidity among U.S. college students ages 18–24. *Journal of Studies on Alcohol* **63**, 136–144.

Hurlbut, S. C. and Sher, K. J. (1992) Assessing alcohol problems in college students. *Journal of American College Health* **41**, 49-58.

Knight, J. R., Wechsler, H., Kuo, M. *et al*. (2002) Alcohol abuse and dependence among U.S. college students. *Journal of Studies on Alcohol* **63**, 263–270.

Kokotailo, P. K., Egan, J., Gangnon, R. *et al*. (2004) Validity of the alcohol use disorders identification test in college students. *Alcoholism: Clinical and Experimental Research* **28**, 914–920.

Koski-Janne, A. and Cunningham, J. (2001) Interest in different forms of self-help in a general population sample of drinkers. *Addictive Behaviors* **26**, 91–99.

Kypri, K., Saunders, J. B., Williams, S. M. *et al*. (2004) Web-based screening and brief intervention for hazardous drinking: a double-blind randomized controlled trial. *Addiction* **99**, 1410–1417.

Larimer, M. E. and Cronce, J. M. (2002) Identification, prevention and treatment: a review of individual-focused strategies to reduce problematic alcohol consumption by college students. *Journal of Studies on Alcohol* **14**, 148–163.

Marlatt, G. A., Baer, J. S., Kivlahan, D. R. *et al*. (1998) Screening and brief intervention for high-risk college student drinkers: results from a 2-year follow-up assessment. *Journal of Consulting and Clinical Psychology* **66**, 604–615.

Miller, E. T. (2001) Preventing alcohol abuse and alcohol-related negative consequences among freshmen college students: using emerging computer technology to deliver and evaluate the effectiveness of brief intervention efforts. *Dissertation Abstracts International: Section B: The Sciences & Engineering* **61**, 4417.

Miller, W. R. and Rollnick, S. (2002) *Motivational Interviewing: Preparing People to Change*, Guilford Press, New York.

Moore, M. J., Soderquist, J. and Werch, C. (2005) Feasibility and efficacy of a binge drinking prevention intervention for college students delivered via the Internet versus postal mail. *Journal of American College Health* **54**, 38–44.

National Institute on Alcohol Abuse and Alcoholism.(2002) 2001–2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC).

National Institute on Alcohol Abuse and Alcoholism. (2003) *Helping Patients With Alcohol Problems: A Health Practitioner's Guide. NIH Pub. No. 03-3769*. NIAAA, Bethesda, MD.

Neighbors, C., Larimer, M. E. and Lewis, M. A. (2004) Targeting misperceptions of descriptive drinking norms: efficacy of a computer-delivered personalized normative feedback

intervention. *Journal of Consulting and Clinical Psychology* **72**, 434–447.
Perkins, H. W., Meilman, P. W., Leichliter, J. S. *et al*. (1999) Misperceptions of the norms for the frequency of alcohol and other drug use on college campuses. *Journal of American College of Health* **47**, 253–258.
Pope, H. G., Jr., Ionescu-Pioggia, M., Aizley, H. G. *et al*. (1990) Drug use and life style among college undergraduates in 1989: a comparison with 1969 and 1978. *American Journal of Psychiatry* **147**, 998–1001.
Reinert, D. F. and Allen, J. P. (2002) The Alcohol Use Disorders Identification Test (AUDIT): a review of recent research. *Alcoholism: Clinical and Experimental Research* **26**, 272–279.
SAS/STAT Software, Version 8.2 of the SAS System for Windows, Copyright © 1999–2001.
Schroeder, R. C. High-risk drinking: A research based approach to alcohol education on campus. Abstract presented in May, 2001 at the American College Health Association annual conference, Las Vegas, NV.
Sobell, M. B. and Sobell, L. C. (1993) *Problem Drinkers: Guided Self-Change Treatment*, Guilford Press, New York.
Steiner, J., Woodall, W. G. and Yeagley, J. A. (2005) The E-Chug: a randomized, controlled study of a web-based binge drinking intervention with college freshmen. Abstract presented at the Society for Prevention Research 13th Annual Meeting, Washington, DC, May 25–27, 2005.
The Kaiser Family Foundation. (2006) Generation Rx.com: how young people use the internet for health information. Accessed at http://www.kff.org/entmedia/20011211a-index.cfm on March 6, 2006.
Turrisi, R., Padilla, K. K. and Wiersma, K. A. (2000) College student drinking: an examination of theoretical models of drinking tendencies in freshmen and upperclassmen. *Journal of Studies on Alcohol* **61**, 598–602.
U.S. Preventive Services Task Force (2004) Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. *Annals of Internal Medicine* **140**, 554–556.
Walters, S. T. (2000) In praise of feedback: an effective intervention for college students who are heavy drinkers. *Journal of the American College of Health* **48**, 235–238.
Walters, S. T., Bennett, M. E. and Miller, J. H. (2000) Reducing alcohol use in college students: a controlled trial of two brief interventions. *Journal of Drug Education* **30**, 361–372.
Walters, S. T., Hester, R. K., Chiauzzi, E. *et al*. (2005a) Demon rum: high-tech solutions to an age-old problem. *Alcoholism: Clinical and Experimental Research* **29**, 270–277.
Walters, S. T., Miller, E. and Chiauzzi, E. (2005b) Wired for wellness: e-interventions for addressing college drinking. *Journal of Substance Abuse Treatment* **29**, 139–145.
Wechsler, H. (1991) *College Alcohol Study Questionnaire*. Harvard School of Public Health. Boston, MA, 199, p. 12.
Wechsler, H., Lee, J. E., Kuo, M. *et al*. (2002) 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. *Journal of American College of Health* **50**, 203–217.
Wolfson, A. R., Carskadon, M. A., Acebo, C. *et al*. (2003) Evidence for the validity of a sleep habits survey for adolescents. *Sleep* **26**, 213–216.

Alcohol Consumption and HIV Disease Progression

Jeffrey H. Samet, MD, MA, MPH,† Debbie M. Cheng, ScD,*‡ Howard Libman, MD,§ David P. Nunes, MD,‖ Julie K. Alperen, DrPh,* and Richard Saitz, MD, MPH*¶*

Objective: To assess the relation between alcohol consumption and laboratory markers of HIV disease progression.

Methods: We prospectively assessed CD4 cell counts, HIV RNA levels, and alcohol consumption for up to 7 years in 595 HIV-infected persons with alcohol problems recruited between 1997 and 2003. We investigated the relation of these markers of HIV disease progression to alcohol consumption using longitudinal regression models controlling for known prognostic factors, including adherence and depressive symptoms, and stratified by antiretroviral therapy (ART) use.

Results: Among subjects who were not on ART, heavy alcohol consumption was associated with a lower CD4 cell count (adjusted mean decrease of 48.6 cells/μL compared with abstinence; $P = 0.03$) but not with higher $\log_{10}$ HIV RNA. Among subjects who were on ART, heavy alcohol consumption was not associated with a lower CD4 cell count or higher $\log_{10}$ HIV RNA.

Conclusions: Heavy alcohol consumption has a negative impact on the CD4 cell count in HIV-infected persons not receiving ART. In addition to the known deleterious effects of alcohol on ART adherence, these findings suggest that avoiding heavy alcohol consumption in patients not on ART may have a beneficial effect on HIV disease progression.

Key Words: alcohol, CD4, HIV disease progression, HIV viral load

(*J Acquir Immune Defic Syndr* 2007;46:194–199)

Identifying modifiable factors that affect HIV disease progression provides opportunities to improve HIV treatment. Several such factors have been identified, including medication adherence, depressive symptoms, and hepatitis C coinfection.[1–3] Because antiretroviral therapy (ART) is a key determinant of the clinical course of HIV disease, understanding the impact of modifiable factors in its presence and absence is important.[4]

Alcohol use is common among HIV-infected persons,[5–7] and its impact on HIV disease progression has been examined in in vitro, animal, and human studies. Alcohol can adversely affect immunologic function in HIV-infected persons by various mechanisms,[8] including increased HIV replication in lymphocytes.[9,10] Research on the impact of alcohol on simian immunodeficiency virus (SIV) demonstrates increased viral loads in infected rhesus monkeys given intoxicating levels of alcohol at the time of viral inoculation compared with controls that received no alcohol.[11–16] These data indicate that the viral set point is increased and clinical deterioration is more rapid in alcohol-exposed monkeys.

In humans, the literature shows mixed effects of alcohol. Evidence from a large observational cohort of homosexual men, the Multicenter AIDS Cohort Study (MACS), before the advent of highly active ART, found no association between alcohol use and HIV disease progression.[17] We previously published cross-sectional data suggesting a negative impact of alcohol use on HIV disease progression in the era of highly active ART.[18] Alcohol consumption was associated with lower CD4 cell counts and higher HIV viral loads among those receiving ART. No comparable association was found for similar patients not on ART, however.

In this study, we report on the impact of alcohol use among HIV-infected persons with current or past alcohol problems prospectively assessed for up to 7 years to test the following hypothesis: alcohol consumption is independently associated with more rapid HIV disease progression as measured by the CD4 cell count and HIV viral load (HIV RNA).

Received for publication March 26, 2007; accepted June 14, 2007.

From the *Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA; † Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA; ‡Department of Biostatistics, Boston University School of Public Health, Boston, MA; §Divisions of General Medicine and Primary Care and Infectious Diseases, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA; ‖Section of Gastroenterology, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA; and the ¶Department of Epidemiology and Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA.

Support for this study came from the following grants from the National Institute of Alcohol Abuse and Alcoholism (NIAAA) of the National Institutes of Health: R01-AA13216 (Clinical Impact of HCV and Alcohol in HIV-Infected Persons), R01-AA11785 (Medication Adherence in Alcohol Abusing HIV Patients), R01-AA10870 (Enhanced Linkage of Alcohol Abusers to Primary Care), and K24-AA015674 (Impact of Alcohol Use on HIV Infection in US and Russia). This research was conducted in part at the General Clinical Research Center at Boston University School of Medicine (US Public Health Service [PHS] grant M01 RR00533) and at the Beth Israel Deaconess General Clinical Research Center (USPHS grant M01 RR01032).

Parts of these data were presented at the following scientific meetings: XV International AIDS Conference, Bangkok, Thailand, July 11–16, 2004; Research Society on Alcoholism National Meeting, Santa Barbara, CA, June 25–30, 2005; and the International Society for Biomedical Research on Alcoholism (ISBRA) International Congress, Sydney, Australia, September 10–13, 2006.

Correspondence to: Jeffrey H. Samet, MD, MA, MPH, Section of General Internal Medicine, Boston Medical Center, 91 East Concord Street, Suite 200, Boston, MA 02118 (e-mail: jsamet@bu.edu).

METHODS

Study Design

We analyzed the effect of alcohol consumption on CD4 cell counts and HIV RNA levels in 2 prospective cohorts of HIV-infected patients with alcohol problems.

Subject Recruitment

Subjects recruited between 1997 and 2003 were participants in the HIV Alcohol Longitudinal Cohort (ALC) or HIV Longitudinal Interrelationships of Viruses and Ethanol (LIVE) study. Both were prospective observational cohort studies of HIV-infected persons with current or past alcohol problems. Data were collected at study enrollment and then every 6 months from 1997 to 2001 for the HIV-ALC study and from 2001 to 2006 for the HIV-LIVE study.

The recruitment sites of the HIV-ALC (n = 349) have been previously reported[18] and were comparable to those in the HIV-LIVE study. HIV-LIVE subjects (n = 400) were recruited from the HIV-ALC study (n = 154, 38%) and the following sites: (1) the Diagnostic Evaluation Unit,[19] an intake clinic for HIV-infected patients at Boston Medical Center (BMC) (n = 88, 22%); (2) HIV Primary Care and Specialty Clinics at Beth Israel Deaconess Medical Center (BIDMC), Boston, MA (n = 31, 8%); and (3) other sites (n = 127, 32%), including a respite facility for homeless persons, a methadone clinic, BMC's primary care practices, referrals by friends, newspaper advertisements, and posted flyers at homeless shelters and HIV/AIDS social service agencies in the Boston area.

Eligibility criteria included (1) documented HIV antibody test by enzyme-linked immunosorbent assay (ELISA) and confirmed by Western blot (medical record or tested at enrollment), (2) affirmative responses to 2 or more CAGE alcohol screening questions[7] or physician investigator diagnosis of alcohol abuse or dependence, (3) ability to speak English or Spanish, and (4) at least 1 contact person to assist with follow-up. Exclusion criteria included (1) score <21 on the 30-item Mini-Mental State Examination,[20,21] (2) inability to provide informed consent or answer the interview questions, and (3) plans to move from the Boston area in the subsequent 12 months.

All subjects provided written informed consent before enrollment. The Institutional Review Boards of the BMC and BIDMC approved this study.

Subject Assessment

Subjects received an interviewer-administered baseline assessment in English or Spanish. For the Spanish interview, standardized scales were used when available; the remainder of the questionnaire was translated from English into Spanish and back-translated. The CD4 cell count and HIV viral load were obtained at the study visit if not available by review of medical records within the past 4 months.

Outcomes

The primary outcomes were CD4 cell count per microliter and $\log_{10}$ HIV RNA copies per milliliter, which are 2 laboratory markers of HIV disease progression. Measurement of HIV RNA (viral load) was performed using a branched-chain DNA (bDNA) assay or polymerase chain reaction (PCR).[22] The lower threshold of detection varied between 50 and 500 copies/mL over the course of the study. In analyses, undetectable HIV RNA levels were assigned half the value of the lower limit of detection. One secondary outcome, the CD4% (CD4 count/total lymphocyte count), was examined as a complementary measure of immune status that may better reflect one's risk for opportunistic infection in the setting of hypersplenism or postsplenectomy status.[23,24]

Primary Independent Variable

The main independent variable, alcohol consumption in the past 30 days, which was categorized as heavy, moderate, or abstinent, was assessed using a validated calendar method.[25] The "heavy" category was derived from the National Institute on Alcohol Abuse and Alcoholism definition of amounts that risk consequences (>14 drinks per week or ≥ 5 drinks on a single occasion for men <66 years old and >7 drinks per week or ≥ 4 drinks on a single occasion for men ≥ 66 years old and all women).[26] "Moderate" alcohol use was defined as any drinking less than heavy amounts.

Potential Confounders

Other variables included in regression analyses were gender; age; race (black, white, or other); HIV risk behavior (injection drug use, men having sex with men [MSM], or heterosexual sex); homelessness; 3-day adherence to ART; depressive symptoms; time since study enrollment; year of study entry; and participation in the HIV-ALC study, HIV-LIVE study, or both. Homelessness was defined as at least 1 night in a shelter or on the street in the past 6 months.[27] Current ART use was assessed with the question, "Are you currently taking antiretroviral medications for HIV?" Three-day adherence to antiretroviral medications was determined with the AIDS Clinical Trials Group Questionnaire for Adherence to Antiretroviral Medications.[28] Adherence was a dichotomous variable, where patients $<100\%$ adherent during the previous 3 days were considered not adherent. Depressive symptoms were measured using the Center for Epidemiologic Studies Depression (CES-D) scale.[29]

Additional information collected to characterize subjects more completely included current and lifetime alcohol dependence (Composite International Diagnostic Interview [CIDI] Lifetime and 12 Month),[30] current drug dependence (CIDI 12 Month), and use of ART before study entry.

Statistical Analysis

Descriptive baseline statistics (proportions, means, and standard deviations) were used to characterize the study sample overall and by alcohol consumption status. Continuous variables were compared between alcohol consumption groups using analysis of variance for normally distributed variables and, otherwise, by the Kruskal-Wallis test. χ^2 tests were used to compare categoric variables based on alcohol consumption level. We examined the relation between alcohol consumption and HIV disease progression outcomes by fitting 2 separate multivariable longitudinal regression models (on and not on ART) for each of 3 outcomes (CD4 cell count, CD4%, and $\log_{10}$ HIV RNA). Generalized linear mixed effects models

were used to account for the correlation from repeated observations on the same subject. Models were stratified based on ART status, a factor previously shown to modify the effect of alcohol consumption level on HIV progression.[18] Subjects consistently on or not on ART contributed to that ART status, respectively. For subjects who went on or off ART during the study, we took an approach to avoid potential carryover effects from ART. If a subject went from on to off, the on observations were included but the off observations were not. If a subject went from off to on, both observations were included. Observations from subjects who changed multiple times were only counted the first time a change was made.

Potential confounding factors were included in all regression models. Baseline CD4 cell count was a covariate in analyses of the CD4 outcome. Baseline CD4% was handled in a similar manner. Analyses of HIV viral load did not adjust for baseline $\log_{10}$ HIV RNA, because subjects with established infection had likely achieved their viral set points; thus, HIV RNA level would be expected to remain fairly constant over time. To minimize the potential for colinearity, we verified that no pair of variables included in the same regression model was highly correlated ($r > 0.40$). All analyses were conducted using 2-sided tests and a significance level of 0.05. Analyses were performed with the use of SAS software (version 9.1.3; SAS Institute, Cary, NC).

RESULTS

Characteristics for the cohort (n = 595) at study entry and stratified by level of alcohol consumption are listed in Table 1. The cohort included 25% women, 66% nonwhites, and 27% homeless persons, with a mean age of 41 years. Injection drug use was the HIV risk behavior in more than half of the subjects; for the remainder, the risk behavior was divided between MSM and heterosexual sex. Assessment of alcohol use in the past 30 days revealed heavy consumption in 30% of subjects and moderate consumption in 10%; 59% of subjects were abstinent. One quarter (128 of 595) of all subjects used alcohol and illicit drugs (heroin or cocaine), 19% used alcohol alone, and 15% used heroin or cocaine alone. The average daily alcohol consumption of subjects drinking in the past 30 days was 5.4 drinks. Of subjects with a history of injection drug use, 32% (114 of 354) had injected drugs in the previous 6 months.

The median number of assessment time points per subject was 9, and the maximum was 15. The number of observations included in the analyses for CD4 cell count was 1475 for the on ART group and 384 for the not-on-ART group (1273 and 350 respectively for the CD4% secondary outcome). The number of observations included in the analyses for HIV RNA was 1808 and 603 for the on ART group and for the not-on-ART group, respectively. The number of outcomes in the CD4 analyses was less than that in the HIV RNA analyses, because the former adjusted for baseline CD4 cell count. Thus, CD4 analyses only included outcome data from follow-up assessments. Seventy-three percent (1360 of 1859) of the CD4 test results and 71% (1710 of 2411) of the HIV RNA test results were collected within 1 month of the interview.

The results of the 6 regression analyses are presented in Table 2. Among those subjects not on ART, CD4 cell counts averaged 48.6 cells/μL lower for those with heavy alcohol use compared with abstinence ($P = 0.03$). Note that this difference does not represent change over time within individual subjects but, instead, the average difference between subjects with heavy alcohol use compared with abstinence across all time points, adjusting for baseline CD4 cell count among other covariates. There was not a significant association, however, between heavy alcohol consumption and viral load for subjects not on ART. No association was observed for moderate drinking and any of the outcomes in the not-on-ART group. For subjects on ART, no significant association was found for the primary outcomes at any level of alcohol consumption. Because depressive symptoms have been shown to be associated with HIV disease progression, we controlled for the CES-D score in our main analyses as a potential confounder. Because depression may be in the causal pathway between alcohol and HIV disease progression, however, we fit additional models without the CES-D score. In these analyses, among subjects not on ART, similar results were found (CD4 cell counts were an average of 58.4 cells/μL lower for those with heavy alcohol use compared with abstinence; $P = 0.02$). Among subjects on ART, nonsignificantly higher $\log_{10}$ HIV RNA was found in those with heavy alcohol consumption compared with those who abstained (adjusted mean increase = 0.13; $P = 0.09$). The secondary outcome, CD4%, did not show any significant relation with alcohol consumption in either group (in the analyses adjusted for depressive symptoms; see Table 2).

DISCUSSION

The findings demonstrate a clinically and statistically significant effect of heavy alcohol consumption on the CD4 cell count among persons not on ART. Controlling for medication adherence, this effect was not observed in those receiving ART.

Previous efforts to examine this relation in HIV-infected persons have been limited by the study design (eg, cross-sectional)[18] or timing (eg, before the advent of highly active ART).[17,31] The current study prospectively examined alcohol consumption at 6-month intervals among HIV-infected persons selected on the basis of alcohol use. Use of a cohort with past or current alcohol problems was chosen, because a hypothesized modest effect of alcohol on HIV disease progression would require a relatively large study population, careful alcohol measurement, and subjects with heavier levels of alcohol consumption. Other factors, including psychosocial issues, known to be associated with HIV disease progression were measured and controlled for in the multivariable analyses to assess the independent effect of heavy or moderate alcohol consumption on markers of HIV disease progression. This study comprehensively assessed alcohol consumption using validated self-report instruments;[25] extended follow-up; and inclusion of multiple time-varying covariates in the regression analyses, including measures of ART adherence and depressive symptoms.

The finding of an impact on the CD4 cell count in subjects not on ART supports the study hypothesis that alcohol consumption has a direct effect on HIV disease progression. The impact seems to be relatively modest, however, and was

TABLE 1. Baseline Characteristics of 595 Adults With HIV Infection and Current or Past Alcohol Problems, Overall and by Alcohol Consumption Status

	Total Sample 595	None 353 (59%)	Abstinent 61 (10%)	Heavy 180 (30%)	*P*
Age [y], mean (SD)	41 (7.4)	41 (7.3)	41 (8.2)	41 (7.4)	0.54
Male, No. (%)	446 (75)	252 (71)	53 (87)	141 (78)	0.02
Race, No. (%)					
Black	245 (41)	141 (40)	24 (39)	80 (44)	
White	202 (34)	115 (33)	21 (34)	66 (37)	0.30
Other	147 (25)	97 (27)	16 (26)	34 (19)	
Homeless, No. (%)	163 (27)	92 (26)	12 (20)	59 (33)	0.09
CD4 count [cells/μL], median (IQR)*	372 (209, 565.5)	378.5 (209, 564)	408.5 (268, 624)	337 (184, 539)	0.15
CD4%, median (IQR)†	22 (14, 30)	22 (14, 30)	25 (17, 32)	21 (14, 30)	0.42
Log_{10} HIV RNA, mean (SD)‡	3.3 (1.2)	3.2 (1.2)	3.0 (1.2)	3.5 (1.2)	0.02
HIV viral load [copies/mL], median (IQR)‡	1400 (0, 19,662)	963 (0, 17,988)	976 (0, 9554)	3031 (94, 31,490)	0.96
HIV risk group, No. (%)					
Heterosexual/blood	140 (24)	76 (22)	12 (20)	52 (30)	
Injection drug use	313 (54)	207 (61)	28 (46)	78 (45)	0.001
MSM	123 (21)	59 (17)	21 (34)	43 (25)	
CES-D, mean (SD)	22 (13)	22 (13)	20 (14)	23 (13)	0.28
Current alcohol dependence, No. (%)§	29 (12)	14 (9)	0 (0)	15 (22)	0.005
Lifetime alcohol dependence, No. (%)§	222 (91)	135 (89)	23 (96)	64 (93)	0.42
Current drug dependence, No. (%)§	103 (42)	65 (43)	6 (25)	32 (46)	0.18
Drinks per day, median (IQR)	0 (0, 0.7)	0 (0, 0)	0.17 (0.07, 0.40)	2.6 (0.8, 6.8)	<0.0001
Drinks per day, mean (SD)	2.2 (8.2)	0 (0)	0.4 (0.5)	7.2 (13.7)	<0.0001
Currently receiving ART, No. (%)	355 (60)	218 (62)	37 (61)	100 (56)	0.38
100% adherent, past 3 days, No. (%)	249 (70)	170 (78)	20 (54)	59 (59)	0.0002
ART adherence status, No. (%)					
Not on medications	239 (40)	135 (38)	24 (39)	80 (44)	
On medications, not adherent	105 (18)	47 (13)	17 (28)	41 (23)	0.0007
On medications, adherent	249 (42)	170 (48)	20 (33)	59 (33)	
ART ever, No. (%)‖	467 (81)	283 (83)	47 (77)	137 (79)	0.39
Study cohort, No. (%)					
HIV-LIVE only	245 (41)	152 (43)	24 (39)	69 (38)	
HIV-ALC only	195 (33)	125 (35)	17 (28)	53 (29)	
Both	154 (26)	76 (22)	20 (33)	58 (32)	0.06

*n = 572 (Total), n = 338 (Abstinent), n = 58 (Moderate), n = 175 (Heavy).
†n = 568 (Total), n = 336 (Abstinent), n = 58 (Moderate), n = 173 (Heavy).
‡n = 557 (Total), n = 330 (Abstinent), n = 55 (Moderate), n = 171 (Heavy).
§n = 245 (Total), n = 152 (Abstinent), n = 24 (Moderate), n = 69 (Heavy).
‖n = 577 (Total), n = 342 (Abstinent), n = 61 (Moderate), n = 174 (Heavy).

not observed for moderate drinking or among persons on ART after controlling for adherence and other known associations with HIV disease progression. It may be difficult to detect a modest effect of heavy drinking on CD4 cell count among those on ART, because the large beneficial effect of ART on CD4 cell count may overcome any deleterious effect of alcohol. Only 10% of the cohort reported moderate levels of alcohol consumption. Such a low prevalence of moderate alcohol use is not surprising, given the fact that most of the subjects in this cohort met criteria for lifetime alcohol dependence. Individuals with current or past alcohol dependence are likely to abstain or drink heavily.[32] The modest number of moderate drinkers in the cohort limited our ability to assess the impact of moderate alcohol consumption on HIV disease progression, however. In particular, among subjects not on ART, we are unable to conclude that there is no effect of moderate drinking because it is likely that the study was underpowered to detect effects of the observed magnitude. Nevertheless, it is noteworthy that the magnitude and direction of the CD4 cell count change among those on ART with moderate drinking suggest that a larger study would be unlikely to find a detrimental effect.

The mechanism of alcohol's impact on the CD4 cell count is unclear. Attributing it to decreased medication adherence or to likelihood of receiving ART would not explain the finding of lower CD4 cell counts in persons not on ART. Attributing lower CD4 cell counts to a decreased likelihood of receiving ART should be adequately addressed by controlling for baseline CD4 cell count. Evidence from Bagby et al[16] in macaque monkeys demonstrating clinical deterioration of immune function in SIV-infected animals with chronic alcohol exposure suggests a more direct effect on immune function.

TABLE 2. Adjusted Mean Differences in CD4% Cell Count and Log_{10} HIV RNA for Subjects With Moderate and Heavy Alcohol Consumption Compared With Abstinent in the Past Month, Stratified by Receipt of ART

		CD4 Cell Count*†		CD4% Cell Count*†		Log_{10} HIV RNA*‡	
ART Status	Alcohol Consumption Status	Adjusted Mean Difference (vs. Abstinent)* (SE)	*P*	Adjusted Mean Difference (vs. Abstinent) (SE)	*P*	Adjusted Mean Difference (vs. Abstinent) (SE)	*P*
On ART‡	Abstinent	—	—	—	—	—	—
	Moderate	11.5 (13.8)	0.40	0.35 (0.35)	0.29	0.032 (0.08)	0.67
	Heavy	−1.2 (11.0)	0.92	0.35 (0.33)	0.32	0.12 (0.07)	0.10
Not on ART	Abstinent	—	—	—	—	—	—
	Moderate	−25.8 (17.7)	0.15	−0.80 (0.83)	0.35	−0.113 (0.08)	0.16
	Heavy	−48.6 (21.9)	0.03	−1.0 (5.1)	0.23	0.007 (0.08)	0.92

*Adjusts for gender; age; race; HIV risk behavior; homelessness; depressive symptoms; time since study enrollment; year of study entry; and participation in the HIV-ALC study, HIV-LIVE study, or both.
†Also adjusts for baseline value of the outcome.
‡Also adjusts for 3-day self-reported adherence.

The absence of a correlation of heavy alcohol use with HIV viral load in our study contrasts with the findings of an increased viral set point associated with heavy drinking in these animal studies. This difference may be attributable to the fact that under experimental conditions, the monkeys were all intoxicated at the time of SIV inoculation, a scenario that was not likely the case for all subjects in this observational cohort of persons with chronic HIV infection. The absence of impact on HIV viral load among those not on ART also suggests that the decrement in CD4 cell count is not mediated by an increased viral load, which is an important determinant of the trajectory of CD4 cell count decline. Findings from this cohort raise the possibility that the effect of alcohol on the CD4 cell count is CD4 cell specific or related to the nonspecific lymphopenia associated with alcohol use.[33] The absence of a significant association of alcohol consumption on CD4% suggests that the alcohol effect observed may relate to a decrease of lymphocytes overall as opposed to a selective decrement of cells.

The observation that heavy alcohol consumption has an effect on the CD4 cell count in HIV-infected patients not receiving ART is in contrast to that previously described in the medical literature. One of the earliest reports on this subject was by Kaslow et al,[17] in which MACS participants were examined every 6 months to assess the relation between alcohol and AIDS-defining illnesses. No effect was found among the 1706 HIV-infected men; however, the level of alcohol exposure was unclear. The highest alcohol consumption considered was 2 or more drinks per day, and mean consumption was not described. This level of consumption contrasts with an average of 7 drinks per day in the current study's heavy drinkers.

In the analyses excluding depressive symptoms, the finding of an association of borderline significance between heavy drinking and a modest increase of HIV viral load among subjects on ART is consistent with the findings of Chander et al,[34] in which heavy alcohol consumption correlated with significantly less viral suppression in cross-sectional analyses. These authors attributed this finding to the poorer adherence expected among heavy drinkers on ART. This is a possibility, and such an association in the current study may have been attenuated by the inclusion of ART adherence as a covariate in the regression analysis.

One limitation of this study was that we did not enroll HIV-infected persons in a cohort at the time of seroconversion (ie, an inception cohort) and follow their alcohol use, CD4 cell counts, and viral loads over time. To address this limitation, analyses controlled for baseline CD4 cell count. Lack of an inception cohort also allows the possibility that participants in the not-on-ART group may have been exposed to ART in the past but were no longer receiving it at the time of study enrollment. It is unclear how such exposure would alter the current findings. As in any observational study, the effects of heavy alcohol use could be confounded. It is possible that we did not adequately control for all potential contributing factors to HIV disease progression. Multiple characteristics known to be associated with HIV disease progression were included as covariates in the multivariable analyses, however. Another limitation is the fact that alcohol use was assessed 30 days before the interview and CD4 cell counts and HIV RNA levels, by study design, could have been obtained up to 4 months before the interview. Most of these outcomes were within 1 month of the subject interview, however.

Although alcohol consumption in HIV-infected persons is common,[5,7,35] heavy consumption in US cohorts is less frequent than in countries in which HIV infection and high per capita alcohol consumption coexist (eg, Russia, South Africa).[36,37] A modest impact of a common problem (heavy alcohol use) in patients with a prevalent disease (HIV infection) can have major public health consequences. Heavy alcohol use is a potentially modifiable factor that seems to have a modest impact on HIV disease progression. Based on our findings, we believe that HIV-infected persons who drink alcohol heavily and are not on ART might decrease their risk of disease progression if they abstain from alcohol use. There is extensive evidence of the efficacy of a brief intervention for unhealthy alcohol use in nondependent drinkers in medical settings and of the efficacy of psychosocial and pharmacologic treatments for alcohol dependence.[38,39] Although limited evidence demonstrates that intervention for alcohol problems in people with HIV is effective, its implementation among HIV-infected populations seems to be a worthwhile goal.[40] In

addition to the known deleterious effects of alcohol on ART adherence, these findings suggest that avoiding heavy alcohol consumption in patients not on ART may have a beneficial effect on HIV disease progression.

ACKNOWLEDGMENTS

The authors appreciate the contributions of Caitlin McDonnell and Mary Christine Sullivan as research associates on the project; John Vidaver, Vincent Faber, and Emily Quinn for data management and statistical programming; and Carly Bridden for assistance in the preparation of the manuscript.

REFERENCES

1. Greub G, Ledergerber B, Battegay M, et al. Clinical progression, survival, and immune recovery during antiretroviral therapy in patients with HIV-1 and hepatitis C virus coinfection: the Swiss HIV Cohort Study. *Lancet*. 2000;356:1800–1805.
2. Cook JA, Grey D, Burke J, et al. Depressive symptoms and AIDS-related mortality among a multisite cohort of HIV-positive women. *Am J Public Health*. 2004;94:1133–1140.
3. Ickovics JR, Hamburger ME, Vlahov D, et al. Mortality, CD4 cell count decline, and depressive symptoms among HIV-seropositive women: longitudinal analysis from the HIV Epidemiology Research Study. *JAMA*. 2001;285:1466–1474.
4. Palella FJ Jr, Delaney KM, Moorman AC, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med*. 1998;338: 853–860.
5. Conigliaro J, Gordon AJ, McGinnis KA, et al. How harmful is hazardous alcohol use and abuse in HIV infection: do health care providers know who is at risk? *J Acquir Immune Defic Syndr*. 2003;33:521–525.
6. Krupitsky EM, Horton NJ, Williams EC, et al. Alcohol use and HIV risk behaviors among HIV-infected hospitalized patients in St. Petersburg, Russia. *Drug Alcohol Depend*. 2005;79:251–256.
7. Samet JH, Phillips SJ, Horton NJ, et al. Detecting alcohol problems in HIV-infected patients: use of the CAGE questionnaire. *AIDS Res Hum Retroviruses*. 2004;20:151–155.
8. Goforth HW, Lupash DP, Brown ME, et al. Role of alcohol and substances of abuse in the immunomodulation of human immunodeficiency virus disease: a review. *Addictive Disorders and Their Treatment*. 2004;3: 174–182.
9. Cook RT, Stapleton JT, Ballas ZK, et al. Effect of a single ethanol exposure on HIV replication in human lymphocytes. *J Investig Med*. 1997;45:265–271.
10. Bagasra O, Kajdacsy-Balla A, Lischner HW. Effects of alcohol ingestion on in vitro susceptibility of peripheral blood mononuclear cells to infection with HIV and of selected T-cell functions. *Alcohol Clin Exp Res*. 1989;13:636–643.
11. Bagby GJ, Stoltz DA, Zhang P, et al. The effect of chronic binge ethanol consumption on the primary stage of SIV infection in rhesus macaques. *Alcohol Clin Exp Res*. 2003;27:495–502.
12. Poonia B, Nelson S, Bagby GJ, et al. Chronic alcohol consumption results in higher simian immunodeficiency virus replication in mucosally inoculated rhesus macaques. *AIDS Res Hum Retroviruses*. 2006;22:589–594.
13. Kumar R, Perez-Casanova AE, Tirado G, et al. Increased viral replication in simian immunodeficiency virus/simian-HIV–infected macaques with self-administering model of chronic alcohol consumption. *J Acquir Immune Defic Syndr*. 2005;39:386–390.
14. Stoltz DA, Nelson S, Kolls JK, et al. In vitro ethanol suppresses alveolar macrophage TNF-alpha during simian immunodeficiency virus infection. *Am J Respir Crit Care Med*. 2000;161:135–140.
15. Stoltz DA, Nelson S, Kolls JK, et al. Effects of in vitro ethanol on tumor necrosis factor-alpha production by blood obtained from simian immunodeficiency virus-infected rhesus macaques. *Alcohol Clin Exp Res*. 2002;26:527–534.
16. Bagby GJ, Zhang P, Purcell JE, et al. Chronic binge ethanol consumption accelerates progression of simian immunodeficiency virus disease. *Alcohol Clin Exp Res*. 2006;30:1781–1790.
17. Kaslow RA, Blackwelder WC, Ostrow DG, et al. No evidence for a role of alcohol or other psychoactive drugs in accelerating immunodeficiency in HIV-1-positive individuals. A report from the Multicenter AIDS Cohort Study. *JAMA*. 1989;261:3424–3429.
18. Samet JH, Horton NJ, Traphagen ET, et al. Alcohol consumption and HIV disease progression: are they related? *Alcohol Clin Exp Res*. 2003; 27:862–867.
19. Samet JH, Libman H, Labelle C, et al. A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. *Arch Intern Med*. 1995;155:1629–1633.
20. Folstein MF, Folstein SE, McHugh PR."Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189–198.
21. Smith KL, Horton NJ, Saitz R, et al. The use of the mini-mental state examination in recruitment for substance abuse research studies. *Drug Alcohol Depend*. 2006;82:231–237.
22. Pachl C, Todd JA, Kern DG, et al. Rapid and precise quantification of HIV-1 RNA in plasma using a branched DNA signal amplification assay. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1995;8:446–454.
23. Fairley CK, Spelman D, Street A, et al. CD4 lymphocyte numbers after splenectomy in patients infected with the human immunodeficiency virus. *Int J STD AIDS*. 1994;5:177–181.
24. Zurlo JJ, Wood L, Gaglione MM, et al. Effect of splenectomy on T lymphocyte subsets in patients infected with the human immunodeficiency virus. *Clin Infect Dis*. 1995;20:768–771.
25. Sobell L, Sobell M. *Alcohol Timeline Followback (TLFB) Users' Manual*. Toronto, Ontario, Canada: Addiction Research Foundation; 1995.
26. National Institute on Alcohol Abuse and Alcoholism. *The Physician's Guide to Helping Patients with Alcohol Problems*. Washington, DC: National Institutes of Health; 1995.
27. Kertesz SG, Horton NJ, Friedmann PD, et al. Slowing the revolving door: stabilization programs reduce homeless persons' substance use after detoxification. *J Subst Abuse Treat*. 2003;24:197–207.
28. 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 and Adherence Working Group of the Outcomes Committee of the Adult AIDS Clinical Trials Group (AACTG). *AIDS Care*. 2000;12:255–266.
29. Radloff L. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1:385–401.
30. Robins LN, Wing J, Wittchen HU, et al. The Composite International Diagnostic Interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Arch Gen Psychiatry.* 1988;45:1069–1077.
31. Crum RM, Galai N, Cohn S, et al. Alcohol use and T-lymphocyte subsets among injection drug users with HIV-1 infection: a prospective analysis. *Alcohol Clin Exp Res*. 1996;20:364–371.
32. Schuckit MA, Tipp JE, Smith TL, et al. Periods of abstinence following the onset of alcohol dependence in 1,853 men and women. *J Stud Alcohol*. 1997;58:581–589.
33. Kapasi AA, Patel G, Goenka A, et al. Ethanol promotes T cell apoptosis through the mitochondrial pathway. *Immunology.* 2003;108:313–320.
34. Chander G, Lau B, Moore RD. Hazardous alcohol use: a risk factor for nonadherence and lack of suppression in HIV infection. *J Acquir Immune Defic Syndr*. 2006;43:411–417.
35. Lefevre F, O'Leary B, Moran M, et al. Alcohol consumption among HIV-infected patients. *J Gen Intern Med*. 1995;10:458–460.
36. Krupitsky EM, Horton NJ, Williams EC, et al. Alcohol use and HIV risk behaviors among HIV-infected hospitalized patients in St. Petersburg, Russia. *Drug Alcohol Depend*. 2005;79:251–256.
37. World Health Organization. *Global Status Report on Alcohol 2004*. Geneva, Switzerland: World Health Organization; 2004.
38. Moyer A, Finney JW, Swearingen CE, et al. Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. *Addiction*. 2002;97:279–292.
39. Saitz R. Clinical practice. Unhealthy alcohol use. *N Engl J Med*. 2005; 352:596–607.
40. Aharonovich E, Hatzenbuehler ML, Johnston B, et al. A low-cost, sustainable intervention for drinking reduction in the HIV primary care setting. *AIDS Care*. 2006;18:561–568.

PANMINERVA MED 2007;49:67-77

Illicit drugs, alcohol, and addiction in human immunodeficiency virus

J. H. SAMET [1, 2], A. Y. WALLEY [1], C. BRIDDEN [1]

Drug and alcohol use complicate both the prevention and treatment of human immunodeficiency virus (HIV) infection. Substance use is one of the major engines driving HIV transmission, directly, through the sharing of injection drug use equipment and indirectly, through increasing risky sexual behaviors. Drug and alcohol dependence compromise effective HIV treatment by influencing both access and adherence to antiretroviral therapy. Exposure to addictive substances may have direct immunosuppressive effects independent of their impact on access and adherence to treatment. Measures effective at minimizing HIV transmission attributable to drug and alcohol use include HIV testing and referral to treatment, syringe and needle exchange programs, opioid replacement therapy (i.e., methadone and buprenorphine), and behavioral interventions targeting HIV risk behaviors among both HIV-infected and HIV-uninfected people. Measures effective at optimizing HIV treatment among alcohol and drug-dependent patients include HIV testing with referral to treatment and substance use treatment that is linked to or integrated into HIV treatment. Due to the intertwining problems of substance use and HIV infection, physicians and other health care providers must address the issues of illicit drugs and alcohol use as mainstream medical problems in order to provide optimal care for HIV-infected patients.

KEY WORDS: HIV infections - Substance-related disorders - Risk factors.

[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]*Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA*

Substance use and human immunodeficiency virus epidemiology: prevalence and transmission

Substance use impacts many dimensions of human immunodeficiency virus (HIV) infection, beginning with transmission of the virus as injection drug use is a major risk factor for HIV infection. Although injection drug use most commonly implies the use of heroin, other drugs such as cocaine and methamphetamines are parenterally administered. Globally, injection drug use has been identified as a mode of HIV transmission in over 80 countries.[1, 2] Use of contaminated injection drug equipment, especially needles and syringes, is estimated to account for approximately 1/3 of new infections outside sub-Saharan Africa. It is the major route of HIV transmission in eastern Europe and central Asia.[1, 3-5] Approximately 35% of the almost 1 million acquired immune deficiency syndrome (AIDS) cases reported in the United States through 2004 are attributable directly to injection drug use or to having a sex partner [6] who injects drugs. In the past two decades, the proportion of AIDS cases by transmission *via* male to male sexual contact has decreased in the United States, while transmission *via* injection drug use has remained between 20-30%.[7]

Portions of this work were previously presented at the Harvard Medical School, Beth Israel Deaconess Hospital, 11th Annual HIV Update: Contemporary Issues in Management, Boston, MA, USA, 2007.

Funding.—Dr. Samet is supported by the National Institute of Alcohol Abuse and Alcoholism (NIAAA), National Institutes on Health (NIH), K24-AA015674 (Impact of Alcohol Use on HIV Infection - In the United States and Russia).

Acknowledgements.—We would like to acknowledge the contributions of N. Mupier, MPH, particularly for her assistance with the references for this manuscript.

Address reprint requests to: J. H. Samet, MD, MA, MPH, Professor of Medicine and Public Health, 91 E. Concord Street, Ste. 200, Boston, MA 02118, USA. E-mail: jsamet@bu.edu

Injection drug use, however, is not the only substance use issue associated with HIV-infection. Alcohol-related problems are common among HIV-infected persons.[8] Numerous studies have documented an increased prevalence of current and past problems related to alcohol among HIV-infected persons. In the United States, the Veterans Aging Cohort Study (VACS) found that among 881 veterans with HIV-infection, 36% were current "hazardous drinkers".[9] In another study of 664 patients establishing primary care for HIV-infection, 42% screened positive for a history of alcohol problems.[9] In a study of 201 HIV-infected hospitalized Russian patients, only 4% (9/201) had no diagnosis of alcohol or drug abuse or dependence, while such diagnoses were present in 9% (19/201) for alcohol alone, 39% (78/201) for alcohol and drugs and 47% (95/201) for drugs alone.[10] In an epidemiological surveillance study in rural Uganda, HIV-infected persons were twice as likely to report a history of alcohol consumption as compared to HIV-uninfected persons.[11] An understanding of the implications of drug and alcohol use on an individual's subsequent clinical course is key to providing excellent HIV medical care.

Human immunodeficiency virus risk behaviors and substance use

Substance use can be a risk factor for HIV infection by a number of different pathways in addition to the sharing of injection equipment.[12] It is also possible that an altered mental state due to the use of alcohol or drugs can result in risky sexual behaviors, such as reduced condom use or increased unprotected anal intercourse. Among drug users, the use of stimulants, such as cocaine and methamphetamine, has been linked to increased sex and drug risk behaviors in several populations.[13-15] In a study by Booth *et al.*, of over 26 000 drug users, individuals who smoked crack were more likely to have multiple sex partners and exchange sex for drugs and money.[16] Methamphetamine use has been described among men who have sex with men and has been associated with unprotected anal intercourse [17] and increased HIV seroconversion.[18] Among heterosexual men, methamphetamine use has been associated with unprotected vaginal sex with multiple female partners, anal sex, and sex exchanged for money and drugs. Among women, methamphetamine use has been associated with both increased risky sex and drug behaviors.[14, 19] A study among HIV-uninfected injection drug users (IDUs) in St. Petersburg, Russia, demonstrated increased HIV seroconversion among IDUs who inject stimulants.[20] Injection drug use and crack cocaine use were associated with increased HIV prevalence in a study of 1 449 drug users in Brazil.[21] It is hypothesized that decreased experience of pain, enhanced sexual desire, or increased confidence that sometimes occur with the use of drugs may result in an increased number of partners, prolonged duration of sexual activity, and increased tissue damage, resulting in more bodily fluid contact (*i.e.*, blood or semen) and consequent HIV transmission.[12]

The relationship of alcohol and HIV risk behaviors is complex and may vary by gender and age of the individual, as well as by context. In several African countries, strong associations have been demonstrated between alcohol use, HIV sexual risk behaviors, and increased HIV prevalence.[22] In a population-based study in Botswana where the HIV prevalence is 24%, Weiser *et al.* found that heavy alcohol use was associated with 3 to 8 times the odds of having unprotected sex with a non-monogamous partner, having multiple sexual partners, and paying for or selling sex in exchange for money or other resources.[23] Similar associations have been found among sexually-transmitted infection clinic patients in South Africa.[24] In a population-based study of women in the Arusha region of Northern Tanzania, where HIV prevalence was 8.5%, the odds of HIV infection was more than 2 times higher among women who had engaged in sexual intercourse under the influence of alcohol compared to non-drinkers.[25] Among female drug users, increased alcohol consumption is associated with sexual HIV risk-taking behavior.[26] Based on current data, addressing alcohol problems among drug users, particularly women, may be important for reducing HIV sexual risk behaviors and possibly drug risk behaviors.

Alcohol and drugs are frequently used in combination with each other enhancing their disinhibiting effects and moderating their negative effects. Mixed results concerning the impact of alcohol use on risky drug behavior have been found.[26, 27] No significant association was found between alcohol consumption and risky injection drug behavior in a study among 354 drug users.[26] In a cohort of Russian IDUs, an association was found between cannabis use, but not heavy alcohol use, and increased risky injection behavior.[28] Stein *et al.* found that in a cohort of IDUs with an

alcohol abuse prevalence of 28% at-risk alcohol consumption and alcohol abuse were associated with needle sharing.[27]

There is a growing body of literature suggesting an association between the use of alcohol and illicit substances with HIV-risk behaviors among HIV-infected persons. In the HIV Cost and Services Utilization Study (HCSUS), which included a sample of 1 278 HIV-infected persons in medical care, marijuana, hard drugs, moderate alcohol use (*i.e.*, 6-29 drinks in 1 month), and previous drug dependence were related to being sexually active in the past 6 months.[29] Among a cohort of 4 016 HIV-infected men and women from 15 US cities stimulant use was associated with increased rates of unprotected vaginal or anal sex.[30] A study of HIV-infected men in treatment for methamphetamine use found that using methamphetamine increased the likelihood of more unprotected intercourse with more partners than non-infected men.[31] In a cohort of 345 HIV-infected individuals with a history of alcohol problems, 38% reported inconsistent condom use at baseline.[32] Among active IDUs from this cohort, at-risk drinking was also associated with inconsistent condom use. Thus substance use plays a fundamental role in the transmission of HIV infection in diverse populations worldwide by enhancing sexual risk and the sharing of injection equipment.

Substance use impact on human immunodeficiency virus disease clinical manifestations and progression

HIV-infected IDUs do have some distinctive clinical characteristics. The AIDS-related opportunistic infection, tuberculosis is more common and Kaposi's sarcoma is less common among IDUs.[33] In 1995-1996, the era just prior to highly active antiretroviral therapy (HAART), no differences in AIDS defining illnesses were found between 827 IDUs and 1 314 non-IDUs.[34] However, in the period of 2000-2002, AIDS defining illnesses among IDUs increased 50% compared with non-IDUs.[34] Since 1996, several studies have demonstrated a smaller mortality improvement in patients infected with HIV from injection drug use compared to other transmission factors.[35-37]

Comorbidities, such as chronic psychiatric disorders and hepatitis C, are more common among IDUs.[33]

It is well established that psychiatric conditions, such as depression, anxiety, and severe mental illness, are prevalent among substance using HIV-infected individuals.[38, 39] Both HIV-infected and uninfected women who inject drugs may be at higher risk for depressive symptoms.[40, 41] The relationship between mental health disorders, substance abuse, and HIV manifests in a number of ways, including increased high risk sex and drug practices, more rapid HIV disease progression, and decreasing use of HAART.[38, 40, 42-44]

The impact of substance use on HIV disease progression may be related to receipt of and adherence to antiretroviral therapy (ART) was well as direct consequences of the substances on immune function and other clinical outcomes. Investigators have examined the relationship of active substance use and HIV disease progression and mortality. In a longitudinal study of HIV-infected clinic patients, Lucas *et al.* found that subjects who persistently or intermittently used drugs had increased mortality compared to subjects who were non-drug users.[45] Of the subjects classified as intermittent drug users, those with recent active use had rates of HIV disease progression similar to those with persistent drug use, whereas those in a period of abstinence had HIV disease progression rates similar to those who never used drugs. In a 6-site longitudinal study of HIV-infected women, Kapadia *et al.* reported that women who consistently used non-injection drugs were at higher risk for progressing to AIDS than women who had never used non-injection drugs.[46] Among 595 HIV-infected patients with alcohol problems, Walley *et al.* found a 2.4 times increased hazard for 6-month mortality for patients with recent heroin or cocaine use after adjustment for age, CD4 cell count, ART use, homelessness, and alcohol use.[47]

Several reasons may explain why active substance users with HIV infection may have worse mortality and health outcomes. HIV-infected people with alcohol and drug use delay presentation to medical care after learning about an HIV diagnosis and start treatment later in their disease course.[45, 48, 49] In some countries substance users experience increased barriers to accessing health care, including poverty, lack of insurance, and stigmatization.[50-52] Adherence to ART may be more difficult for these patients because of the competing daily struggles to cope with an addiction or find housing.[53-61] Overdose and injury contribute to increased mortality among drug users.[62, 63]

The role of alcohol in HIV disease progression has been examined. Alcohol use may affect mortality through increasing susceptibility to injury, pneumonia, suicide, and liver disease.[64, 65] The direct toxic effects of alcohol and other drugs may lead to further immunosuppression and HIV disease progression.[46, 59, 61, 66] Alcohol use is associated with liver disease and may have a direct effect on the immune system, resulting in HIV disease progression.[61, 67, 68] In the era before HAART, there was no association found between alcohol and HIV progression.[69] In the HAART era, an association between heavy alcohol consumption and HIV disease progression was found.[70] Among those not receiving ART, heavy alcohol use was associated with a mean decrement in CD4 cell count of approximately 50 cells/mm^3 compared to non-drinkers. There was no impact of heavy alcohol use on HIV disease progression among those receiving ART. Thus, there is some evidence that heavy drinking in those not on ART may have a deleterious effect on HIV disease progression.

Physicians' attitudes may be a barrier to HIV care, particularly when patients are substance users. In the HCSUS Study, negative attitudes by physicians toward treating HIV infected IDUs were common (23%).[71] Factors associated with more positive physician attitudes were seeing more IDUs, having more HIV treatment knowledge, and treating fewer patients per week. Receipt of ART was related to these attitudes. Whereas 14% of IDUs with a physician with negative attitudes toward substance users receive ART, 32% of IDUs with physicians with positive attitudes received ART. This compares to ART receipt in 36% of patients of non-IDU patients who had a physician with negative attitudes.

Substance use and antiretroviral therapy adherence

Should antiretroviral therapy be initiated in substance users who are actively using drugs or alcohol? Several studies have demonstrated that injection drug use, cocaine use, and heavy drinking worsen ART adherence. A 12-month study among HIV-infected former and current heroin users found that cocaine use was the most important predictor of non-adherence.[53] Cocaine users were 27% adherent compared to 68% adherent for non-users. Consistent with these findings, 13% of cocaine users maintained viral suppression over a 6-month follow-up compared to 46% for non-users. Similarly, Hinkin *et al.* found that active stimulant use was associated with lower adherence.[72] Two other studies found decreased ART adherence among current drug users and alcohol users, but no decrease in adherence in patients with past drug use only.[55, 56] A cohort study of 578 HIV-infected persons first prescribed ART between 1996-2000 classified drug users as current IDUs, former IDUs, and non-drug users.[61] The current IDUs were less likely to suppress their HIV RNA to <500 copies/mL compared to non-drug users (adjusted odds ratio [AOR]: 0.3; 95% confidence interval [CI]: 0.13-0.67). Former IDUs were not less likely to achieve HIV suppression compared to non-drug users. Lucas *et al.* studied transitions between abstinence and substance use in substance-dependent HIV-infected persons.[57] Becoming abstinent was associated with better ART adherence, improved virologic control, and higher CD4 cell count; resuming use was associated with worse adherence, poorer virologic control, and decreased CD4 cell count.

Alcohol use has also been shown to negatively impact medication adherence. In a 1990's survey of 212 HIV-infected persons from two outpatient clinics, problem drinkers were significantly more likely to report taking medications off schedule (45% *versus* 26%; P=0.02).[73] Reasons for missing medications included forgetting, running out of medications, and consuming alcohol or drugs. More recently, alcohol use was found to be associated with decreased ART adherence and decreased use of ART.[54, 74, 75]

However, some studies have shown that some active drug users can successfully take medications.[76] Techniques, such as directly observed therapy, may provide the additional support active drugs users need to maintain adherence to their medication regimen.[77, 78] Among patients with a history of drug or alcohol dependence, but currently in recovery, ART should be considered in the same manner as a patient without a history of substance abuse.[79] While adherence may be compromised in patients with active alcohol or drug abuse or dependence, clinicians should weigh the risks and benefits of starting ART on a case by case basis in each individual patient.[80] Patient-specific factors, such as the patient's personal adherence history, availability of social and adherence support, motivation to engage in treatment, CD4 cell count, viral load, history of opportunistic infections, and the resistance profile of the patient's virus, should all be considered in making this decision.

Human immunodeficiency virus prevention, human immunodeficiency virus medical care, and substance abuse treatment

HIV seroprevalence among IDUs varies greatly depending on the local prevalence of drug use, the level of awareness of HIV risk in the community, and the availability of HIV prevention programs.

Needle exchange programs

One of the best documented prevention strategies for HIV infection among substance users is needle exchange programs. Comprehensive reviews of the scientific literature reveal substantial evidence that needle exchange programs decrease risky behavior and HIV seroconversion among IDUs and are cost-effective.[81, 82] In one evaluation, a randomized controlled trial compared needle exchange programs with improved access to syringes from pharmacies.[83] Needle exchange programs were not associated with an increase in injection drug use, the number of injections a day, or positive urine toxicology tests.

Globally HIV prevalence among IDUs is highly variable. The notably low 1% seroprevalence of HIV infection among IDUs in Australia is striking in comparison to the 25% seroprevalence in the United States and Europe and 30-80% in Asia, Russia and Eastern Europe.[2, 84] In Australia the institution of needle exchange programs occurred early in the evolution of the epidemic (mid-1980s) and methadone maintenance treatment for opioid dependence has been readily available.[85] The use of such needle exchange programs and pharmacotherapy for opioid dependence are key strategies that have been successfully utilized in parts of the world and not so effectively utilized in others.[86]

Effective human immunodeficiency virus risk reduction interventions

HIV prevention interventions have been demonstrated to be effective for both HIV-infected patients and high-risk HIV-negative patients (*e.g.*, Project RESPECT).[87, 88] Few interventions have been developed that demonstrate significant sexual and drug-related behavioral risk reduction.[89-92] Though the original Project RESPECT did not specifically address substance using or HIV-infected populations, more recent utilization of the program, has been conducted with these populations and are being assessed.[93, 94] In South Africa, an HIV and alcohol risk reduction behavioral skills intervention among alcohol using sexually transmitted infection clinic patients showed a 65% reduction in unprotected sex and a 25% increase in condom use at 6 month follow-up compared to a control condition.[22]

Human immunodeficiency virus testing among substance users

Does HIV testing have any impact on one's substance use? In a survey of patients undergoing HIV testing at 5 public alcohol treatment centers, which also addressed HIV risk reduction, 26% reported a reduction in having sex with an injection drug user, 58% reported a reduction in the use of injection drugs, and 77% reported improvement in consistent condom use with multiple sexual partners at 13 months follow-up.[95] Thus, there was a clear association among HIV testing and subsequent diminished HIV risky behavior. In the HCSUS Study of 2 864 HIV-infected persons, 80% of substance users quit or cut-down after their HIV diagnosis.[96] Persons with nadir CD4 counts of <50 were also more likely to quit or reduce their substance use. In the past year, the Centers for Disease Control and Prevention (CDC) has made major revisions in recommendations for HIV testing in the US, promoting opt-out screening for patients in all healthcare settings, with high risk persons being screened at least annually.[97] High-risk individuals include IDUs, particularly those who share needles, and substance abuse users who use stimulants. The World Health Organization guidelines recommend opt-out screening when testing is offered, routine testing for groups with clear indications (*e.g.*, presenting with HIV-related symptoms), and universal testing in generalized epidemics only.[98] Data to support the notion that those with risky alcohol use should be considered "high risk" with regards to these recommendations are less clear.

Linking human immunodeficiency virus-infected patients to medical care

Increasingly, it has been recognized that effective therapy and prevention for HIV-infected persons is only possible if these individuals engage in medical care.[99] HIV testing is essential to engaging HIV-infected persons in medical care, but not sufficient.[100]

Linking newly-diagnosed HIV-infected persons with substance use to care can be accomplished with post-test counseling and referral. In the 1990's, it was demonstrated that presentation for initial HIV care after a positive test was commonly >1 year.[49] A history of injection drug use was associated with this delay (P<0.001); men with a history of alcohol problems delayed 14.6 months longer than men who did not have such an alcohol history.[101]

Linkage of HIV-infected patients with substance use problems to medical and substance use treatment can help coordinate care between these important clinical services. Models of care linkage and coordination range from centralized, comprehensive one-stop sites where patients receive all needed care and services to distributive models where referrals to needed care and services are facilitated with case managers.[102] On-site substance abuse treatment in primary care has been shown to be safe and effective.[103-107] Integrated care for alcohol and drug dependence coordinated with medical care improves addiction treatment outcomes and increases the possibility of improved health care utilization.[102, 107-111]

Case management/multidisciplinary care for human immunodeficiency virus -infected substance users

Case management is a coordinated approach to services delivery. Patient assessment, care planning, linkage to services, outcome monitoring, and advocacy for patients are addressed in a single point of contact, for substance abuse treatment, medical services and a variety of social services (*e.g.* housing, self-help groups, and employment). Case management decreases relapse and increases retention in addiction treatment.[112, 113] Twice weekly outpatient group counseling was compared with counseling plus case management in a *quasi*-experimental study of patients admitted to 1 of 12 addiction treatment programs. At 4 weeks, case management was associated with greater receipt of addiction, medical, psychiatric, employment, and family services, and at 6 months, with less intoxication and fewer days of psychiatric and medical problems.[112] Case management has also been found to be associated with improved ART adherence and CD4 cell counts.[114]

Substance abuse treatment

An efficient way to reduce HIV transmission *via* substance use-related risky behaviors is to provide effective substance abuse treatment. Effective behavioral and pharmacological treatments exist and often combinations of these approaches may be optimal.

Screening, counseling and behavioral treatments

Physicians can play an important role in the early detection and intervention of substance abuse among patients with HIV infection.[115] For those identified with substance use problems, clinicians can assess and address a patient's readiness to change alcohol and drug use behaviors with the use of brief interventions and appropriate referral to treatment. The structure of these interventions include the elements included in the mnemonic FRAMES: feedback about substance use (*e.g.*, quantity relative to norms); responsibility to stop use of substances is the patient's; argumentation is to be avoided; menu of options to address recovery should be offered; empathy should be expressed; and self-efficacy to change one's behavior needs to be supported.[116] Taken together, this approach involves giving clear advice about the need to change behavior, providing the patient options to help facilitate that change, having this discussion in a manner in which empathy concerning the patient's condition is expressed while at the same time supporting the patient's self efficacy concerning their perception and their ability to achieve change.

Integral to substance abuse treatment with pharmacotherapy or without it, is substance abuse counseling or talk therapy. In the United States, methadone maintenance clinics are required to provide counseling, and outpatient buprenorphine prescribers generally provide or refer patients to counseling.

Twelve step programs are another component of substance abuse treatment that can be highly effective. These programs include alcoholics anonymous (AA), narcotic anonymous (NA), and cocaine anonymous. Such programs focus on abstinence and emphasize lifelong participation. However, abstinence is not required for participation, but rather a desire to quit drug and alcohol use. Through going to regular meetings, attendees build a support network and an abstinence mentor, called a sponsor. Some communities offer day treatment programs that include AA or NA group meetings, as well as individualized counseling and case management. Halfway houses or sober houses, where people in treatment and recovery live together, are particularly valuable resources for patients who are homeless or whose use is triggered by where they

live. Residential or therapeutic communities are programs where patients stay for weeks to months and provide more structure, usually restricting patients to the community grounds.[117]

Focusing on contingencies is an increasingly important part of substance abuse treatment. Drug courts offer substance users suspended criminal penalties, contingent upon entry and completion of substance abuse treatment.[118] Contingency management, a treatment strategy that rewards alcohol and drug users for decreased use or abstinence has been demonstrated to be efficacious in many settings.[119] The practice in methadone maintenance of giving "take home" doses of methadone as a reward for urine toxicology results free of illicit substances is an example of a contingency management technique.

Pharmacotherapy

The growing pharmacotherapy for addressing substance use includes methadone, buprenorphine, and naltrexone for opioid dependence and naltrexone, acamprosate, and disulfram for alcohol dependence. Because naltrexone, acamprosate, and disulfram do not have the re-enforcing characteristics that buprenorphine and methadone have, HIV-infected and comparably addicted non-HIV-infected patients less commonly use these medications. For cocaine, metamphetamine and other psychostimulants, there are, as yet, no effective medications.

Both methadone and buprenorphine are opioid replacement therapies that are highly effective for opioid dependence by treating withdrawal symptoms, blocking the euphoric effects of other opioids, and by reducing craving for opioids. Methadone has been used for opioid dependence in the United States for 4 decades. Its efficacy is unequivocal. It improves overall survival, improves retention in treatment, decreases heroin and other drug use, decreases HIV drug risk behaviors,[120] decreases hepatitis and HIV seroconversion, decreases criminal activity, increases social functioning, improves birth outcomes, and decrease obstetric complications.[121] Among HIV-infected patients on ART in the Vancouver IDU Study (VIDUS) methadone maintenance treatment was associated with better adherence (AOR: 1.52; 95% CI: 1.16-2), HIV ribonucleic acid (RNA) suppression (AOR: 1.34; 95% CI: 1-1.79), and CD4 cell count rise (AOR: 1.58; 95% CI: 1.26-1.99).[122] Structured methadone programs with daily dosing have the potential to offer directly observed antiretroviral therapy, which has been associated with improved viral suppression rates in one observational study.[123]

Buprenorphine is a partial opioid agonist that has been available in France for over a decade and more recently became available (2002) through outpatient primary care physicians and psychiatrists with limited additional training in the United States. It is most commonly prescribed co-formulated with naloxone and administered sublingually.[124] The purpose of the naloxone component is to discourage intravenous use. Buprenorphine's efficacy is comparable to methadone.[125] Patients who stop buprenorphine endure milder withdrawal symptoms than with methadone and encounter a lower risk of overdose than other opioids due to the agonist ceiling effect. A 12-week pilot study of buprenorphine in 16 HIV-infected patients demonstrated a decrease in opioid-positive urine tests and a decrease in HIV viral load.[126]

Interactions between antiretroviral therapy and opioid agonist therapy

Providers of patients taking ART and methadone or buprenorphine should consider the known drug-drug interactions and be sure to engage in regular communication if more than one provider prescribes these medications.[127, 128] In general, since methadone preceded the AIDS epidemic, its interactions with ART have been more thoroughly studied compared to buprenorphine. Within the nucleoside reverse transcriptase class, zidovudine (AZT) concentrations are increased in patients taking methadone. Though uncommonly reported, clinicians should be aware of methadone patients' potential for AZT-related toxicities, such as cytopenias. Because methadone decreases the concentration of didanosine (DDI) in patients taking the non-enteric coated tablets, patients on methadone and DDI should be prescribed the enteric-coated tablets only. Within the non-nucleoside transcriptase inhibitor class, both efavirenz and nevirapine reduce methadone concentrations and can induce opioid withdrawal in methadone-maintained patients. This withdrawal may be prevented by increasing the methadone dose based on patient symptoms at the time of initiating efavirenz or nevirapine. Within the protease inhibitor class, atazanavir boosted with ritonavir has been associated with opioid excess in one case report of buprenorphine patients, thus patients should be monitored for a methadone dose reduction

when starting atazanavir and ritonavir. Loprinavir/ritonavir, nelfinavir, and tipranavir are each associated with decreased methadone levels, so patients starting these medications should be monitored for withdrawal and have their methadone increased as needed. Methadone does decrease an active metabolite of nelfinavir, though no clinically significant effects have been reported. Therapeutic drug monitoring of nelfinavir in patients receiving both may be useful particularly in the setting of virologic failure in adherent patients.[129]

Chronic pain/pain management

Pain management is another important dimension that at times is particularly complicated in patients with substance use. Patients with opioid dependence may be undertreated for pain due to their dependency and/or their methadone maintenance treatment.[130, 131] Providers may consider methadone a treatment in and of itself for pain. However, if a patient is in methadone treatment, then pain should be treated by maintaining methadone and using shorter acting narcotics given at higher doses. Treatment should be given as needed preferably on a fixed schedule to relieve the pain. Therapeutic use of marijuana has been examined in the Positive Health Study, a longitudinal cohort of HIV-infected persons. Of 244 individuals using marijuana, 44% reported both recreational and therapeutic use.[132] In a randomized placebo controlled trial of 223 adults with HIV and symptomatic HIV sensory neuropathy, smoking cannabis reduced daily pain by 34% compared to 17% in the placebo group (P=0.03).[133] More subjects in the cannabis group experienced a reduction >30% as compared to placebo (52% *vs* 24%; P=0.04). There were no serious adverse events reported.

Conclusions

In summary, drug and alcohol use are commonly encountered in subjects with or at risk of HIV infection. Therefore, it is important to address drug and alcohol use in the context of prevention and treatment of HIV. Targeting and tailoring HIV prevention efforts to drug and alcohol users can effectively reduce risky sexual and drug behaviors that can lead to HIV transmission. Providing behavioral and pharmacologic treatment for substance abuse can contribute to reduction in HIV risk behaviors. Optimal HIV treatment with access to ART and other medical services is crucial in the management of the disease. A coordinated effort to integrate substance use treatment, HIV treatment, and prevention activities is necessary to effectively address patients' intertwining issues of substance use and HIV infection.

References

1. UNAIDS. Report on the global AIDS epidemic 2006. 2006. Available at: http://www.unaids.org/en/HIV_data/2006GlobalReport/default.asp. Accessed June 26, 2007.
2. Aceijas C, Stimson GV, Hickman M, Rhodes T. Global overview of injecting drug use and HIV infection among injecting drug users. AIDS 2004;18:2295-303.
3. UNAIDS. At risk and neglected: four key populations. 2006. Available at: http://data.unaids.org/pub/GlobalReport/2006/2006_GR_CH05_en.pdf. Accessed June 26, 2007.
4. Dehne KL, Pokrovskiy V, Kobyshcha Y, Schwartlander B. Update on the epidemics of HIV and other sexually transmitted infections in the newly independent states of the former Soviet Union. AIDS 2000;14 Suppl 3:S75-S84.
5. Hamers FF, Downs AM. HIV in central and eastern Europe. Lancet 2003;361:1035-44.
6. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2004. Atlanta, US Department of Health and Human Services, Centers for Disease Control and Prevention. 2005. Available at: http://www.cdc.gov/hiv/topics/surveillance/resources/reports/2004report/pdf/2004SurveillanceReport.pdf. Accessed June 26, 2007.
7. Centers for Disease Control and Prevention. AIDS Surveillance - General Epidemiology (through 2005). Atlanta, US Department of Health and Human Services, Centers for Disease Control and Prevention. 2007. Available at: http://www.cdc.gov/hiv/topics/surveillance/resources/slides/epidemiology/index.htm. Accessed June 26, 2007.
8. Lefevre F, O'Leary B, Moran M, Mossar M, Yarnold PR, Martin GJ *et al.* Alcohol consumption among HIV-infected patients. J Gen Intern Med 1995;10:458-60.
9. Conigliaro J, Gordon AJ, McGinnis KA, Rabeneck L, Justice AC. How harmful is hazardous alcohol use and abuse in HIV infection: do health care providers know who is at risk? J Acquir Immune Defic Syndr 2003;33:521-5.
10. Krupitsky EM, Horton NJ, Williams EC, Lioznov D, Kuznetsova M, Zvartau E *et al.* Alcohol use and HIV risk behaviors among HIV-infected hospitalized patients in St. Petersburg, Russia. Drug Alcohol Depend 2005;79:251-6.
11. Mbulaiteye SM, Ruberantwari A, Nakiyingi JS, Carpenter LM, Kamali A, Whitworth JA. Alcohol and HIV: a study among sexually active adults in rural southwest Uganda. Int J Epidemiol 2000;29:911-5.
12. Drumright LN, Patterson TL, Strathdee SA. Club drugs as causal risk factors for HIV acquisition among men who have sex with men: a review. Subst Use Misuse 2006;41:1551-601.
13. Buchanan D, Tooze JA, Shaw S, Kinzly M, Heimer R, Singer M. Demographic, HIV risk behavior, and health status characteristics of "crack" cocaine injectors compared to other injection drug users in three New England cities. Drug Alcohol Depend 2006;81:221-9.
14. Lorvick J, Martinez A, Gee L, Kral AH. Sexual and injection risk among women who inject methamphetamine in San Francisco. J Urban Health 2006;83:497-505.
15. Molitor F, Ruiz JD, Flynn N, Mikanda JN, Sun RK, Anderson R. Methamphetamine use and sexual and injection risk behaviors among out-of-treatment injection drug users. Am J Drug Alcohol Abuse 1999;25:475-93.
16. Booth RE, Kwiatkowski CF, Chitwood DD. Sex related HIV risk behav-

iors: differential risks among injection drug users, crack smokers, and injection drug users who smoke crack. Drug Alcohol Depend 2000;58:219-26.
17. Mansergh G, Purcell DW, Stall R, McFarlane M, Semaan S, Valentine J, Valdiserri R. CDC consultation on methamphetamine use and sexual risk behavior for HIV/STD infection: summary and suggestions. Public Health Rep 2006;121:127-32.
18. Plankey MW, Ostrow DG, Stall R, Cox C, Li X, Peck JA *et al.* The relationship between methamphetamine and popper use and risk of HIV seroconversion in the multicenter AIDS cohort study. J Acquir Immune Defic Syndr 2007;45:85-92.
19. Centers for Disease Control and Prevention (CDC). Methamphetamine use and HIV risk behaviors among heterosexual men: preliminary results from five northern California counties, December 2001-November 2003. MMWR Morb Mortal Wkly Rep 2006;55:273-7.
20. Kozlov AP, Shaboltas AV, Toussova OV, Verevochkin SV, Masse BR, Perdue T *et al.* HIV incidence and factors associated with HIV acquisition among injection drug users in St Petersburg, Russia. AIDS 2006;20:901-6.
21. Pechansky F, Woody G, Inciardi J, Surratt H, Kessler F, Von Diemen L *et al.* HIV seroprevalence among drug users: an analysis of selected variables based on 10 years of data collection in Porto Alegre, Brazil. Drug Alcohol Depend 2006;82 Suppl 1:S109-13.
22. Kalichman SC, Simbayi LC, Vermaak R, Cain D, Jooste S, Peltzer K. HIV/AIDS risk reduction counseling for alcohol using sexually transmitted infections clinic patients in Cape Town, South Africa. J Acquir Immune Defic Syndr 2007;44:594-600.
23. Weiser SD, Leiter K, Heisler M, McFarland W, Percy-de Korte F, DeMonner SM *et al.* A population-based study on alcohol and high-risk sexual behaviors in Botswana. PLoS Med 2006;3:e392.
24. Simbayi LC, Kalichman SC, Jooste S, Mathiti V, Cain D, Cherry C. Alcohol use and sexual risks for HIV infection among men and women receiving sexually transmitted infection clinic services in Cape Town, South Africa. J Stud Alcohol 2004;65:434-42.
25. Mnyika KS, Klepp KI, Kvale G, Ole-King'ori N. Risk factors for HIV-1 infection among women in the Arusha region of Tanzania. J Acquir Immune Defic Syndr Hum Retrovirol 1996;11:484-91.
26. Rees V, Saitz R, Horton NJ, Samet J. Association of alcohol consumption with HIV sex- and drug-risk behaviors among drug users. J Subst Abuse Treat 2001;21:129-34.
27. Stein MD, Hanna L, Natarajan R, Clarke J, Marisi M, Sobota M *et al.* Alcohol use patterns predict high-risk HIV behaviors among active injection drug users. J Subst Abuse Treat 2000;18:359-63.
28. Walley AY, Krupitsky EM, Cheng DM, Raj A, Edwards EM, Bridden C *et al.* Implications of cannabis use and heavy alcohol use on HIV drug risk behaviors in Russian heroin users. AIDS Behav 2007; [Epub ahead of print].
29. Beckett M, Burnam A, Collins RL, Kanouse DE, Beckman R. Substance use and high-risk sex among people with HIV: a comparison across exposure groups. AIDS Behav 2003;7:209-19.
30. Morin SF, Myers JJ, Shade SB, Koester K, Maiorana A, Rose CD. Predicting HIV transmission risk among HIV-infected patients seen in clinical settings. AIDS Behav 2007; [Epub ahead of print].
31. Shoptaw S, Reback CJ, Freese TE. Patient characteristics, HIV serostatus, and risk behaviors among gay and bisexual males seeking treatment for methamphetamine abuse and dependence in Los Angeles. J Addict Dis 2002;21:91-105.
32. Ehrenstein V, Horton NJ, Samet JH. Inconsistent condom use among HIV-infected patients with alcohol problems. Drug Alcohol Depend 2004;73:159-66.
33. Cohn JA. HIV-1 infection in injection drug users. Infect Dis Clin North Am 2002;16:745-70.
34. Moore RD, Keruly JC, Chaisson RE. Differences in HIV disease progression by injecting drug use in HIV-infected persons in care. J Acquir Immune Defic Syndr 2004;35:46-51.
35. Keiser O, Taffe P, Zwahlen M, Battegay M, Bernasconi E, Weber R *et al.*; And the Swiss HIV Cohort Study. All cause mortality in the Swiss HIV Cohort Study from 1990 to 2001 in comparison with the Swiss population. AIDS 2004;18:1835-43.
36. Perez-Hoyos S, del Amo J, Muga R, del Romero J, Garcia de Olalla P, Guerrero R *et al.*; GEMES (Spanish Multicenter Study Group of Seroconverters). Effectiveness of highly active antiretroviral therapy in Spanish cohorts of HIV seroconverters: differences by transmission category. AIDS 2003;17:353-9.
37. Poundstone KE, Chaisson RE, Moore RD. Differences in HIV disease progression by injection drug use and by sex in the era of highly active antiretroviral therapy. AIDS 2001;15:1115-23.
38. Chander G, Himelhoch S, Moore RD. Substance abuse and psychiatric disorders in HIV-positive patients: epidemiology and impact on antiretroviral therapy. Drugs 2006;66:769-89.
39. Marimoutou C, Carrieri P, Poizot-Martin I, Loundou A, Tremolieres F, Rey D *et al.*; MINIF 2000 Study Group. Hospitalization for depressive syndrome in a cohort of HIV-infected patients contaminated through injecting drug use: MANIF 2000 cohort, France, 1995-1999. AIDS Care 2003;15:729-34.
40. Johnson ME, Yep MJ, Brems C, Theno SA, Fisher DG. Relationship among gender, depression, and needle sharing in a sample of injection drug users. Psychol Addict Behav 2002;16:338-41.
41. Wisniewski AB, Apel S, Selnes OA, Nath A, McArthur JC, Dobs AS. Depressive symptoms, quality of life, and neuropsychological performance in HIV/AIDS: the impact of gender and injection drug use. J Neurovirol 2005;11:138-43.
42. Disney E, Kidorf M, Kolodner K, King V, Peirce J, Beilenson P *et al.* Psychiatric comorbidity is associated with drug use and HIV risk in syringe exchange participants. J Nerv Ment Dis 2006;194:577-83.
43. Ickovics JR, Hamburger ME, Vlahov D, Schoenbaum EE, Schuman P, Boland RJ *et al.*; HIV Epidemiology Research Study Group. Mortality, CD4 cell count decline, and depressive symptoms among HIV-seropositive women: longitudinal analysis from the HIV Epidemiology Research Study. JAMA 2001;285:1466-74.
44. Lundgren LM, Amodeo M, Chassler D. Mental health status, drug treatment use, and needle sharing among injection drug users. AIDS Educ Prev 2005;17:525-39.
45. Lucas GM, Griswold M, Gebo KA, Keruly J, Chaisson RE, Moore RD. Illicit drug use and HIV-1 disease progression: a longitudinal study in the era of highly active antiretroviral therapy. Am J Epidemiol 2006;163:412-20.
46. Kapadia F, Cook JA, Cohen MH, Sohler N, Kovacs A, Greenblatt RM *et al.* The relationship between non-injection drug use behaviors on progression to AIDS and death in a cohort of HIV seropositive women in the era of highly active antiretroviral therapy use. Addiction 2005;100:990-1002.
47. Walley A, Cheng DM, Libman H. Recent drug use and homelessness are associated with increased short-term mortality in HIV-infected persons with alcohol problems. J Gen Intern Med 2007;22:158.
48. Celentano DD, Galai N, Sethi AK, Shah NG, Strathdee SA, Vlahov D *et al.* Time to initiating highly active antiretroviral therapy among HIV-infected injection drug users. AIDS 2001;15:1707-15.
49. Samet JH, Freedberg KA, Stein MD, Lewis R, Savetsky J, Sullivan L *et al.* Trillion virion delay: time from testing positive for HIV to presentation for primary care. Arch Intern Med 1998;158:734-40.
50. Clarke S, Delamere S, McCullough L, Hopkins S, Bergin C, Mulcahy F. Assessing limiting factors to the acceptance of antiretroviral therapy in a large cohort of injecting drug users. HIV Med 2003;4:33-7.
51. Hsu LC, Vittinghoff E, Katz MH, Schwarcz SK. Predictors of use of highly active antiretroviral therapy (HAART) among persons with AIDS in San Francisco, 1996-1999. J Acquir Immune Defic Syndr 2001;28:345-50.
52. O'Connor PG, Selwyn PA, Schottenfeld RS. Medical care for injection-drug users with human immunodeficiency virus infection. N Engl J Med 1994;331:450-9.
53. Arnsten JH, Demas PA, Grant RW, Gourevitch MN, Farzadegan H, Howard AA *et al.* Impact of active drug use on antiretroviral therapy adherence and viral suppression in HIV-infected drug users. J Gen Intern Med 2002;17:377-81.
54. Chander G, Lau B, Moore RD. Hazardous alcohol use: a risk factor for non-adherence and lack of suppression in HIV infection. J Acquir Immune Defic Syndr 2006;43:411-7.

55. Golin CE, Liu H, Hays RD, Miller LG, Beck CK, Ickovics J *et al.* A prospective study of predictors of adherence to combination antiretroviral medication. J Gen Intern Med 2002;17:756-65.
56. Howard AA, Arnsten JH, Lo Y, Vlahov D, Rich JD, Schuman P *et al.*; HER Study Group. A prospective study of adherence and viral load in a large multi-center cohort of HIV-infected women. AIDS 2002;16:2175-82.
57. 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;16:767-74.
58. Moss AR, Hahn JA, Perry S, Charlebois ED, Guzman D, Clark RA *et al.* Adherence to highly active antiretroviral therapy in the homeless population in San Francisco: a prospective study. Clin Infect Dis 2004;39:1190-8.
59. Palepu A, Tyndall MW, Li K, Yip B, O'Shaughnessy MV, Schechter MT *et al.* Alcohol use and incarceration adversely affect HIV-1 RNA suppression among injection drug users starting antiretroviral therapy. J Urban Health 2003;80:667-75.
60. Palepu A, Tyndall M, Yip B, O'Shaughnessy MV, Hogg RS, Montaner JS. Impaired virologic response to highly active antiretroviral therapy associated with ongoing injection drug use. J Acquir Immune Defic Syndr 2003;32:522-6.
61. Samet JH, Horton NJ, Traphagen ET, Lyon SM, Freedberg KA. Alcohol consumption and HIV disease progression: are they related? Alcohol Clin Exp Res 2003;27:862-7.
62. Joe GW, Simpson DD. Mortality rates among opioid addicts in a longitudinal study. Am J Public Health 1987;77:347-8.
63. Wang C, Vlahov D, Galai N, Cole SR, Bareta J, Pollini R *et al.* The effect of HIV infection on overdose mortality. AIDS 2005;19:935-42.
64. Hwang SW. Mortality among men using homeless shelters in Toronto, Ontario. JAMA 2000;283:2152-7.
65. Rees VW, Horton NJ, Hingson RW, Saitz R, Samet JH. Injury among detoxification patients: alcohol users' greater risk. Alcohol Clin Exp Res 2002;26:212-7.
66. Mientjes GH, Miedema F, van Ameijden EJ, van den Hoek AA, Schellekens PT, Roos MT *et al.* Frequent injecting impairs lymphocyte reactivity in HIV-positive and HIV-negative drug users. AIDS 1991;5:35-41.
67. Bagby GJ, Zhang P, Purcell JE, Didier PJ, Nelson S. Chronic binge ethanol consumption accelerates progression of simian immunodeficiency virus disease. Alcohol Clin Exp Res 2006;30:1781-90.
68. Pavia CS, La MM, Kavanagh M. Influence of alcohol on antimicrobial immunity. Biomed Pharmacother 2004;58:84-9.
69. Dingle GA, Oei TP. Is alcohol a cofactor of HIV and AIDS? Evidence from immunological and behavioral studies. Psychol Bull 1997;122:56-71.
70. Samet J, Cheng D, Libman H, Nunes D, Alperen J, Saitz R. Alcohol consumption and HIV disease progression. J Acquir Immune Defic Syndr. In press 2007.
71. Ding L, Landon BE, Wilson IB, Wong MD, Shapiro MF, Cleary PD. Predictors and consequences of negative physician attitudes toward HIV-infected injection drug users. Arch Intern Med 2005;165:618-23.
72. Hinkin CH, Barclay TR, Castellon SA, Levine AJ, Durvasula RS, Marion SD *et al.* Drug use and medication adherence among HIV-1 infected individuals. AIDS Behav 2007;11:185-94.
73. 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;16:83-8.
74. 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;28:572-7.
75. Berg KM, Demas PA, Howard AA, Schoenbaum EE, Gourevitch MN, Arnsten JH. Gender differences in factors associated with adherence to antiretroviral therapy. J Gen Intern Med 2004;19:1111-7.
76. Marco A, Cayla JA, Serra M, Pedro R, Sanrama C, Guerrero R *et al.* Predictors of adherence to tuberculosis treatment in a supervised therapy programme for prisoners before and after release. Study Group of Adherence to Tuberculosis Treatment of Prisoners. Eur Respir J 1998;12:967-71.
77. Altice FL, Mezger JA, Hodges J, Bruce RD, Marinovich A, Walton M, Springer SA, Friedland GH. Developing a directly administered antiretroviral therapy intervention for HIV-infected drug users: implications for program replication. Clin Infect Dis 2004;38 Suppl 5:S376-87.
78. McCance-Katz EF, Gourevitch MN, Arnsten J, Sarlo J, Rainey P, Jatlow P. Modified directly observed therapy (MDOT) for injection drug users with HIV disease. Am J Addict 2002;11:271-8.
79. Sherer R. Adherence and antiretroviral therapy in injection drug users. JAMA 1998;280:567-8.
80. Ware NC, Wyatt MA, Tugenberg T. Adherence, stereotyping and unequal HIV treatment for active users of illegal drugs. Soc Sci Med 2005;61:565-76.
81. Gibson DR, Flynn NM, Perales D. Effectiveness of syringe exchange programs in reducing HIV risk behavior and HIV seroconversion among injecting drug users. AIDS 2001;15:1329-41.
82. Wodak A, Cooney A. Do needle syringe programs reduce HIV infection among injecting drug users: a comprehensive review of the international evidence. Subst Use Misuse 2006;41:777-813.
83. Fisher DG, Fenaughty AM, Cagle HH, Wells RS. Needle exchange and injection drug use frequency: a randomized clinical trial. J Acquir Immune Defic Syndr 2003;33:199-205.
84. Burak J, Bansberg D. Epidemiology and HIV transmission in injection drug users. HIV InSite; 1998.: http://hivinsite.ucsf.edu/InSite?page=kb-07&doc=kb-07-04-01. Accessed June 26, 2007.
85. Wodak A. Harm reduction: Australia as a case study. Bull N Y Acad Med. 1995;72:339-47.
86. Des Jarlais DC, Hagan H, Friedman SR, Friedmann P, Goldberg D, Frischer M *et al.* Maintaining low HIV seroprevalence in populations of injecting drug users. JAMA 1995;274:1226-31.
87. Kamb ML, Fishbein M, Douglas JM Jr, Rhodes F, Rogers J, Bolan G *et al.* Efficacy of risk-reduction counseling to prevent human immunodeficiency virus and sexually transmitted diseases: a randomized controlled trial. Project RESPECT Study Group. JAMA 1998;280:1161-7.
88. Merson MH, Dayton JM, O'Reilly K. Effectiveness of HIV prevention interventions in developing countries. AIDS 2000;14 Suppl 2:S68-S84.
89. Margolin A, Avants SK, Warburton LA, Hawkins KA, Shi J. A randomized clinical trial of a manual-guided risk reduction intervention for HIV-positive injection drug users. Health Psychol 2003;22:223-8.
90. Rotheram-Borus MJ, Lee MB, Murphy DA, Futterman D, Duan N, Birnbaum JM *et al.*; Teens Linked to Care Consortium. Efficacy of a preventive intervention for youths living with HIV. Am J Public Health 2001;91:400-5.
91. Wingood GM, DiClemente RJ, Mikhail I, Lang DL, McCree DH, Davies SL, Hardin JW, Hook EW 3rd, Saag M. A randomized controlled trial to reduce HIV transmission risk behaviors and sexually transmitted diseases among women living with HIV: The WiLLOW Program. J Acquir Immune Defic Syndr 2004;37 Suppl 2:S58-67.
92. Kalichman SC, Rompa D, Cage M, DiFonzo K, Simpson D, Austin J *et al.* Effectiveness of an intervention to reduce HIV transmission risks in HIV-positive people. Am J Prev Med 2001;21:84-92.
93. Krupitsky EM, Cheng DM, Raj A. Reduction of risky sexual behavior among hospitalized Russian substance-dependent patients: the Russian partnership to reduce the epidemic *via* engagement in narcology treatment (RUSSIAN PREVENT). Alcohol Clin Exp Res 2006;0:78A.
94. Rietmeijer CA. Risk reduction counselling for prevention of sexually transmitted infections: how it works and how to make it work. Sex Transm Infect 2007;83:2-9.
95. Avins AL, Lindan CP, Woods WJ, Hudes ES, Boscarino JA, Kay J *et al.* Changes in HIV-related behaviors among heterosexual alcoholics following addiction treatment. Drug Alcohol Depend 1997;44:47-55.
96. Collins RL, Kanouse DE, Gifford AL, Senterfitt JW, Schuster MA, McCaffrey DF *et al.* Changes in health-promoting behavior following diagnosis with HIV: prevalence and correlates in a national probability sample. Health Psychol 2001;20:351-60.
97. Centers for Disease Control and Prevention. The Global HIV/AIDS Pandemic, 2006. 8-11-2006. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5531a1.htm. Accessed June 26, 2007.

98. World Health Organization. Guidance on provider-initiated HIV testing and counselling in health facilities. 2007. Available at: http://www.who.int/hiv/topics/vct/PITCguidelines.pdf. Accessed June 26, 2007.
99. Samet JH. Engaging the HIV-infected substance abuse—the classic academic approach to changing clinical practice: excruciatingly slow, extraneous, and/or effective? MedGenMed 2001;3:16.
100. Samet JH, Freedberg KA, Savetsky JB, Sullivan LM, Stein MD. Understanding delay to medical care for HIV infection: the long-term non-presenter. AIDS 2001;15:77-85.
101. Turner BJ, Cunningham WE, Duan N, Andersen RM, Shapiro MF, Bozzette SA, Nakazono T, Morton S, Crystal S, St Clair P, Stein M, Zierler S. Delayed medical care after diagnosis in a US national probability sample of persons infected with human immunodeficiency virus. Arch Intern Med 2000;160:2614-22.
102. 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.
103. Friedmann PD, Alexander JA, Jin L, D'Aunno TA. On-site primary care and mental health services in outpatient drug abuse treatment units. J Behav Health Serv Res 1999;26:80-94.
104. Friedmann PD, D'Aunno TA, Jin L, Alexander JA. Medical and psychosocial services in drug abuse treatment: do stronger linkages promote client utilization? Health Serv Res 2000;35:443-65.
105. Friedmann PD, Lemon SC, Stein MD, Etheridge RM, D'Aunno TA. Linkage to medical services in the Drug Abuse Treatment Outcome Study. Med Care 2001;39:284-95.
106. Friedmann PD, Zhang Z, Hendrickson J, Stein MD, Gerstein DR. Effect of primary medical care on addiction and medical severity in substance abuse treatment programs. J Gen Intern Med 2003;18:1-8.
107. Weisner C, Mertens J, Parthasarathy S, Moore C, Lu Y. Integrating primary medical care with addiction treatment: a randomized controlled trial. JAMA 2001;286:1715-23.
108. Rothman J, Rudnick D, Slifer M, Agins B, Heiner K, Birkhead G. Colocated substance use treatment and HIV prevention and primary care services, New York State, 1990-2002: a model for effective service delivery to a high-risk population. J Urban Health 2007;84:226-42.
109. 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.
110. Takizawa C, Cheng DM, Samet JH, Winter M, Larson MJ, Saitz R. Primary medical care can reduce HIV risk behaviors in adults with addictions. J Addict Dis. In press 2007.
111. Committee on Crossing the Quality Chasm: Adaptation to Mental Health and Addictive Disorders-Institute of Medicine. Coordinating Care for Better Mental, Substance-Use, and General Health. Improving the quality of health care for mental and substance-use conditions. Washington, DC: National Academy Press; 2006.p.210-58.
112. McLellan AT, Hagan TA, Levine M, Gould F, Meyers K, Bencivengo M, Durell J. Supplemental social services improve outcomes in public addiction treatment. Addiction 1998;93:1489-99.
113. Shwartz M, Baker G, Mulvey KP, Plough A. Improving publicly funded substance abuse treatment: the value of case management. Am J Public Health 1997;87:1659-64.
114. Kushel MB, Colfax G, Ragland K, Heineman A, Palacio H, Bangsberg DR. Case management is associated with improved antiretroviral adherence and CD4+ cell counts in homeless and marginally housed individuals with HIV infection. Clin Infect Dis 2006;43:234-42.
115. Samet JH, Rollnick S, Barnes H. Beyond CAGE. A brief clinical approach after detection of substance abuse. Arch Intern Med 1996;156:2287-93.
116. Miller WR, Rollnick S. Motivational interviewing. Preparing people to change addictive behavior. New York: The Guilford Press; 1991.
117. National Institute on Drug Abuse. Therapeutic Community. 2002. Available at: http://www.drugabuse.gov/PDF/RRTherapeutic.pdf. Accessed June 26, 2007.
118. Harrison LD, Scarpitti FR. Introduction: progress and issues in drug treatment courts. Subst Use Misuse 2002; 37:1441-67.
119. Prendergast M, Podus D, Finney J, Greenwell L, Roll J. Contingency management for treatment of substance use disorders: a meta-analysis. Addiction 2006;101:1546-60.
120. Strain EC, Stitzer ML. Methadone treatment for opioid dependence. Baltimore: Johns Hopkins University Press; 1999.
121. Strain EC, Bigelow GE, Liebson IA, Stitzer ML. Moderate- *vs* high-dose methadone in the treatment of opioid dependence: a randomized trial. JAMA 1999;281:1000-5.
122. Palepu A, Tyndall MW, Joy R, Kerr T, Wood E, Press N *et al.* Antiretroviral adherence and HIV treatment outcomes among HIV/HCV co-infected injection drug users: the role of methadone maintenance therapy. Drug Alcohol Depend 2006;84:188-94.
123. Lucas GM, Mullen BA, 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 concurrent comparison groups. Clin Infect Dis 2006;42:1628-35.
124. Fiellin DA, O'Connor PG. Clinical practice. Office-based treatment of opioid-dependent patients. N Engl J Med 2002;347:817-23.
125. Johnson RE, Chutuape MA, Strain EC, Walsh SL, Stitzer ML, Bigelow GE. A comparison of levomethadyl acetate, buprenorphine, and methadone for opioid dependence. N Engl J Med 2000;343:1290-7.
126. Sullivan LE, Barry D, Moore BA, Chawarski MC, Tetrault JM, Pantalon MV *et al.* A trial of integrated buprenorphine/naloxone and HIV clinical care. Clin Infect Dis 2006;43 Suppl 4:S184-90.
127. McCance-Katz EF, Moody DE, Smith PF, Morse GD, Friedland G, Pade P *et al.* Interaction between buprenorphine and antiretrovirals.II. The protease inhibitors nelfinavir, lopinavir/ritonavir, and ritonavir. Clin Infect Dis. 2006; 43:S235-S246.
128. Gourevitch MN, Friedland GH. Interactions between methadone and medications used to treat HIV infection: a review. Mt Sinai J Med 2000;67:429-36.
129. 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.
130. Alford DP, Compton P, Samet JH. Acute pain management for patients receiving maintenance methadone or buprenorphine therapy. Ann Intern Med 2006;144:127-34.
131. Scimeca MM, Savage SR, Portenoy R, Lowinson J. Treatment of pain in methadone-maintained patients. Mt Sinai J Med 2000;67:412-22.
132. Fogarty A, Rawstorne P, Prestage G, Crawford J, Grierson J, Kippax S. Marijuana as therapy for people living with HIV/AIDS: social and health aspects. AIDS Care 2007;19:295-301.
133. Abrams DI, Jay CA, Shade SB, Vizoso H, Reda H, Press S, Kelly ME, Rowbotham MC, Petersen KL. Cannabis in painful HIV-associated sensory neuropathy: a randomized placebo-controlled trial. Neurology 2007;68:515-21.

Available online at www.sciencedirect.com
ScienceDirect
Preventive Medicine 44 (2007) 198–201
Preventive Medicine
www.elsevier.com/locate/ypmed

The effect of a primary care exercise intervention for rural women

Bonnie J. Sherman [a,*], Gina Gilliland [b], Jeanne L. Speckman [c], Karen M. Freund [a]

[a] *The Women's Health Unit, Evans Department of Medicine, and the Women's Health Interdisciplinary Research Center, Boston University Medical Center, 720 Harrison Avenue, #1108, Boston, MA 02118, USA*
[b] *Missouri Community Center of Excellence In Women's Health, USA*
[c] *Health Care Research Unit, Evans Department of Medicine, Boston University Medical Center, USA*

Available online 19 December 2006

Abstract

Objective. Rural women have limited exercise opportunities and significant barriers to engaging in physical activity. This study assessed the effect of a brief primary care based walking intervention in rural women.

Methods. The participants were recruited in March, 2003 by a primary care nurse at three locations in rural Missouri. The enrolled subjects were given a pedometer, exercise videotape and provided exercise counseling at intake and four time points over 6 months. The week 1 pedometer step counts were compared with step counts at 6-month follow-up.

Results. Of the initial 75 participants, 61 completed at least one follow up encounter. The participant's mean age was 42.5 years. At intake, the majority of women (90%) exhibited one or more risk factors for cardiovascular disease; 78% were obese or overweight. Although most (62%) women reported being physically active, the mean pedometer reading was low at 6337 steps per day at week 1. Over the follow-up period, participants increased their step counts by a mean of 2573 steps per day ($p < .001$). Increases in step counts were seen in normal weight, overweight and obese participants.

Conclusions. A simple walking intervention through a primary care practice was effective in increasing the short term walking rates of rural women.

Keywords: Rural health; Female; Exercise; Brief intervention; Obesity

Introduction

The combination of physical inactivity and poor diet is the second leading cause of death in the United States (Mokdad et al., 2004), and a major contributor to disease and disability. Women report personal, social and environmental barriers to engaging in sustained physical activity (Ainsworth et al., 2003). Rural women often experience greater barriers than urban women, due to higher rates of poverty, chronic disease and illness, lower levels of education and greater distances to travel for health care (Mulder et al., 2006). The Centers for Disease Control and Prevention (CDC) indicated that rural women are more sedentary than urban and suburban women based on data from the Behavioral Risk Factor Surveillance System Survey (BRFSS) (Centers for Disease Control and Prevention, 1998). Occupational physical activity of rural farm workers has also decreased due to the increased automation of the agriculture industry (Hill and Melanson, 1999). Rural women report that the lack of formal exercise programs, caregiving responsibilities, lack of streetlights and sidewalks and little opportunity to incorporate physical exercise into daily activity prevent them from being more physically active (Wilcox et al., 2000; Eyler, 2003; Deshpande et al., 2005). This has resulted in an epidemic of sedentary behavior and obesity among rural women.

Primary care providers are on the front lines of managing the medical complications of obesity and physical inactivity, however their potential role in prevention and health promotion has not been clearly demonstrated. Despite their unique position, there have been few primary care practice-based programs that have addressed physical inactivity. The Activity Counseling Trial (ACT) (Activity Counseling Trial Research Group, 2001) compared simple physician advice to physician advice combined with an ongoing educator intervention, telephone behavioral counseling, regular activity logs, and pedometer feedback. They found that among inactive primary

* Corresponding author. Fax: +1 617 638 8026.
E-mail address: bsherman@bu.edu (B.J. Sherman).

0091-7435/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ypmed.2006.10.009

care patients, simple exercise advice may provide benefit to men but not to women. Women required behavioral counseling, feedback and telephone contact to increase exercise activity, and improve their physiologic measures of physical fitness. The intensity of the ACT intervention and the homogenous demographics of the urban, educated, middle-income study population raise questions as to its generalizability to rural populations. The purpose of this study is to assess the feasibility of a simple primary care-based walking intervention for rural women.

Methods

We conducted an exercise intervention in a cohort of rural women from March to September, 2003. This study was a collaboration between the Northeast Missouri Health Council and Boston University School of Medicine and was approved by the Boston University Institutional Review Board. Women were recruited through a nurse at the Northeast Missouri Health Council Primary Care Clinic, or through advertisements placed by the primary care providers at a local hair salon and food establishment. As the study was conducted in a small town (population under 17,000), the main source of recruitment came from advertisements within the community and through word of mouth. The recruitment was conducted at all sites concurrently, and the first 75 women who expressed interest in the study and completed the baseline assessment were enrolled. All enrolled participants were patients at the Northeast Missouri Health Council Primary Care Clinic. Recruitment was completed within 2 day, indicating the interest in the community for such a program. The baseline assessment was conducted at the primary care practice site. A registered nurse within the clinic obtained informed consent, demographic information and asked a single question of whether participants exercised regularly. In addition, blood was taken at this baseline appointment to assess non-fasting serum glucose and total cholesterol. The registered nurse also took a blood pressure, height, weight and calculated a Body Mass Index (Weight in pounds ÷ Height in inches ÷ Height in inches) × 703. All women were given free of charge, an exercise videotape, a pedometer (Accusplit model AX-120, San Jose, California) and written information on the importance of exercise in maintaining good health. The exercise tape was a commercially available walking program by a well-known fitness expert, which encouraged viewers to walk two miles over the course of the tape. Participants were also asked to keep a daily log of their step counts, and instructed to utilize the exercise videotape as part of their exercise regime. Participants were contacted by telephone by the program coordinator, a nurse practitioner within the clinic, at set time intervals (1 week, 1 month, 3 months and 6 months) and asked for their step counts from the preceding 3 days. Previous research has indicated that a 3-day pedometer recall provides a sufficient estimate of mean pedometer-determined steps/day (Tudor-Locke et al., 2005). Data was collected at all four time points to encourage continued participant engagement in the intervention. The change over the course of the study was calculated by comparing the pedometer counts at week 1 of the intervention with counts at the 6-month follow up telephone contact. If patients could not be reached at the 6-month time point, we carried forward the last known pedometer count since the 1 week assessment, either at 1 month or 3 months of the study as the 6 month time point. Participants were included in the analysis if they had both a 1-week and at least 1 month of follow-up data. We repeated our analysis excluding those who failed to complete the 6 month protocol.

Descriptive statistics were calculated, and normally distributed variables (i.e. age) were compared with *t*-tests or analysis of variance. Non-normally distributed variables (i.e. days on protocol) were compared between groups with Kruskal–Wallis tests of the Wilcoxon Rank Sum. Each woman was categorized based upon her calculated body mass index (BMI). Normal weight was classified as a BMI under 25, overweight was a BMI of 25 or more but less than 30, and obese was a BMI 30 or more.

Comparisons of mean pre and post intervention pedometer readings were assessed by analysis of variance (ANOVA). Individual paired *t*-tests were conducted for each of the three BMI groups and by whether they exercised regularly at baseline.

Results

Of the 75 women enrolled, 44 women completed the entire 6-month protocol, 10 completed 3 months of the protocol and 7 completed 1 month of the protocol for a total of 61 (80%). The data from one participant was removed from the analysis as the step counts reported were statistical outliers.

Among the 60 women included in the final analysis, the average age was 42 years (range 22–64 years). All participants were Caucasian, which is representative of the demographics of this town. Twenty-five women (42%) had Medicare or Medicaid insurance, 26 women (43%) had a private form of insurance, and 9 women (15%) had unknown insurance status or were self-pay. The mean BMI of participants was 30.6 (S.D. = 7.4), with 13 women normal weight (22%), 17 women were overweight (28%) and 30 women being obese (50%). Thirty-seven women (62%) reported at intake that they engaged in physical activity. There was no statistically significant difference in number of days in the program or in age between the normal, overweight and obese groups (*p*-values of <0.90 and <0.63). Overall 54 women (90%) had one of more risk factors for developing cardiovascular disease. Thirty-two women (53%) had high blood pressure (defined as over 120 or greater systolic), three (5%) had elevated random serum glucose levels (over 200 mg/dL), 25 women (42%) had elevated total serum cholesterol (over 200 mg/dL) and 47 women (78%) were classified as overweight or obese.

As shown in Table 1, the mean week 1 pedometer reading was 6337 steps for the entire group. Stratifying women by BMI, a statistically significant inverse trend between exercise and weight was noted: mean baseline step counts were 7415 for normal weight women, 6908 for overweight and 5545 in obese women ($f = 3.15$, $p < .05$). The average length of time from the start date to end date was 146 days (standard deviation (S.D.) = 46, range 41–195). The overall step counts rose by 2573 after the intervention. Assuming a stride length of 2.25 ft per step, the widely accepted norm for women, the

Table 1
Mean daily pedometer counts before and after intervention in rural women in Kirksville, Missouri from March–September, 2003

Group	*N*	Pre intervention pedometer count mean (S.D).	Post intervention pedometer count mean (S.D)	Difference between pre and post count mean (S.D.)	*P*-value*
Entire group	60	6337 (2591)	8910 (3113)	2573 (2638)	0.001
BMI					
Normal weight	13	7415 (2897)	9900 (2373)	2484 (2790)	0.005
Overweight	17	6903 (2159)	10,256 (3092)	3348 (2527)	0.0005
Obese	30	5545 (2502)	7718 (3023)	2173 (2626)	0.0005
Activity level					
Active	37	6972 (2576)	9899 (2872)	2927 (2312)	0.0005
Inactive	23	5314 (2374)	7318 (2862)	2003 (2059)	0.0005

S.D = standard deviation; BMI = body mass index; *Individual paired *t*-test.

step count amounts to an increase of approximately 1.1 miles (2573 steps × 2.25 ft per step/5280 ft per mile = 1.09 miles). There was a statistically significant difference in final mean step count between BMI groups, with overweight women performing best ($f = 5.06$, $p < .01$); however, the group difference between pre and post measurements was not statistically significant. Comparison of individual changes in step counts, both overall and in each BMI group showed a significant increase in mean step count, with greatest increase found in overweight women. The mean pedometer reading at week 1 was statistically different for women reporting physical activity (mean 6972) and those who did not (mean 5314) ($f = 6.3$, $p < 0.01$). The individual paired *t*-tests of the increase in step count were statistically significant for both women who reported at baseline being active and those who were inactive, indicating that both groups benefited from the intervention. The paired *t*-tests remain statistically significant even when patients who failed to complete the 6 month protocol are excluded.

Discussion

This study suggests that a brief intervention based within a primary care setting can achieve short-term increases in physical activity in rural women. A videotape with home based exercises, a pedometer to provide feedback on physical activities, and reinforcement through telephone contact with nursing staff within the primary care practice were used in combination as the intervention within this study.

Given the epidemic of obesity in this country, and its medical sequelae, there is a need for multiple interventions at the individual, group and community level. Physician advice and awareness of weight as a health risk have been found to be motivational factors for improving diet and exercise concurrently (Huang et al., 2004). While the benefit of a preventive primary care-based intervention has been proven for smoking cessation (Abdullah and Simon, 2006; Fiore et al., 2004), there is less evidence to support the role of the primary care practice in obesity management. The rapid enrollment in this study underscores the need for such an intervention within this rural community as well as the participants' trust of the primary care practice. The use of nursing staff within the primary care practice for short telephone follow-up and brief behavioral interventions is a cost effective utilization of resources, and builds upon preexisting relationships to promote exercise.

The use of pedometers to assess and monitor activity levels in large groups of people is appealing, as they are simple to use, inexpensive and measure walking distances with acceptable accuracy (Tudor-Locke et al., 2002). Pedometers have shown to assist in goal setting and increase motivation in minimal contact exercise interventions (Heesch et al., 2005). Pedometers are an effective way to increase physical activity levels (Stovitz et al., 2005; Croteau, 2004) even though their role as a biofeedback tool and motivational aid is unclear. The mean step counts of approximately 6300 steps per day at the start of this study indicated that the study population was relatively inactive for their age compared to other Americans. Previous research has found that the average 20 to 50 year old walks between 7000 and 13,000 steps per day, and average person over 50 years of age walks 6000 to 8500 steps per day (Tudor-Locke and Myers, 2001). Women engage in fewer leisure activities and take fewer steps compared to men, so these numbers may overestimate the activity levels of women (Chan et al., 2004; Bennett et al., 2006). During the 6-month follow-up period, there was an average increase of over 2500 steps per day, which translates into participants walking over a mile more per day at the end of the intervention. All groups of participants benefited from the intervention and had significant increases in their step count, regardless of Body Mass Index or self-reported exercise status at intake. We only collected vital signs at intake and therefore are unable to comment on whether this increase in exercise resulted in a change in weight or blood pressure. However, previous research has indicated that any increase in exercise level can decrease cardiovascular risk factors regardless of a change in BMI (Rimmer et al., 2002; Tully et al., 2005). Over 78% of participants in this study can be categorized as overweight or obese, a group who has previously shown the greatest benefits from engaging in a physical activity program (Williams, 2005).

While patients with obesity and weight-related conditions access health care services more frequently (Bertakis and Azari, 2005), the role of the primary care practice in supporting physical activity has received little attention. Primary care physicians and nursing staff are essential to motivate patients to adopt a healthy lifestyle, but there are many barriers that exist that prevent them from being successful. Competing demands for time, inadequate counseling skills and knowledge, and the lack of reimbursement prevent primary care providers from engaging in meaningful interaction with patients regarding health promotion and disease prevention (Huang et al., 2004; Brotonsc et al., 2005). Despite these challenges, the physical consequences associated with physical inactivity and obesity are reduced with even small amounts of physical activity. A modest 5% improvement in cardiorespiratory fitness has been associated with a 9% reduction in mortality risk and improved disease outcome (Activity Counseling Trial Research Group, 2001; Blair and Connelly, 1996). The average increase in step count in our study would yield cardiorespiratory changes greater than 5%, based on previous research which has correlated increases in step count with cardiorespiratory fitness (Perna et al., 2005). Other work has shown that walking even 3 miles per week can greatly improve disease outcome (Holmes et al., 2005). This underscores the importance of even modest improvements in fitness level in overall cardiovascular risk reduction.

Study limitations

Our study has several limitations. One is that it only assessed short-term outcomes. The study findings would have been stronger if a control group was used and had the groups been randomized to their group. We did not have sufficient information to generalize these findings to morbidly obese or very inactive women. An additional limitation is that the self-reported step counts were not validated by another measure and

were the only physical activity outcome reported. Finally our study participants were comprised of a convenience sample of self-selected Caucasian women and were not a randomized sample. Studies are needed to determine if long-term benefit can be derived from instituting such an intervention in the primary care setting, with intermittent reinforcement during other primary care visits.

Conclusion

These initial findings are promising that primary care practices can begin to address obesity through specific interventions targeted at modifying physical activity.

Acknowledgments

This work was supported, in part, by a grant by the Department of Health and Human Services, and the Office of Women's Health (#01T020142).

References

Abdullah, A.S., Simon, J.L., 2006. Health promotion in older adults: evidence-based smoking cessation programs for use in primary care settings. Geriatrics 61, 30–34.

Activity Counseling Trial Research Group, 2001. Effects of physical activity counseling in primary care. The Activity Counseling Trial: a randomized controlled trial. J. Am. Med. Assoc. 286, 677–687.

Ainsworth, B.E., Wilcox, S., Thompson, W.W., Richter, D.L., Henderson, K.A., 2003. Personal, social and physical environmental correlates of physical activity in African-American women in South Carolina. Am. J. Prev. Med. 25, 23–29.

Bennett, G., Wolin, K.Y., Puleo, E., Emmons, K.M., 2006. Pedometer-determined physical activity among multiethnic low-income housing residents. Med. Sci. Sports Exerc. 28, 768–773.

Bertakis, K.D., Azari, R., 2005. Obesity and the use of health care services. Obes. Res. 13, 372–379.

Blair, S.N., Connelly, J.C., 1996. How much physical activity should we do? The case for moderate amounts and intensities of physical activity. Med. Sci. Sports Exerc. 67, 193–205.

Brotonsc, C., Bjorkelund, C., Bule, M., et al., 2005. Prevention and health promotion in clinical practice: the views of the general practitioners in Europe. Prev. Med. 40, 595–601.

Centers for Disease Control and Prevention, 1998. Self-reported physical activity by degree of urbanization—United States, 1996. Morb. Mortal. Wkly. Rep. 47, 1097–1100.

Chan, C.B., Ryan, D.A., Tudor-Locke, C., 2004. Health Benefits of a pedometer-based physical activity intervention in sedentary workers. Prev. Med. 39, 1215–1222.

Croteau, K.A., 2004. A preliminary study on the impact of pedometer based intervention on daily steps. Am. J. Health Promot. 18, 217–220.

Deshpande, A.A., Baker, E.A., Lovegreen, S.L., Brownson, R.C., 2005. Environmental correlates of physical activity among individuals with diabetes in the rural Midwest. Diabetes Care 28, 1012–1018.

Eyler, A.A., 2003. Personal, social and environmental correlates of physical activity in rural midwestern White women. Am. J. Prev. Med. 25, 86–92.

Fiore, M.C., McCarthy, D.E., Jackson, T.C., et al., 2004. Integrating smoking cessation treatment into primary care: an effectiveness study. Prev. Med. 38, 412–420.

Heesch, K.C., Dinger, M.K., McClary, K.R., Rice, K.R., 2005. Experiences of women in a minimal contact pedometer-based intervention: a qualitative study. Women Health 41, 97–116.

Hill, J.O., Melanson, E.L., 1999. Overview of the determinants of overweight and obesity: current evidence and research issues. Med. Sci. Sports Exerc. 31, S515–S521.

Holmes, M.D., Chen, W.Y., Feskanich, D., Kroenke, C.H., Colditz, G.A., 2005. Physical activity and survival after breast cancer diagnosis. J. Am. Med. Assoc. 293, 2479–2486.

Huang, J., Yu, H., Marin, E., et al., 2004. Physicians weight loss counseling in two public hospital primary care clinics. Acad. Med. 79, 156–161.

Mokdad, A.H., Marks, J.S., Stroup, D.F., Gerberding, J.L., 2004. Actual causes of death in the United States, 2000. J. Am. Med. Assoc. 291, 1238–1245.

Mulder, P.L., Shellenberger, S., Streigel, R. et al. The behavioral health care needs of rural women. Available at: http://www.apa.org/rural/ruralwomen.pdf (Accessed July 10, 2006).

Perna, F.M., Craft, L., Skrinar, G., et al., 2005. Exercise adherence counseling effects on physical activity among breast cancer patients. Ann. Behav. Med. 29, 43.

Rimmer, J.H., Nicola, R., Riley, B., Creviston, T., 2002. Exercise training for African Americans with disabilities residing in difficult social environments. Am. J. Prev. Med. 23, 290–295.

Stovitz, S.D., VanWormer, J.J, Center, B.A, Bremer, K.L., 2005. Pedometers as a means to increase ambulatory activity for patients seen at a family medicine clinic. J. Am. Board Fam. Med. 18, 335–343.

Tudor-Locke, C.E., Myers, A.M., 2001. Methodological consideration for researchers and practitioners using pedometers to measure physical (ambulatory) activity. Res. Q. Exerc. Sport 72, 1–12.

Tudor-Locke, C., Williams, J.E, Reis, J.P., Pluto, D., 2002. Utility of pedometers for assessing physical activity; convergent validity. Sports Med. 31, 795–808.

Tudor-Locke, C., Burkett, L., Reis, J.P., et al., 2005. How many days of pedometer monitoring predict weekly physical activity in adults? Prev. Med. 40 (3), 293–298.

Tully, M.A., Cupples, M.E., Chan, W.S., McGlade, K., Young, I.S., 2005. Brisk walking, fitness and cardiovascular risk: a randomized controlled trial in primary care. Prev. Med. 41, 622–628.

Wilcox, S., Castro, C., King, A.C., Housemann, R., Brownson, R.C., 2000. Determinants of leisure time physical activity in rural compared with urban older and ethnically diverse women in the United States. J. Epidemiol. Commun. Health 54, 667–672.

Williams, T., 2005. Nonlinear relationships between weekly walking distance and adiposity in 27,596 women. Med. Sci. Sports Exerc. 37, 1893–1901.

Primary Medical Care and Reductions in HIV Risk Behaviors in Adults with Addictions

Carolina Takizawa, MD
Debbie Cheng, ScD
Jeffrey Samet, MD
Michael Winter, MPH
Mary Jo Larson, PhD
Richard Saitz, MD

ABSTRACT. Human immunodeficiency virus (HIV) risk behaviors are prevalent in persons with addictions.

Objectives: To assess whether exposure to primary medical care is associated with decreases in HIV risk behaviors.

Design: Prospective 2-year cohort of 298 adults with addictions.

Outcomes: Sex and drug-related HIV risk behaviors, measured by the Risk Assessment Battery.

Predictors: Cumulative number of primary care visits (0, 1, $\geq$ 2). Associations were tested using regression models for correlated data.

Results: In women, receipt of primary care was associated with less sex risk behavior (mean decrease 2.1, $p \leq 0.1$). Among women and men, $\geq$ 2 primary care visits was associated with lower odds of any drug risk behavior (OR = 0.37, p = 0.03).

Conclusions: Exposure to primary care can impact HIV risk behavior favorably among adults with addictions. doi:10.1300/J069v26n03_03 *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <http://www.HaworthPress.com>*

Carolina Takizawa is affiliated with Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center & Boston University School of Medicine, Boston, MA 02118.

Debbie Cheng is affiliated with the Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118.

Jeffrey Samet is affiliated with Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center & Boston University School of Medicine, and is also affiliated with the Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA 02118.

Michael Winter is affiliated with Data Coordinating Center, Boston University School of Public Health, Boston, MA 02118.

Mary Jo Larson is affiliated with New England Research Institutes, Watertown, MA 02472.

Richard Saitz is affiliated with Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center & Boston University School of Medicine, and also the Youth Alcohol Prevention Center, and the Department of Epidemiology, Boston University School of Public Health, Boston, MA 02118.

Address correspondence to: Carolina Takizawa, MD, 519 Washington Street #7, Brookline, MA 02446 (E-mail: kinacmt@hotmail.com).

The authors have no conflict of interest to report with regards to this manuscript.

This research was supported by the National Institute on Drug Abuse (RO1-DA10019) and National Institute on Alcohol Abuse and Alcoholism (RO1-AA10870). Dr. Takizawa was supported by a Health Resources and Services Administration (HRSA) development grant (5-D55-HP00056-05). A limited portion of the manuscript was presented as a poster at the 66th annual Scientific Meeting of the College on Problems of Drug Dependence.

KEYWORDS. Primary care, HIV, risk behavior, addictions

INTRODUCTION

In the U.S., 5 million adults engage in HIV risk behavior.[1] These behaviors, such as unprotected sex and injection of illicit drugs, are more prevalent in persons with addictions.[2,3] There are 800,000-900,000 HIV-infected persons in the U.S.; approximately 40,000 new cases are diagnosed yearly, most of whom are persons engaging in high-risk behavior.[4]

Fortunately, randomized controlled trials and observational studies suggest that HIV prevention programs decrease the likelihood of engaging in risky behaviors. Such programs have targeted high-risk populations.[5-10] HIV prevention programs in alcohol[11,12] and drug abuse[13-17] treatment settings have also demonstrated success. Programs that offer, in addition, on-site primary medical care have shown improvements in other medical and social aspects in persons with addictions.[18-20] However, few people at risk receive health care in these settings.

The United States Preventive Services Task Force recommends that all adolescents and adults be assessed and counseled regarding HIV risk behavior.[21] However, such practices are not, as yet, routinely performed in primary care.[22-24] Nonetheless, it is likely that some efforts in HIV risk reduction are occurring in such settings. It is possible that exposure to primary care for persons with addictions plays a role in moderating health risk behaviors, including those that are HIV-related. For example, primary medical care offered in an addiction program setting has been associated with lower addiction severity.[19]

Although the efficacy of HIV-prevention educational tools in primary care has been studied, the impact of HIV risk reduction counseling delivered outside of research settings (e.g., as delivered in usual primary care offices) has not been well described.[25,26] This study aimed to test the hypothesis that linkage to primary care in the community by persons with addictions without primary care is associated with reduced HIV sex and drug risk behavior.

METHODS

Study Design

This study is a prospective cohort of adults with addictions, recruited from a residential detoxification program that represents a secondary data analysis of a randomized controlled trial. Subjects were participants in the Health Evaluation and Linkage to Primary care (HELP) study, a randomized controlled trial of subjects (adults with addictions) to either a single multidisciplinary evaluation and referral to facilitate linkage of patients to primary care (enhanced care) or to usual care at the detoxification program. Between 1997 and 1999, 470 of 642 eligible subjects provided informed consent and agreed to participate in the HELP study; 69% of those randomized to enhanced care and 53% of those randomized to usual care, linked to primary care. Eighty-five percent of all subjects had at least one follow-up visit after study entry. Sixty percent of subjects did not have health insurance; half were homeless; and all resided in proximity to the detoxification center (Boston). The intervention (enhanced care) and the control (usual care) groups did not focus on HIV risk counseling. Details of the trial have been published elsewhere.[27] This current study was a secondary analysis of data collected from subjects in the HELP study.

The study was approved by the Institutional Review Board at the Boston University Medical Center; additional confidentiality protection was provided by a certificate of confidentiality from the National Institute on Alcohol Abuse and Alcoholism. All subjects provided written informed consent.

Subjects

Subjects were enrolled from June 1997 to March 1999. Inclusion criteria were the following: 18 years of age or older; alcohol, cocaine, or heroin as first or second drug of choice; and no primary care (PC) provider. Patients were excluded if they were unable to speak either English or Spanish, were pregnant, unable to provide contacts for follow-up, had plans to

leave the area, or had evidence of dementia (Mini-Mental state examination score < 21). Research interviews were scheduled at 6-month intervals for a total of 2 years. The current cohort analysis was restricted to subjects who had at least 2 follow-up research interviews after study entry to ensure that HIV risk behaviors could be assessed at an interview subsequent to that where the number of PC visits was recorded. Of the 470 subjects in the study, 298 met this last criterion and comprised our full study cohort.

Assessments

Trained research associates interviewed subjects at the detoxification unit after enrollment and resolution of acute withdrawal symptoms. Assessments, which were repeated at each follow-up research interview, included demographics, receipt of primary care (PC), HIV risk behaviors, alcohol and drug addiction severity,[28] and health-related quality of life.[29] Receipt of PC was assessed by asking subjects, at each follow-up interview, if they had visited a doctor, nurse practitioner, or physician assistant. Subjects were asked how many times in the past 6 months (or since the last research interview) they had seen that clinician. Subjects were asked the clinician's specialty, whether they would call that clinician for non-emergent issues, or whether that clinician could be their regular personal doctor. When patients did not know the specialty, attempts to contact the clinician were made to determine it.

HIV sex and drug risk behavior were assessed by the short version of the Risk Assessment Battery (RAB).[30-32] Items for each component (sex and drug risk) asked about HIV risk behavior in the past six months. The RAB-sex component is composed of six items, addressing the number of partners, sexual orientation, use of condoms, sex trade, and visits to crack houses. The RAB-drug component is composed of 8 items, addressing injection of drugs, needle sharing, and visits to a shooting gallery. The validity and reliability of the RAB instrument has previously been established.[30]

Study entry alcohol and drug addiction severity were assessed using the Addiction Severity Index (ASI). Health-related quality of life was assessed using the Short Form Health Survey (SF-36).[28,29]

Dependent Variables/Outcomes

There were two main outcomes of interest assessed at the time point subsequent to the primary care exposure assessment. The RAB-sex score was modeled as a continuous variable, with scores ranging from 0 (least risk) to 18 (most risk). The RAB-drug score, with scores ranging from 0 (least risk) to 22 (most risk), was dichotomized to 0 (no HIV drug risk behavior) and $\geq$ 1 (any HIV drug risk behavior) for two reasons: (1) Most (63%) of the RAB-drug scores were 0, and (2) a RAB-drug score of 0 indicates that the subject did not inject drugs in the past 6 months. A RAB-drug score of 1 indicates injecting drugs in the past 6 months (except for 4 subjects who did not report injecting in the past 6 months but had visited a shooting gallery) and a score of two or more indicates that the individual injected drugs in the past six months and had some degree of needle sharing.

Independent Variables/Predictors

The primary independent variable was the self-reported cumulative number of primary care (PC) visits received after study entry as assessed at each follow-up interview. For the visit to be defined as PC, two criteria needed to be met: (1) the subject reported having a regular personal doctor, would call the doctor for a non-emergent issue, or saw a doctor that could be their regular personal doctor; and (2) the doctor was in a specialty that could be considered PC, including obstetrics/gynecology, family medicine, adolescent medicine, internal medicine, pediatrics, AIDS medicine, pulmonology, cardiology, or gastroenterology. For 92% of all subject visits defined as PC, the clinician's specialty was confirmed either by direct contact or by self-report. The remaining 8% of visits met the first criterion, but the specialty of the provider remained unverified; visits to these clinicians were also considered PC.

The cumulative number of PC visits between baseline and follow-up interviews was categorized as 0 (no PC visits), 1 (1 PC visit) and $\geq$ 2 (2 or more PC visits), due to the skewed distribution of the number of PC visits.[19] The cumula-

tive number of PC visits was modeled as a time-varying covariate that predicted HIV risk behaviors assessed at the subsequent follow-up research interview.

Self-report of PC was compared to administrative data sources from PC practices based at the Boston Medical Center and the Boston Health Care for the Homeless Program. Ninety-five percent of all study subjects had data available from these sources; of those who had PC visits in these data, 81% self-reported receipt of PC.

In the analyses, other variables (assessed at study entry) were considered: age, marital status, race/ethnicity, ASI (Addiction Severity Index) drug and alcohol scales, the mental (MCS) and physical (PCS) component summary scores from the SF-36, and randomization group (usual care or enhanced evaluation and referral at study entry). The alcohol and drug scales of the ASI were used to summarize addiction severity, with scores ranging from 0 (no severity) to 1 (highest severity). The MCS and PCS scores (scale of 0-100) are standardized to the U.S. general population; mean scores for each are 50, with a standard deviation of 10. Scores below this mean represent individuals with worse physical or mental health-related quality of life.

Statistical Analysis

Analyses were performed using SAS/STAT software (Version 8.2.). The analyses of the RAB-sex score focused on the overall sample, referred to as the "HIV sex risk sample" (n = 298), and analyses of the RAB-drug score focused on a subgroup that reported ever injecting drugs at study entry ("HIV drug risk sample," n = 110).

Descriptive statistics for study entry characteristics were obtained separately for each sample. RAB scores were analyzed using longitudinal regression models to account for correlation due to repeated observations on the same subject. Fully adjusted regression models controlled for time since study entry, gender, race/ethnicity, randomization group, ASI-alcohol, ASI-drug, MCS, PCS, and marital status (RAB-sex score analyses only). The main independent variable, cumulative number of primary care (PC) visits as assessed at each follow-up interview, was time-varying. The expected effect of a particular dose or threshold of PC on HIV risk behavior was not known, therefore each level of PC exposure were compared to no receipt of PC in all analyses.

For the RAB-sex score, a continuous outcome, data were analyzed using generalized linear models for correlated data implemented with PROC MIXED. Unadjusted (adjusting only for time since study entry) and fully adjusted models were fit. A third model that included baseline RAB-sex scores in the fully adjusted models was also fit to the data. Because of likely differential response of women to PC and sexual risk behavior,[33,34] an interaction between gender and receipt of PC was tested. If there was evidence of an interaction ($p < 0.10$), analyses were subsequently stratified on gender.

For the RAB-drug score, a dichotomous outcome, data were analyzed using generalized estimating equations logistic regression models implemented with PROC GENMOD. Three models were fit: unadjusted, fully adjusted, and fully adjusted controlling additionally for RAB-drug scores at study entry. An interaction between gender and receipt of PC was not tested, as there was no hypothesis of differential response.

RESULTS

Patient Characteristics

Subject characteristics at study entry and subsequent receipt of primary care (PC) are presented in Table 1. These characteristics were similar to those of the 470 subjects in the original HELP study, with the exception of baseline RAB-drug scores (which as expected were higher in the HIV-drug risk sample). The majority of subjects was male, unmarried and had relatively poor mental health-related quality of life.

Of the 298 subjects, 4.7% were 18-24 years of age, 80.2% were 25-44 years, and 15.1% were 45 years or older. In addition, the length of stay in the detoxification center (from which the subjects were recruited) for the 298 subjects was 4.7 days (SD = 1.9 days). Twenty percent of the 298 subjects had asthma, emphysema or chronic lung disease; 16% had high blood pres-

TABLE 1. Subject Characteristics at Study Entry

	HIV-sex risk sample (n = 298)		HIV-drug risk sample (n = 110)*
	men (n = 222)	women (n = 76)	men and women (n = 110)
age, in years**	36.4(7.7)	36.6(7.3)	37.1(7.9)
% male	100	0	73
% married	8	8	10
% African American	54	53	32
% Hispanic	9	9	9
% White	32	32	53
% enhanced carea	54	47	52
ASI-drug**	0.3(0.2)	0.3(0.1)	0.3(0.1)
ASI-alcohol**	0.5(0.3)	0.4(0.4)	0.4(0.4)
PCS**	48.7(10.8)	44.3(10.2)	46.0(10.5)
MCS**	32.6(12.6)	28.2(11.0)	27.1(12.0)
RAB-sex**	4.6 (2.6)	5.1 (3.2)	4.8 (2.9)
RAB-drug**	2.1 (4.7)	2.0 (4.5)	5.4(6.3)

*Subset of the HIV-Sex Risk Sample

**Mean (SD)

[a]Randomization at the detoxification unit (study entry)

PC, primary care; n, number of subjects; ASI, addiction severity index; PCS, physical component summary score; MCS, mental component summary score; RAB, risk assessment battery.

TABLE 2. Receipt of Primary Care Over the 2-Year Period

	HIV-sex risk sample		HIV-drug risk sample
Receipt of PC	Men (n = 222)	Women (n = 76)	men and women (n = 110)
0	44%	22%	46%
1	16%	12%	7%
≥ 2	40%	66%	46%
Total	100%	100%	100%

PC, primary care

TABLE 3. Distribution of Observations by Receipt of Primary Care Over the 2-Year Period

	HIV-sex risk sample		HIV-drug risk sample
Receipt of PC	men #obs (%)	women #obs (%)	men and women #obs (%)
0	197(44)	43(29)	96(45)
1	74(17)	19(13)	22(10)
≥ 2	172(39)	84(58)	94(44)
Total	443(100)	146(100)	212(100)

PC, primary care; #obs, number of observations.

sure; 11% had a seizure disorder; 10% had chronic liver disorder; and 2.4% had diabetes. Fifty-one percent reported ever having a sexually transmitted disease, and 3% reported having been tested positive for HIV.

Over half of the subjects in each study sample received PC during the two-year period at least once (Table 2). The distribution of the cumulative number of PC visits for observations from the HIV-sex and drug risk samples is presented in Table 3. The proportions of 0 and ≥ 2 PC visits were similar in the male HIV-sex risk sample the HIV-drug risk sample. Among females in the HIV-sex risk sample, the majority of observations had ≥ 2 PC visits.

Receipt of Primary Care (PC) and HIV Sex Risk Behavior

Initial regression models of HIV-sex risk behavior suggested a possible interaction (p = 0.08) between gender and the receipt of PC. Therefore, separate longitudinal regression models were fit for women and men.

In women, receipt of PC was associated with a significant decrease in HIV sex risk behavior, when adjusting for all covariates. The mean adjusted decrease in RAB-sex scores was 1.3 and 2.3 for women who received ≥ 2 PC visits and 1 PC visit, respectively, compared to women who did not receive PC (Table 4). When additionally adjusting for HIV sex risk behavior at study entry, receipt of PC remained significantly associated with decreases in HIV sex risk behavior.

In men (Table 5), receipt of ≥ 2 PC visits was significantly associated with a decrease in HIV sex risk behavior in the unadjusted model (mean RAB-sex score decrease of 0.7; compared to men who did not receive PC; p = 0.01). In adjusted analyses, the association between receipt of PC and sex risk behavior was attenuated (p = 0.10).

Receipt of Primary Care (PC) and HIV Drug Risk Behavior

The number of PC visits was associated with a significant decrease in the odds of having any

TABLE 4. HIV-Sex Risk Behavior Decreases Over the Two-Year Period in Women, by Receipt of Primary Care

model	PC group	mean change in RAB-sex score (95% CI)	p-value
unadjusted[a]	PC = 0	(reference)	-
	PC = 1	−2.1 (−3.4,−0.8)	0.002
	PC≥ 2	−1.3 (−2.2,−0.4)	0.007
adjusted[b]	PC = 0	(reference)	-
	PC = 1	−2.3 (−3.6,−0.9)	0.001
	PC≥ 2	−1.3 (−2.3,−0.3)	0.009
adjusted[c]	PC = 0	(reference)	-
	PC = 1	−2.1 (−3.5,−0.8)	0.002
	PC≥ 2	−1.2 (−2.2,−0.3)	0.01

[a]adjusted for time since study entry

[b]adjusted for time since study entry, age, race/ethnicity, randomization group, marital status, ASI-drug, ASI-alcohol, PCS and MCS

[c]adjusted for time since study entry, age, race/ethnicity, randomization group, marital status, ASI-drug, ASI-alcohol, PCS, MCS and study entry RAB-sex score

PC, primary care; RAB, risk assessment battery; CI, confidence interval.

TABLE 5. HIV-Sex Risk Behavior Decreases Over the Two-Year Period in Men, by Receipt of Primary Care

model	PC group	mean change in RAB-sex score (95% CI)	p-value
unadjusted[a]	PC = 0	(reference)	-
	PC = 1	−0.2 (−0.9,0.4)	0.45
	PC≥ 2	−0.7 (−1.2,−0.2)	0.01
adjusted[b]	PC = 0	(reference)	-
	PC = 1	−0.2 (−0.8,0.5)	0.62
	PC≥ 2	−0.4 (−1.0,0.1)	0.14
adjusted[c]	PC = 0	(reference)	-
	PC = 1	−0.1 (−0.8,0.5)	0.63
	PC≥ 2	−0.4 (−0.9,0.1)	0.10

[a]adjusted for time since study entry

[b]adjusted for time since study entry, age, race/ethnicity, randomization group, marital status, ASI-drug, ASI-alcohol, PCS and MCS

[c]adjusted for time since study entry, age, race/ethnicity, randomization group, marital status, ASI-drug, ASI-alcohol, PCS, MCS and study entry RAB-sex score

PC, primary care; RAB, risk assessment battery; CI, confidence interval.

HIV drug risk behavior in some instances (Table 6). In all analyses, receipt of 1 PC visit was not significantly associated with a decrease in HIV drug risk behavior. In the adjusted model, subjects who received ≥ 2 PC visits had

TABLE 6. HIV-Drug Risk Behavior by Receipt of Primary Care

model	PC group	Odds of having any HIV drug risk behavior* OR (95% CI)	p-value
unadjusted[a]	PC = 0	(reference)	-
	PC = 1	1.33 (0.40,4.42)	0.64
	PC≥ 2	0.45 (0.21,0.93)	0.03
adjusted[b]	PC = 0	(reference)	-
	PC = 1	0.88 (0.28,2.83)	0.83
	PC≥ 2	0.37 (0.15,0.92)	0.03
adjusted[c]	PC = 0	(reference)	-
	PC = 1	1.10 (0.34,3.63)	0.87
	PC≥ 2	0.46 (0.17,1.26)	0.13

*odds of having RAB-drug score ≥ 1 (reference group RAB-drug score = 0)

[a]adjusted for time since study entry

[b]adjusted for time since study entry, age, race/ethnicity, randomization group, ASI-drug, ASI-alcohol, PCS and MCS

[c]adjusted for time since study entry, age, race/ethnicity, randomization group, ASI-drug, ASI-alcohol, PCS, MCS and study entry RAB-drug score

PC, primary care; OR, odds ratio; CI, confidence interval.

significantly lower odds (OR = 0.37, 95% CI = 0.15-0.92) of reporting any HIV drug risk behavior, compared to those who did not receive PC. When additionally adjusting for HIV drug risk behavior at study entry, this association was no longer statistically significant (OR = 0.46, 95% CI = 0.17-1.26), although the magnitude of association remained notable.

DISCUSSION

In this study, receipt of primary care (PC) was associated with a decrease in HIV sex risk behavior among women with addictions. In men, receipt of PC appeared to be associated with a decrease in HIV sex risk behavior; however, this decrease was attenuated and no longer statistically significant in adjusted analyses. Men and women with a history of injecting drugs were less likely to engage in HIV drug risk behavior if they received ≥ 2 PC visits over the two-year period. The magnitude of this association was notable but no longer statistically significant in an analysis adjusting for HIV drug risk behavior at study entry. In nearly all analyses, even after adjusting for potential confounders, the RAB-sex and drug scores were

lower among subjects exposed to PC. Although the significance of RAB score decreases did not persist in some adjusted analyses, the consistency of the direction of association between exposure to PC and these scores suggests that PC affects HIV risk behaviors.

We found an association between visiting a primary care provider and HIV risk behavior reduction. While we did not collect information on what counseling occurred (if any) at the primary care visits, there are at least three possible explanations for this association: primary care, either the visit and relationship itself or specific risk-reduction counseling that occurred during the visit, could cause risk behavior reduction; other exposures, such as specialized risk reduction counseling resulting from primary care referral or found by the patient him or herself, could lead to decreased risk behaviors; or, some other factor could cause both an increased likelihood to visit a primary care provider and cause a reduction in HIV risk behavior. To address the latter possibility, we performed analyses that adjusted for factors known to be associated with receipt of primary care (e.g., gender, medical illness). We recognize, however, that these analyses cannot completely eliminate confounding by self-selection to primary care. Only a study that randomized patients to primary care exposure could settle this question. Further studies on the content of primary care and utilization of specific primary care interventions are needed to better understand the association we observed between primary care and HIV risk behaviors. Nonetheless, it does appear that reduction in HIV risk behaviors can now be added to reductions in addiction severity as a potential benefit of primary medical care.[19]

In prior studies, brief counseling interventions have led to decreases in HIV risk behaviors. In a randomized controlled trial of brief counseling (two 20-minute face-to-face HIV education sessions) in sexually transmitted disease (STD) clinics, condom use increased and STD rates decreased in the subsequent year.[7] Similarly, prospective cohort studies in adults attending drug treatment programs, that included HIV risk counseling, observed decreases in HIV risk behavior.[11,17]

The magnitude of decrease in risk behavior associated with PC exposure in our cohort (e.g., 1-2 points in the RAB-sex score) is clinically important and similar to that seen in studies of risk reduction counseling. Cocaine-dependent subjects randomized to HIV risk reduction counseling (i.e., 2-3 sessions/week for 6 months), had mean RAB-sex score decreases of 1.5-2.7.[35] A 1-point decrease in the RAB-sex score could represent an array of changes: from using condoms most of the time to always, from having two sexual partners to one, from exchanging sex for drugs to no such exchange. A 1-point decrease in RAB-drug scores could represent a change from injecting to not injecting drugs, from sharing to not sharing needles, or from sharing to not sharing a cooker. A 2-point or more decrease in either score signifies either a greater decrease in a single risk behavior or a decrease in multiple risk behaviors.

There were strengths and limitations of this study. As an observational study, causality between receipt of PC and HIV risk behavior could not be established. Consistent dose-response relationships between PC and HIV risk behavior reductions were not observed. However, it is unclear whether PC exposure would operate in such a fashion or via a threshold effect. The observed associations could be due to uncontrolled confounding (i.e., unmeasured HIV risk reduction counseling or effects of the multidisciplinary assessment and referral intervention). However, data were collected prospectively, and possible confounders such as health and addiction severity, demographics, and risk behaviors at study entry were accounted for. Furthermore, the intervention (randomization to usual or enhanced care) is unlikely to explain these findings; more than half of the control (usual care) group received PC and the intervention, adjusted for in the analyses, did not affect HIV risk behavior.[27]

The outcome in this study was not a specific risk behavior. While this choice may present a challenge for interpretation, it is a strength in that a validated scale that represents a range of risk behaviors was used.

Small sample size may have limited our ability to detect statistically significant results in some comparisons. For example, in the HIV drug risk analysis, there were only 22 observations in which subjects received one PC visit. Despite small subgroups and what may appear to be a relatively weak exposure in the context of many other exposures (e.g., to substance use,

other health care), an effect of PC on HIV risk behaviors was detected. The fact that any effect was detected is remarkable as it was not likely that the content of any particular PC visit focused on HIV risk behavior assessment and counseling. Generalizability may be limited to urban, inner city, or high-risk groups; however, these are the populations that would particularly benefit from HIV risk reduction interventions. Furthermore, the severity of HIV risk behavior of subjects in this study was similar to those observed in other study populations with addictions.[35-37] Finally, self-report of PC as defined in this study did not agree perfectly with administrative data. Although concordance was good, neither source is perfect.

While HIV risk reduction interventions in research or clinical settings can decrease HIV risk behaviors, this study suggests that receipt of PC, as it is currently delivered in the community, can have a role in improving HIV risk behavior in persons with addictions. These findings support linkage to PC for such patients and national recommendations that PC clinicians counsel adult and adolescent patients regarding HIV risk behavior. Further research should focus on facilitating referral of patients (especially those with addictions) to PC, incorporating HIV risk reduction counseling into these settings and studying what components of PC are key for improving HIV risk behavior.

REFERENCES

1. Anderson JE, Wilson RW, Barker P, Doll L, Jones TS, Holtgrave D. Prevalence of sexual and drug-related HIV risk behaviors in the U.S. adult population: Results of the 1996 National Household Survey on Drug Abuse. J Acquir Immune Defic Syndr 1999;21(2):148-56.

2. Kwiatkowski C, Booth R. HIV risk behaviors among older American drug users. J Acquir Immune Defic Syndr 2003;33(Supplement):S131-S7.

3. Tyndall MW, Patrick D, Spittal P, Li K, O'Shaughnessy MV, Schechter MT. Risky sexual behaviors among injection drugs users with high HIV prevalence: Implications for STD control. Sex Transm Infect 2002; 78(Supplement 1):i170-i5.

4. HIV prevalence trends in selected populations in the United States: Results from National Serosurveillance, 1993-1997. Available at: www.cdc.gov/hiv/pubs/hivprevalence/toc.htm. Accessed June, 2004.

5. Belcher L, Kalichman S, Topping M, Smith S, Emshoff J, Norris F, et al. A randomized trial of a brief HIV risk reduction counseling intervention for women. J Consult Clin Psychol 1998;66(5):856-61. *doi:10.1037/0022-006X.66.5.856*

6. Kalichman SC, Rompa D, Cage M, DiFonzo K, Simpson, D, Austin J, et al. Effectiveness of an intervention to reduce HIV transmission risks in HIV-positive people. Am J Prev Med 2001;21(2):84-92. *doi:10.1016/S0749-3797(01)00324-5*

7. Kamb ML, Fishbein M, Douglas JM, Rhodes F, Rogers J, Bolan G, et al. Efficacy of risk-reduction counseling to prevent human immunodeficiency virus and sexually transmitted diseases; A randomized controlled trial. JAMA 1998;280(13):1161-7. *doi:10.1001/jama.280.13.1161*

8. Kegeles SM, Hays RB, Pollack LM, Coates TJ. Mobilizing young gay and bisexual men for HIV prevention: a two-community study. AIDS 1999;13(13): 1753-62. *doi:10.1097/00002030-199909100-00020*

9. Raj A, Amaro H, Cranston K, Martin B, Cabral H, Navarro A, et al. Is a general women's health promotion program as effective as an HIV-intensive prevention program in reducing HIV risk among hispanic women? Public Health Rep 2001;116(6):599-607. *doi:10.1093/phr/116.6.599*

10. The NIMH Multisite HIV prevention trial: Reducing HIV sexual risk behavior. Science 1998;280(5371): 1889-94.

11. Avins A, Lindan C, Woods W, Hudes E, Boscarino J, Kay J, et al. Changes in HIV-related behaviors among heterosexual alcoholics following addiction treatment. Drug Alcohol Depend 1997;44:47-55. *doi:10.1016/S0376-8716(96)01321-X*

12. Borsari B, Carey K. Effects of a brief motivational intervention with college student drinkers. J Consult Clin Psychol 2000;68(4):728-33. *doi:10.1037/0022-006X.68.4.728*

13. Des Jarlais DC, Perlis T, Friedman SR, Chapman T, Kwok J, Rockwell R, et al. Behavioral risk reduction in a declining HIV epidemic: Injection drug users in New York City, 1990-1997. Am J Public Health 2000; 90(7):1112-6.

14. Gibson DR, Flynn NM, Perales D. Effectiveness of syringe exchange programs in reducing HIV risk behavior and HIV seroconversion among injecting drug users. AIDS 2001;15:1329-41. *doi:10.1097/00002030-200107270-00002*

15. Gostin LO, Lazzarini ZJ, Jones TS, Flaherty K. Prevention of HIV/AIDS and other blood-borne diseases among injection drug users: A national survey on the regulation of syringes and needles. JAMA 1997; 277(1):53-62. *doi:10.1001/jama.277.1.53*

16. Vlahov D, Junge B, Brookmeyer R, Cohn S, Riley E, Armenian H, et al. Reductions in high-risk drug use behaviors among participants in the Baltimore Needle Exchange Program. J Acquir Immune Defic Syndr Hum Retrovirol 1997;16(5):400-6.

17. Woods WJ, Guydish JR, Sorensen JL, Coutts A, Bostrom A, Acampora A. Changes in HIV-related risk

behaviors following drug abuse treatment. AIDS 1999; 13(15):2151-5. *doi:10.1016/S0749-3797(02)00415-4*

18. Friedmann PD, Zhang Z, Hendrickson J, Stein MD, Gerstein DR. Effect of primary medical care on addiction and medical severity in substance abuse treatment programs. J Gen Intern Med 2003;18(1):1-8. *doi:10.1046/j.1525-1497.2003.10601.x*

19. 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(1):70-8. *doi:10.1111/j.1360-0443.2005.00916.x*

20. 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(1):85-91. *doi:10.1001/archinte.161.1.85*

21. U.S. Preventive Services Task Force. Guide to clinical preventive services. 2nd ed. Baltimore: Williams & Wilkins, 1996.

22. Halpern-Felsher BL, Ozer E, Millstein SG, Wibbelsman CJ, Fuster CD, Elster AB, et al. Preventive services in a health maintenance organization: How well do pediatricians screen and educate adolescent patients? Arch Pediatr Adolesc Med 2000;154(2):173-9.

23. Ramsey PG, Curtis JR, Paauw DS, Carline JD, Wenrich MD. History-taking and preventive medicine skills among primary care physicians: An assessment using standardized patients. Am J Med 1998;104(2): 152-8. *doi:10.1016/S0002-9343(97)00310-0*

24. Wenrich MD, Curtis JR, Carline JD, Paauw DS, Ramsey PG. HIV risk screening in the primary care setting. Assessment of physicians skills. J Gen Intern Med 1997;12(2):107-13. *doi:10.1046/j.1525-1497.1997.00015.x*

25. Boekeloo BO, Schamus LA, Simmens SJ, Cheng TL, O'Connor K, D'Angelo LJ. A STD/HIV prevention trial among adolescents in managed care. Pediatrics 1999;103(1):107-15. *doi:10.1542/peds.103.1.107*

26. Whitlock EP, Orleans CT, Pender N, Allan, J. Evaluating primary care behavioral counseling interventions. Am J Prev Med 2002; 22(4):267-84.

27. Samet JH, Larson MJ, Horton N, Winter M, Saitz R. Linking alcohol- and drug-dependent adults to primary medical care: A randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. Addiction 2003;98:509-16. *doi:10.1046/j.1360-0443.2003.00328.x*

28. McLellan AT. Addiction Severity Index (ASI). In: Carson-Dewitt, editor. Encyclopedia of drugs, alcohol and addictive behavior. Farmington Hills, MI: Macmillan; 2001, pp. 19-21.

29. Ware JE, Sherbourne CD. The MOS36 item short-form health survey: Conceptual framework and item selection. Med Care 1992;30:473-83.

30. Metzger D, Woody GE, Navaline HA, McLellan AT, Meyers K, Boney T, et al. The Risk Assessment Battery (RAB): Validity and reliability. In: Sixth Annual Meeting of National Cooperative Development Group for AIDS; 1993.

31. Metzger DS, Navaline HA, Woody GE. Assessment of substance abuse: HIV Risk Assessment Battery (RAB). Farmington Hills, MI: Macmillan; 2001.

32. Navaline HA, Snider EC, Petro CJ, Tobin D, Metzger D, Alterman AI, et al. 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(Suppl 2):S281-3.

33. Amaro H. Love, sex, and power: Considering women's realities in HIV prevention. Am Psychol 1995; 50(6):437-47. *doi:10.1037/0003-066X.50.6.437*

34. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract 2000;49(2):147-52.

35. Woody GE, Gallop R, Luborsky L, Blaine J, Frank A, Salloum I, et al. HIV risk reduction in the National Institute on Drug Abuse Cocaine Collaborative Treatment Study. J Acquir Immune Defic Syndr 2003; 33:82-7.

36. Brunette MF, Rosenberg SD, Goodman LA, Mueser KT, Osher FC, Vidaver R, et al. HIV risk factors among people with severe mental illness in urban and rural areas. Psychiatr Serv 1999;50(4):556-8.

37. Stein MD, Hanna L, Natarajan R, Clarke J, Marisi M, Sobota M, et al. Alcohol use patterns predict high-risk HIV behaviors among active injection drug users. J Subst Abuse Treat 2000;18:359-63. *doi:10.1016/S0740-5472(99)00070-7*

doi:10.1300/J069v26n03_03

Awareness of Hepatitis C Diagnosis is Associated with Less Alcohol Use Among Persons Co-Infected with HIV

Judith I. Tsui, MD MPH[1], Richard Saitz, MD MPH[2,3], Debbie M. Cheng, ScD[2,4], David Nunes, MD[5], Howard Libman, MD[6], Julie K. Alperen, DrPH[2], and Jeffrey H. Samet, MD MA MPH[2,7]

[1]General Internal Medicine Section (111A1), San Francisco VA Medical Center, 4150 Clement St., San Francisco, CA 94121, USA; [2]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA; [3]Department of Epidemiology and Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA; [4]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; [5]Section of Gastroenterology, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA; [6]Divisions of General Medicine and Primary Care and Infectious Diseases, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA; [7]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA.

BACKGROUND AND OBJECTIVE: It is unknown whether testing HIV-infected individuals for hepatitis C virus (HCV) and informing them of their HCV status impacts subsequent alcohol use. We hypothesized that HIV-infected individuals with current or past alcohol problems who reported being told they had HCV were more likely to 1) abstain from alcohol and 2) not drink *unhealthy* amounts compared to individuals who had not been told.

DESIGN, PARTICIPANTS, AND MEASUREMENTS: Data from a prospective, observational cohort study (HIV-Longitudinal Interrelationships of Viruses and Ethanol) were used to assess the association between awareness of having HCV at baseline and subsequent abstinence and not drinking *unhealthy* amounts as reported at 6-month follow-up intervals. General estimating equations logistic regression was used to account for the correlation from using repeated observations from the same subject over time. We adjusted for age, sex, race, homelessness, injection drug use, depressive symptoms, and having abnormal liver tests.

RESULTS: Participants who reported being told they had HCV were more likely to report abstaining from alcohol (AOR=1.60; 95% CI: 1.13 to 2.27) and not drinking *unhealthy* amounts (AOR=1.46; 95% CI: 1.01 to 2.11).

CONCLUSIONS: Among patients infected with HIV who had a history of alcohol problems, reporting being told one had HCV was associated with greater abstinence from alcohol and less *unhealthy* amounts of drinking.

KEY WORDS: alcohol; hepatitis C; HIV; awareness of diagnosis.
DOI: 10.1007/s11606-007-0147-y

Received August 30, 2006
Revised December 7, 2006
Accepted January 30, 2007
Published online February 23, 2007

BACKGROUND

Because of overlapping risk factors, approximately 30% of individuals with HIV are co-infected with hepatitis C virus (HCV).[1] Alcohol and HIV infection are both associated with more rapid progression of HCV-related liver disease;[2,3] therefore, it is recommended that all patients with HCV, and in particular those who are co-infected with HIV/HCV, abstain from or at least moderate their alcohol use.[4] It is unknown whether patients with HIV who use alcohol change their behavior in response to being diagnosed with HCV.

We examined the effect of reporting being told one had HCV on drinking in a cohort of patients with HIV and a history of alcohol problems, hypothesizing that patients who reported being told they had HCV were more likely to have reduced alcohol consumption compared to individuals who had not been told.

METHODS

Design. Data were from a prospective, observational cohort study (HIV-Longitudinal Interrelationships of Viruses and Ethanol [HIV-LIVE]) in which assessments occurred at 6-month intervals over a maximum of 42 months.

Subjects. Recruitment occurred from a previous cohort study, an intake clinic for HIV-infected patients, HIV primary care and specialty clinics, homeless shelters, drug treatment programs, subject referrals, and flyers. Enrollment occurred between August 2001 and July 2003.

Eligibility criteria were as follows: 1) documented HIV antibody test by ELISA and confirmed by Western blot; 2) 2 or more affirmative responses to the CAGE alcohol screening questionnaire[5,6] or physician investigator diagnosis of alcoholism; 3) ability to speak English or Spanish. Exclusion criteria included: 1) scoring <21 on the 30-item Folstein Mini-Mental State Examination[7,8] and 2) inability to provide informed consent. The Institutional Review Boards of Boston Medical Center and Beth Israel Deaconess Medical Center approved this study.

Table 1. Characteristics of HIV-infected Subjects at Baseline by HCV Awareness

	Never told HCV+ (*n*=189)	Told HCV+ (*n*=211)	*P* value
Age, mean (±SD*)	41 (±8)	44 (±7)	<0.01
Race:			
White	30%	35%	0.15
Black	47%	37%	
Other	23%	28%	
Female	20%	30%	0.02
Recent homelessness**	19%	31%	<0.01
Recent injection drug use**	5%	23%	<0.01
Depressive symptoms†	56%	70%	<0.01
Abnormal liver enzymes‡	47%	80%	<0.01
Detectable HCV RNA	12%	85%	<0.01

*Standard deviation
**Within the past 6 months
†CES-D scale score ≥ 16
‡Alanine aminotranferease (ALT) and aspartate aminotransferase (AST)

Outcomes. The two outcomes of interest were abstinence and not drinking unhealthy amounts in the past 30 days. At each 6-month study visit past-month alcohol consumption was assessed using a validated calendar method (Alcohol timeline followback).[9] Abstinence was defined as no drinks in the past 30 days. Unhealthy drinking amounts were defined as: 1) greater than 14 standard drinks per week, or greater than 4 drinks on occasion for men, or 2) greater than 7 drinks in the past week or greater than 3 drinks on occasion for women.[10]

Independent Variables. The primary independent variable was awareness of HCV diagnosis. This was defined as a positive response to the question "Has a doctor ever told you that you had hepatitis C?" which was queried at study entry only. A negative response could either mean that a patient had been tested and told he/she did not have hepatitis C, or that he/she was never tested.

Additional covariates were self-reported age, sex, race, recent (in the past 6 months) homelessness, recent injection drug use, depressive symptoms, and having abnormal liver enzymes. Depressive symptoms were assessed using the Center for Epidemiologic Studies Depression scale (CES-D),[11] and a threshold of ≥16 was used to define a higher level of depressive symptoms. Secondary analyses used information on chronic infection as determined by HCV RNA levels on polymerase chain reaction (PCR) testing.

Statistical Analysis. Chi-square and Student's *t* tests were used to compare baseline characteristics of subjects who reported being told by a doctor they had hepatitis C compared to those who did not report being told. General estimating equations (GEE) logistic regression was used to calculate odds ratios and 95% confidence intervals, adjusting for other covariates. The GEE approach was used to account for the correlation from using repeated observations from the same subject over time. An exchangeable working correlation structure was used and empirical standard errors are reported for all analyses. Collinearity of covariates was assessed by calculating the correlation between independent variables and no pair of variables had a correlation >0.40. A two-tailed *p* value <0.05 was considered statistically significant for all hypothesis testing. To assess whether the effect of being told was 1) independent from having actual infection and 2) did not differ by actual infected status, we did secondary analyses adjusting for and stratifying by HCV RNA status (detectable vs undetectable). We also performed additional analyses, adjusting for site of recruitment. Analyses were performed using SAS software (version 9.1; SAS Institute, Cary, NC, USA).

RESULTS

The study sample was comprised of 400 participants, of which 211 (53%) reported being told by a physician that they had hepatitis C. The latter patients were more likely to be older, female, homeless, currently using injection drugs, have more depressive symptoms, and have abnormal liver enzymes (Table 1).

At baseline, 64% of patients who were told they had hepatitis C reported abstinence from alcohol versus 52% of patients who were never told they had hepatitis C (*p* value= 0.02). Fewer individuals who had been told they had hepatitis C reported drinking unhealthy amounts (27% told vs 36% never told, *p* value=0.09).

The median follow-up time for patients was 18 months. Reporting being told one had hepatitis C was positively associated with abstinence and not drinking unhealthy amounts in multivariable logistic regression models (Table 2). Participants who reported being told they had hepatitis C had approximately one and a half times the odds of abstaining from alcohol and not drinking unhealthy amounts compared to participants who had not been told they had hepatitis C (AOR for abstinence=1.60; 95% CI: 1.13 to 2.27; AOR for no unhealthy alcohol use=1.46; 95% CI: 1.01 to 2.11). Secondary analyses adjusting for recruitment site produced similar results (data not shown).

Table 2. Adjusted Odds Ratios for Abstinence and No Unhealthy Alcohol Use for HIV-Infected Individuals With Current or Past Alcohol Problems*

	Abstinence (95% CI)	No unhealthy alcohol use (95% CI)
Told HCV Positive	1.60 (1.13, 2.27)	1.46 (1.01, 2.11)
Age	1.01 (0.99, 1.04)	1.04 (1.01, 1.06)
Female	1.83 (1.24, 2.71)	1.51 (1.01, 2.26)
Black race	0.79 (0.54, 1.16)	0.86 (0.57, 1.3)
Other race	1.34 (0.87, 2.05)	1.35 (0.84, 2.17)
Recent homelessness**	0.7 (0.53, 0.92)	0.6 (0.44, 0.82)
Recent injection drug use**	0.88 (0.62, 1.24)	1.01 (0.65, 1.58)
Depressive symptoms†	0.92 (0.73, 1.16)	0.84 (0.65, 1.09)
Abnormal liver enzymes‡	0.86 (0.69, 1.06)	0.77 (0.59, 1.00)

*Results from GEE logistic regression models with adjustment for all variables: analyses based on 1,650 observations and 400 subjects.
**Within the past 6 months
†CES-D scale score ≥ 16
‡Alanine aminotranferease (ALT) and aspartate aminotransferase (AST)

Most (396) of the 400 study subjects had HCV RNA test results available. Of the individuals who had been told they had HCV, 85% had detectable HCV RNA; of those never told they had HCV, 88% had undetectable HCV RNA. Adjusting for HCV RNA did not attenuate the impact of being told of HCV infection on abstinence (AOR=1.59, 95% CI: 0.98 to 2.58) or not drinking unhealthy amounts (AOR=1.56, 95% CI: 0.79 to 3.1), thus it appeared the effect of being told was independent of actual infection. Also, the effect of being told did not differ by status of HCV RNA: odds ratios were similar and confidence intervals overlapped.

DISCUSSION

Among patients infected with HIV who had current or past alcohol problems, reporting being told one had hepatitis C was associated with greater odds of abstaining from alcohol and a lower odds of drinking unhealthy amounts. The results from a secondary analysis suggest that it is being told of one's hepatitis C diagnosis, rather than actually having infection, that is linked to the effect on drinking. This finding is clinically relevant, as it gives indirect support to the hypothesis that telling patients that they have HCV may lead to less unhealthy alcohol use, which in turn can impact long-term outcomes among patients who are co-infected with HIV and HCV.

Our findings appear to be consistent with limited prior research. A study of HIV positive veterans also found a higher rate of abstinence among patients who were co-infected with HCV (based on antibody testing) compared to patients with HIV alone.[12] Studies in other populations not specifically HIV-infected have been mixed, finding either less alcohol use for those aware of their HCV status or no effect.[13–16]

Although the absolute difference in abstinence and unhealthy use between the 2 groups was modest, these differences could nonetheless translate into substantial gains on a population level. Approximately 4 million Americans—1.6% of the population—are believed to be infected with HCV.[17] It is estimated that 5–20% of those infected will develop cirrhosis after 20 years; once diagnosed with cirrhosis, the risk of developing hepatocellular cancer may be 1–4% per year.[4,18] Heavy alcohol use has a synergistic affect with HCV infection and greatly increases the odds of developing cirrhosis and liver cancer.[19,20] Therefore, if screening and notifying individuals of their HCV positive status leads to even modest improvements in alcohol use, a substantial number of individuals might avert liver complications.

This study has several important limitations. We did not assess whether patients received advice about their drinking so we cannot determine whether patients changed their behavior in response to physician counseling. Drinking was based on participants' self-report; despite using a validated tool and assuring confidentiality, patients who were told they had hepatitis C and were aware that they should not be drinking may have been more likely to misrepresent their alcohol use (social desirability bias). Finally, although we adjusted for confounders, there may have been residual confounding.

Among a cohort of HIV-infected individuals with current or past alcohol problems, reporting being told that one had hepatitis C was associated with greater abstinence and less unhealthy amounts of drinking. Testing HIV-infected patients for HCV and informing them of their status may lead them to drink less, and thus reduce their likelihood of developing HCV-associated liver disease.

Acknowledgments: *The authors appreciate the contribution of Vincent Faber for data management assistance. Support for this study came from the following grants from the National Institute of Alcohol Abuse and Alcoholism (NIAAA) of the NIH: R01-AA13766, R01-AA11785, R01-AA10870, and K24 AA015674. This research was conducted in part in the General Clinical Research Center at Boston University School of Medicine, USPHS Grant M01 RR00533 and the Clinical Research Center at Beth Israel Deaconess Medical Center, USPHS Grant M01 RR01032. Support for Dr. Tsui comes from Grant Number KL2 RR024130 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) and NIH Roadmap for Medical Research. This paper was presented at the 29th Society of General Internal Medicine Annual Meeting in April 2006.*

Conflict of interest: *None disclosed.*

Corresponding Author: *Judith I. Tsui, MD, MPH; General Internal Medicine Section (111A1), San Francisco VA Medical Center, 4150 Clement St., San Francisco, CA 94121, USA (e-mail: Judith.Tsui@ucsf.edu).*

REFERENCES

1. **Alter MJ.** Epidemiology of viral hepatitis and HIV co-infection. J Hepatol. 2006;44:S6–9.
2. **Peters MG, Terrault NA.** Alcohol use and hepatitis C. Hepatology. 2002;36:S220–5.
3. **Sulkowski MS, Thomas DL.** Hepatitis C in the HIV-infected person. Ann Intern Med. 2003;138:197–207.
4. Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. Centers for Disease Control and Prevention. Morb Mort Wkly Rep Recomm Rep. 1998;47:1–39.
5. **Mayfield D, McLeod G, Hall P.** The CAGE questionnaire: validation of a new alcoholism screening instrument. Am J Psychiatry. 1974;131:1121–3.
6. **Samet JH, Phillips SJ, Horton NJ, Traphagen ET, Freedberg KA.** Detecting alcohol problems in HIV-infected patients: use of the CAGE questionnaire. AIDS Res Hum Retroviruses. 2004;20:151–5.
7. **Folstein M, Folstein S, McHugh P.** "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.
8. **Smith KL, Horton NJ, Saitz R, Samet JH.** The use of the mini-mental state examination in recruitment for substance abuse research studies. Drug Alcohol Depend. 2006;82:231–7.
9. **Sobell L, Sobell M.** Alcohol Timeline Followback. In: Association AP, ed. Handbook of Psychiatric Measures. Washington, DC; 1996:477–9.
10. **Saitz R.** Clinical practice. Unhealthy alcohol use. N Engl J Med. 2005;352:596–607.
11. **Radloff L.** The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas. 1977;1:385–401.
12. **Conigliaro J, Madenwald T, Bryant K, et al.** The Veterans Aging Cohort Study: observational studies of alcohol use, abuse, and outcomes among human immunodeficiency virus-infected veterans. Alcohol Clin Exp Res. 2004;28:313–21.
13. **Kwiatkowski CF, Fortuin Corsi K, Booth RE.** The association between knowledge of hepatitis C virus status and risk behaviors in injection drug users. Addiction. 2002;97:1289–94.
14. **McCusker M.**Influence of hepatitis C status on alcohol consumption in opiate users in treatment. Addiction. 2001;96:1007–14.
15. **Ompad DC, Fuller CM, Vlahov D, Thomas D, Strathdee SA.** Lack of behavior change after disclosure of hepatitis C virus infection among

young injection drug users in Baltimore, Maryland. Clin Infect Dis. 2002;35:783–8.

16. **Nalpas B, Martin S, Fontaine H, Fabbro-Peray P, Brechot C, Pol S.** Impact of medical recommendations on alcohol consumption in HCV positive patients. J Hepatol. 2001;35:312–3.
17. **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.
18. **Seeff LB.** Natural history of chronic hepatitis C. Hepatology. 2002;36: S35–46.
19. **Corrao G, Arico S.** Independent and combined action of hepatitis C virus infection and alcohol consumption on the risk of symptomatic liver cirrhosis. Hepatology. 1998;27:914–9.
20. **Donato F, Tagger A, Gelatti U, et al.** Alcohol and hepatocellular carcinoma: the effect of lifetime intake and hepatitis virus infections in men and women. Am J Epidemiol. 2002; 155:323–31.

The American Journal of Bioethics, 7(11): 1–6, 2007

ISSN: 1526-5161 print / 1536-0075 online
DOI: 10.1080/15265160701638520

Target Article

Health Literacy, Health Inequality, and a Just Healthcare System

Angelo E. Volandes, Center for Bioethics, Brigham & Women's Hospital
Michael K. Paasche-Orlow, Boston University School of Medicine

Limited health literacy is a pervasive and independent risk factor for poor health outcomes. Despite decades of reports exhibiting that the healthcare system is overly complex, unneeded complexity remains commonplace and endangers the lives of patients, especially those with limited health literacy. In this article, we define *health literacy* and describe the empirical evidence associating health literacy and poor health outcomes. We recast the issue of poor health literacy from within the ethical perspective of the least well-off and argue that poor health outcomes deriving from limited health literacy ought to be understood as a fundamental injustice of the healthcare system. We offer three proposals that attempt to rectify this injustice, including: universal precautions that presume limited health literacy for all healthcare users; expanded use of technology supported communication; and clinical incentives that account for limited health literacy.

Keywords: Health Literacy; health disparities; justice; health policy; health outcomes; pay for performance

Approximately 90 million American adults lack the literacy skills needed to use the healthcare system (Nielsen-Bohlman et al. 2004). The prevalence of limited literacy is particularly high among those with lower levels of education, the elderly, minorities, and those with chronic disease (Paasche-Orlow et al. 2005b). An emerging literature has begun to describe the myriad health consequences of limited literacy (DeWalt et al. 2004). Indeed, limited literacy has been shown to be an independent risk factor for worse health status, hospitalization and mortality (DeWalt et al. 2004; Sudore et al. 2006b; Wolf 2006a). Though more than 300 articles in the medical literature exhibit ways healthcare is overly complex for people with limited literacy; the recent shift toward shared decision-making, consumer-oriented healthcare, and programs such as Medicare Part D are evidence of increasing complexity (Parker et al. 2003). Despite the clear injustice of a healthcare system that is organized for the most literate and powerful members of our society, the medical ethics literature has neglected some of our most vulnerable patients by remaining largely quiet about the ethical implications of health literacy.

In this article, we will define *health literacy*, a concept closely linked to literacy, and briefly review the empirical evidence for the association between health literacy and poor health outcomes. Then we will explore the moral implications of limited health literacy for a healthcare system based on justice using one influential criterion, the position of the least well-off. Finally, we will offer three proposals to address this basic injustice. Our hope is that these ideas will bring greater attention to a significant limitation in the current healthcare system and that instigating a far-reaching discourse on the moral implications of the emerging literature on health literacy may help lead to amelioration of this source of basic inequalities in health.

HEALTH LITERACY AND HEALTH OUTCOMES

Health literacy is the "degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions" (Nielsen-Bohlman et al. 2004, X). While basic literacy skills, such as reading, arithmetic, and using documents are core attributes of the broader concept, health literacy also includes skills such as: how to use the healthcare system; having cognitive dimensions such as memory and attention; and, displaying adequate neurosensory capacities such as vision and hearing (Baker 2006). Another key feature of the definition is that health literacy highlights the contextual demands placed on the individual. Many people who experience literacy barriers would not experience such limitations if the healthcare system were to be streamlined, simplified, and standardized (Paasche-Orlow et al. 2006).

Health literacy may be a critical and underexamined mechanism of health inequalities (Saha 2006; Sentell and Halpin 2006). Limited health literacy has been shown to be a more powerful predictor of health status and health-related behaviors than race or education (Baker et al. 1997; Bennett et al. 1998; Scott et al. 2002; Williams et al. 1998a; Williams et al. 1998b). Furthermore, limited health literacy has been shown to be an independent risk factor of worse outcomes and health disparities independent of race and education

Acknowledgment: We thank Muriel Gillick for commenting on an earlier draft of this article.
Address correspondence to Angelo E. Volandes, Center for Bioethics, Brigham & Women's Hospital, 1620 Tremont Street, Boston, MA 02220. E-mail: avolandes@partners.org

(Baker et al. 1998; Baker et al. 2002; Paasche-Orlow et al. 2005a; Sudore et al. 2006b; Wolf et al. 2006b).

Limited health literacy is associated with low health knowledge, increased incidence of chronic illness, poorer intermediate disease markers, and less than optimal use of preventive health services (Berkman et al. 2004). Numerous studies have exhibited the association between limited health literacy and health outcomes; in two separate studies, limited health literacy has been linked with mortality (DeWalt et al. 2004; Sudore et al. 2006b; Wolf et al. 2006a) For example, in a community sample of 2,512 men and women age 70 to 79 years, Sudore et al. (2006b) found that, over the course of a mean follow-up of 5.1 years, those with limited health literacy had a higher risk of death than those with adequate literacy (19.7% versus 10.6%, respectively). In other words, the rate of death for those with limited health literacy was approximately twice the rate of death among those with adequate literacy (hazard ratio of 2.03; 95% confidence interval 1.62 to 2.55).

Furthermore, in the same study (Sudore et al. 2006b), after adjusting for age, race, gender, income, education, health status, health behaviors, health access, and psychosocial status, health literacy remained an independent risk factor for all-cause mortality with a hazard ratio of 1.75 (95% confidence interval 1.27 to 2.41). To put the size of this effect into context, Malik et al. (2004) found that having diabetes conferred a 1.97 (95% confidence interval 1.59 to 2.43) adjusted hazard ratio for all-cause mortality among 6,255 subjects with an average of 13.4 years of follow-up in the Second National Health and Nutrition Examination Survey (NHANES II), i.e., the health impact of having limited health literacy is on the order of having diabetes.

Justice and Healthcare

Previous critiques of the high literacy demands of the healthcare system have focused primarily on the issue of autonomy. For example, attention has been paid to the overwhelming complexity of informed consent forms and efforts have been made to simplify the consent process in ways that are successful with patients who have limited literacy skills (Paasche-Orlow et al. 2003; Sudore et al. 2006a). The autonomy of healthcare users with limited literacy is thwarted if the forms intended to preserve their individual autonomy are inaccessible.

While concern for patient autonomy is well placed, a more thorough evaluation of the moral consequences of limited health literacy can be assessed through an analysis of the requirements of justice. The complexity of the healthcare system negatively affects the health of people with limited literacy, and much of this complexity is unneeded. The lens of justice helps focus the discussion on how this unneeded complexity and the general assumption of high literacy skills influence health and promote racial and ethnic health disparities. While it is inappropriate to ignore patients' values and preferences and cynical to ignore opportunities to advance patient autonomy by improving patient education materials, consent forms, notices of privacy protection (Paasche-Orlow et al. 2005c), patient's bill of rights documents (Paasche-Orlow 2006), and the like, we believe the problem of limited health literacy should primarily be understood as an issue of health inequality and justice.

One widely used criterion for examining the role of justice in health inequalities is the position of the least well-off in a healthcare system (Daniels 1985; Rawls 1971). Presently, the healthcare system places little importance on the position of the least well-off in the context of health literacy. The average English reading level for American adults is between the eighth and ninth grade; the average reading level of many documents designed for patients, including education materials, explanations of benefits and services, and documents outlining patients' rights such as informed consent forms, notices of privacy protection, and advanced directives documentation, continue to be written at the senior high-school level or higher (Paasche-Orlow et al. 2003; Paasche-Orlow et al. 2005c; Paasche-Orlow 2006). Despite the fact that one in three users of the healthcare system has limited literacy, a high level of literacy is typically assumed and the healthcare system remains fit only for users with the highest literacy.

The primacy and importance of the position of the least well-off in any healthcare system has been an influential criterion for evaluating the justice of a healthcare system, especially in theories of distributive justice. The position of the least well-off more generally has been a common thread in numerous religious philosophies and most closely associated with the ethical theory of prioritarianism (Parfit 1991). Previous uses of this principle in the healthcare context stem from the widely influential maximin principle found in the philosophical and economics literature, most notably in the work of the philosopher John Rawls (1971). In Rawls' framework, decision-makers, who are behind a veil of ignorance and unaware of their positions in a society, would design a system in which the position of the least well-off is maximized regardless of the potentially negative impact on those better off (Rawls 1971). In the healthcare context, decision-makers using the maximin principle would design a system that places the least fortunate healthcare users in the least unfortunate situation. For decision-makers considering the least well-off in terms of health literacy, the most just arrangement would be one that ensured that the healthcare system was designed to benefit users with limited literacy.

The reorganization of healthcare to fulfill the prioritarian objectives raised in discussions of health literacy would involve the elevation of patient education to be a core function of health providers. Instead of assuming literacy and then trying to retrofit care for low literacy patients as some form of specialty service, application of the maximin principle leads us to the conclusion that the standard of care should be reoriented to the needs of health consumers with limited literacy. In the remaining sections, we will explore three examples of how efforts in clinical care and health policy can help shift the standard of care to make the needs of patients with limited literacy a priority and begin to

address the fundamental injustice arising from an unnecessarily complex healthcare system.

UNIVERSAL PRECAUTIONS FOR LIMITED HEALTH LITERACY

Presently, the standard of care or default position of physicians commencing discussions with patients is to assume a high level of literacy. Limited health literacy is some kind of exception to the rule, and patient education is contingent on a patient asking for more information. Indeed, physicians rarely evaluate comprehension (Schillinger et al. 2003; Schillinger et al. 2004). Patients are trained to say that they understand everything clearly when clinicians inquire about questions while rising from a chair with one hand on the doorknob. Similarly, patients with limited literacy often conceal their lack of understanding to avoid shame (Parikh et al. 1996). Consequently, clinicians frequently fail to evaluate comprehension adequately and rarely take the time to teach their patients until they can confirm that critical self-management skills are understood (Schillinger et al. 2003). This approach is detrimental to patients. A standard of care that assumes a high level of literacy should be thought of as a risky venture.

We suggest a change in the standard of care — a 'flipping' of the default — in which the universal assumption is limited literacy. The key added activity for the healthcare team would be to confirm comprehension of the clinical plan with every patient at multiple steps of the care continuum. This confirmation can be done with a "teach-back" model in which each patient is asked to state or exhibit the care plan so the healthcare team can provide additional education focused on items the patient has not understood until the patient is able to confirm comprehension (Schillinger et al. 2003). Ongoing education until a learner can confirm comprehension is known as a "teach-to-goal" or "teach-to-mastery" educational technique and has been shown to be effective (Paasche-Orlow et al. 2005a; Sudore et al. 2006a). For example, a study of confirming comprehension of the medication regimen and inhaler technique among patients with asthma revealed that: 1) patients with limited literacy had worse comprehension at baseline; 2) all patients were able to exhibit mastery after the teach-to-goal process; and 3) comprehension at 2 weeks was similar for patients with and without limited literacy (Paasche-Orlow et al. 2005a).

Available evidence suggests that confirming patient comprehension is worthwhile and would decrease the self-management deficits associated with limited literacy. Under current conventions, the amount of information given to patients seems to be linked to a clinician's judgment of the complexity of the information, the severity of the risks, or other medicolegal attributes of healthcare. We want patient education to become a central priority of healthcare and think that it is appropriate to hold clinicians accountable for educating patients all the time, even for an illness as minor as the common cold.

Some may argue that this proposal is impractical due to the time and resources it would consume. Ultimately, questions about the strict utility of such investments in patient education are empirical questions of the cost-benefit analyses that would need to be evaluated in various settings. We hold that such cost-benefit analyses will need to be judged within the prioritarian framework we have laid out, i.e., that the cost of organizing healthcare teams to teach-to-goal will appropriately provide more benefit for the patients with the highest risk of worse health outcomes.

Furthermore, all learners have better comprehension and retention of information when presented with clear, succinct messages that are reiterated (Coyne et al. 2003; Flory and Emanuel 2004). While simple communication and confirming comprehension will help patients with limited literacy more than it will help patients without limited literacy, there is little downside in reorienting the system for limited literacy. Not only will patients' self-management skills improve, but also, patients may learn that that their providers really care about the specific tasks, which may promote adherence.

TECHNOLOGY-SUPPORTED COMMUNICATION

Our second proposal calls for developing and implementing communication technology models that promote meaningful communication and are successful for people with limited health literacy. The present healthcare system was designed by physicians who fashioned a medical system that works best for users who think, talk, and act like physicians. The dominant iatroculture assumes the average healthcare user reads, writes, and thinks at a highly sophisticated level in which the spoken and written word are used to communicate complex medical and legal information.

The call for improved communication in medicine is hardly new. The medical literature is replete with studies showing the benefits of plain simplified language in the medical setting. This movement also calls attention to needed improvements in training healthcare providers in communication and patient education as well as the need to invest in a multilevel team approach to educate patients, especially in the context of chronic disease where improvement in outcomes is closely linked to self-management (Paasche-Orlow et al. 2006).

However, the complex ideas that users of the healthcare system must understand, such as advance care planning, medication regimens, and medical decision-making, may not be optimally explained simply with verbal or written communication. These complex ideas can be facilitated by pictures, video, multimedia and other decision aids beyond the written and spoken word (Barry 2002; Woolf et al. 2005). We have successfully piloted a video to help patients with limited health literacy better envision and understand hypothetical health states such as dementia (Volandes et al. 2006). Other health information technologies, such as automated phone systems (Friedman et al. 1996; Migneault et al. 2006; Rubin et al. 2006), and automated computer interfaces such as animated embodied-conversational agents and "talking touch-screens" (Bickmore et al. 2005; Bickmore and Giorgino 2006; Hahn et al. 2004), are currently being

refined and tested in intervention studies for patients with limited health literacy.

Although there has been a slow rise in the number of articles promoting innovative media, there needs to be a concerted effort on the part of the healthcare system to invest and develop technologies that better promote meaningful communication. Care is needed, however, in how new communication technologies are implemented. Patients with limited health literacy may have limited access to newer technologies. While well designed and tested systems with simplified user interfaces have the potential to improve healthcare for people with limited health literacy, failure to ensure access to all and a lack of focus on how such systems will be used by people with limited health literacy could result in health information technologies that exacerbate inequalities in the healthcare system.

Some may argue that developing technology platforms to promote meaningful communication is too expensive in an already overly expensive healthcare system that is approaching 15% of the United States Gross Domestic Product. However, the rapidly decreasing price of digital technology and the increasing availability of web-based informational systems make the implementation and development of new technology feasible. Ideally, this technology should be tested for efficacy and cost/benefit but there is reason to think that technology platforms may work as well or surpass healthcare providers. Technology platforms are highly efficient and, unlike physicians, deliver what they are designed to deliver with high fidelity. Furthermore, such systems may be used to collect, store, and compare data, support self-management, as well as promote and facilitate communication between patients and their healthcare team. Considerations of the least well-off in the healthcare system will require creativity in alternative communication and patient education techniques that have been proven to be effective with patients who have limited health literacy. We are in a creative era of exploration in health information technology. Ultimately, while the risk of unequal access is real and the difficulty people have interacting with new technologies needs to be overcome, health information technologies provide a real opportunity to advance patient education for patients with limited health literacy.

A MORE JUST PAY-FOR-PERFORMANCE

Payors have begun to use healthcare process measures to influence physicians' reimbursement in programs loosely referred to as *pay-for-performance*. This trend should be reassessed through the perspectives of health literacy and the least well-off. Currently, such financial reimbursement schemes are 'one size fits all' with respect to patient literacy and for the most part, patients with limited literacy are at a distinct disadvantage compared with high literacy patients through the lens of these various evaluative processes.

The typical pay-for-performance arrangement withholds partial compensation and then offers providers bonuses for appropriately measuring specific laboratory tests (e.g., cholesterol test), conducting cancer screening (e.g., mammography), investing in electronic medical records, and prescribing generic medications. For example, providers will get additional compensation per member per month if they meet or exceed the pre-set target for the percentage of patients with diabetes who have had a hemoglobin A1C (HbA_{1c}) test at least twice annually or the percentage of patients older than age 50 years who have undergone colorectal cancer screening (Rosenthal et al. 2005; Rosenthal and Frank 2006).

The goal is to improve quality. However, quality is defined in a manner that is driven by data that are easily derived from administrative databases, such as HbA_{1c} blood tests or colonoscopies. Patient education is not coded, so it cannot be promoted by current programs. Furthermore, patients who refuse a test are not removed from their clinician's score, i.e., clinicians must convince their patients to undergo colonoscopy. This design will lead clinicians away from patient education and toward marketing whatever targets are currently on the scorecard. Notwithstanding the ways this arrangement generally undercuts informed consent, patients with limited health literacy, who need more education and who are not as likely to use preventive services, will be seen more negatively as a result of pay-for-performance (Scott et al. 2002).

Some pay-for-performance plans categorize providers according to how well they adhere to the targets and link subscriber co-payments to the pay-for-performance tier of their primary care provider. In such a scenario, a provider might learn to view patients with limited health literacy as a professional liability that could cost all other patients higher co-pays as well as give the provider a worse reputation. Providers may attempt to evade penalty by avoiding patients who have lower rates of adherence to the pay-for-performance targets.

Current pay-for-performance programs have been shown to accrue financial benefit only to providers who do well on such measures prior to initiating the program (Rosenthal et al. 2005). To level the playing field for the least fortunate users of the healthcare system, we must be willing to invest disproportionately more resources on patients with limited literacy. To accomplish this, resources must be allocated specifically to healthcare settings that care for a disproportionate share of patients with limited literacy to hire diabetes educators, nutritionists, clinical pharmacologists, and other practitioners who are dedicated to patient education.

It is important to note that current programs that get providers to measure HbA_{1c} tests have not been shown to lead to improvement in the actual values of HbA_{1c} (O'Connor et al. 2005). This finding can be seen to show how some of the quality measures place an inordinate focus on the wrong kinds of behavior. A shift of quality measures away from items such as laboratory tests toward patient education may provide incentives to interact with patients in a way that will have a better chance of improving outcomes.

CONCLUSION

The empirical evidence is clear: limited health literacy negatively impacts health. We have presented three ways clinical care and health policy can be transformed to address the needs of patients with limited literacy. We hope that these ideas can be evaluated on their merits and that they might serve as a catalyst to introduce a broad discourse on unneeded complexity in the healthcare system. This is important because, in failing to meet the needs of patients with limited literacy, the healthcare system propagates a serious injustice. While the healthcare system may not rectify inequalities that transcend healthcare, such as inequalities in primary education, we should expect that the healthcare system not exacerbate underlying injustices. Unfortunately, this is precisely the impact of the current arrangement, as limited literacy has been shown to be an underlying mechanism of ethnic and racial health disparities (Saha 2006; Sentell and Halpin 2006).

Future attention and research focusing on the problem of unneeded complexity in healthcare may help ameliorate the clustering of disadvantage represented by those with limited health literacy, a population that is disproportionately elderly, non-white, less educated, and chronically ill. As Powers and Faden (2006) point out, "inequalities are interactive" and to pursue the remedial task of aspiring for justice in the real world we need to identify through empirical methods how "inequalities of one kind beget inequalities of another kind " (X). The emerging literature on health literacy is providing such evidence. Not to focus attention on patients with limited literacy is to neglect a fundamental obligation of the healthcare system to its most vulnerable constituents.

REFERENCES

Baker, D. W. 2006. The meaning and the measure of health literacy. *Journal of General Internal Medicine* 21(8): 878–883.

Baker, D. W., J. A. Gazmarian, M. V. Williams, et al. 2002. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. *American Journal of Public Health* 92(8): 1278–1283.

Baker, D. W., R. M. Parker, M. V. Williams, and W.S. Clark. 1998. Health literacy and the risk of hospital admission. *Journal of General Internal Medicine* 13(12): 791–798.

Baker, D. W., R. M. Parker, M. V. Williams, W. S. Clark, and J. Nurss. 1997. The relationship of patient reading ability to self-reported health and use of health services. *American Journal of Public Health* 87(6): 1027–1030.

Barry, M. J. 2002. Health decision aids to facilitate shared decision making in office practice. *Annals of Internal Medicine* 136(2): 127–135.

Bennett, C. L., M. R. Ferreira, T. C. Davis, et al. 1998. Relation between literacy, race, and stage of presentation among low-income patients with prostate cancer. *Journal of Clinical Oncology* 16(9): 3101–3104.

Berkman N. D., D. A. DeWalt, M. P. Pignone, et al. 2004. Literacy and health outcomes. Evidence report/technology assessment no. 87. In *Agency for Healthcare Research and Quality Publication No. 04-E007-2.* Rockville, MD: Agency for Healthcare Research and Quality.

Bickmore, T., and T. Giorgino. 2006. Health dialog systems for patients and consumers. *Journal of Biomedical Informatics* 39(5): 556–571.

Bickmore, T., A. Gruber, and R. Picard. 2005. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Education and Counseling* 59(1): 21–30.

Coyne, C. A., R. Xu, P. Raich, et al. 2003. Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: a study of the Eastern Cooperative Oncology Group. *Journal of Clinical Oncology* 21(5): 836–842.

Daniels, N. 1985. *Just Health Care.* Cambridge: Cambridge University Press.

DeWalt, D. A., N. D. Berkman, S. Sheridan, K. N. Lohr, and M. P. Pignone. 2004. Literacy and health outcomes: a systematic review of the literature. *Journal of General Internal Medicine* 19(12): 1228–1239.

Flory, J., and E. Emanuel. 2004. Interventions to improve research participants' understanding in informed consent for research: a systematic review. *Journal of the American Medical Association* 292(13): 1593–1601.

Friedman, R. H., L. E. Kazis, A. Jette, et al. 1996. A telecommunications system for monitoring and counseling patients with hypertension. Impact on medication adherence and blood pressure control. *American Journal of Hypertension* 9(4 Pt 1): 285–292.

Hahn, E. A., D. Cella, D. Dobrez, et al. 2004. The talking touchscreen: a new approach to outcomes assessment in low literacy. *Psychooncology* 13(2): 86–95.

Malik, S., N. D. Wong, S. S. Franklin, G. J. L'Italien, J. R. Pio, and G. R. Williams. 2004. Impact of the metabolic syndrome on mortality from coronary heart disease, cardiovascular disease, and all causes in United States adults. *Circulation* 110(10): 1245–1250.

Migneault, J. P., R. Farzanfar, J. A. Wright, and R. H. Friedman. 2006. How to write health dialog for a talking computer. *Journal of Biomedical Informatics* 39(5): 468–481.

Nielsen-Bohlman, L. T., A. M. Panzer, B. Kindig, and A. David. 2004. *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.

O'Connor, P. J., J. Desai, L. I. Solberg, et al. 2005. Randomized trial of quality improvement intervention to improve diabetes care in primary care settings. *Diabetes Care* 28(8): 1890–1897.

Paasche-Orlow, M. K. 2006. National survey of states' patients' bill of rights. *Journal of General Internal Medicine* 21(S4): 90.

Paasche-Orlow, M. K., D. Schillinger, S. M. Greene, and E. H. Wagner.2006. How health care systems can begin to address the challenge of limited literacy. *Journal of General Internal Medicine* 21(8): 884–887.

Paasche-Orlow, M. K., D. M. Jacob, and J. N. Powell. 2005a. Notices of privacy practices: a survey of the health insurance portability

and accountability act of 1996 documents presented to patients at US hospitals. *Medical Care* 43(6): 558–564.

Paasche-Orlow, M. K., R. M. Parker, J. A. Gazmararian, L. T. Nielsen-Bohlman, and R. R. Rudd. 2005b. The prevalence of limited health literacy. *Journal of General Internal Medicine* 20(2): 175–184.

Paasche-Orlow, M. K., K. A. Riekert, A. Bilderback, et al. 2005c. Tailored education may reduce health literacy disparities in asthma self-management. *American Journal of Respiratory Care* 172(8): 980–986.

Paasche-Orlow, M. K., H. A. Taylor, and F. L. Brancati. 2003. Readability standards for informed-consent forms as compared with actual readability. *New England Journal of Medicine* 348(8): 721–726.

Parfit, D. 1991. *Equality or priority?* Lawrence,KS: University of Kansas Press.

Parikh, N. S., R. M. Parker, J. R. Nurss, D. W. Baker, and M. V. Williams. 1996. Shame and health literacy: the unspoken connection. *Patient Education and Counseling* 27(1): 33–39.

Parker, R. M., S. C. Ratzan, and N. Lurie. 2003. Health literacy: a policy challenge for advancing high-quality health care. *Health Affairs* 22(4): 147–153.

Powers, M., and R. Faden. 2006. *Social justice: the moral foundations of public health and health policy*. New York, NY: Oxford University Press.

Rawls, J. 1971. *A theory of justice*. Cambridge, MA: Harvard University Press.

Rosenthal, M. B., and R. G. Frank. 2006. What is the empirical basis for paying for quality in health care? *Medical Care and Research Review* 63(2): 135–157.

Rosenthal, M. B., R. G. Frank, Z. Li, and A. M. Epstein. 2005. Early experience with pay-for-performance: from concept to practice. *Journal of the American Medical Association* 294(14): 1788–1793.

Rubin, A., J. P. Migneault, L. Marks, E. Goldstein, K. Ludena, and R. H. Friedman. 2006. Automated telephone screening for problem drinking. *Journal of Studies on Alcohol* 67(3): 454–457.

Saha, S. 2006. Improving literacy as a means to reducing health disparities. *Journal of General Internal Medicine* 21(8): 893–895.

Schillinger, D., A. Bindman, F. Wang, A. Stewart, and J. Piette. 2004. Functional health literacy and the quality of physician-patient communication among diabetes patients. *Patient Education and Counseling* 52(3): 315–323.

Schillinger, D., J. Piette, K. Grumbach, et al. 2003. Closing the loop: physician communication with diabetic patients who have low health literacy. *Archives of Internal Medicine* 163(1): 83–90.

Scott, T. L., J. A. Gazmarian, M. V. Williams, and D. W. Baker. 2002. Health literacy and preventive health care use among Medicare enrollees in a managed care organization. *Medical Care* 40(5): 395–404.

Sentell, T. L., and H. A. Halpin. 2006. The Importance of adult literacy in understanding health disparities. *Journal of General Internal Medicine* 21(8): 862–867.

Sudore, R. L., C. S. Landefeld, B. A. Williams, D. E. Barnes, K. Lindquist, and D. Schillinger. 2006a. Use of a modified informed consent process among vulnerable patients: a descriptive study. *Journal of General Internal Medicine* 21(8): 867–873.

Sudore, R. L., K. Yaffe, S. Satterfield, et al. 2006b. Limited literacy and mortality in the elderly: the health, aging, and body composition study. *Journal of General Internal Medicine* 21(8): 806–812.

Volandes A, L. Lehmann, E. F. Cook, S. Shaykevich, E. D. Abbo, and M. R. Gillick. 2006. Using video images of dementia in advance care planning [abstract]. *Society for Medical Decision Making*. Q5

Williams, M. V., D. W. Baker, E. G. Honig, T. M. Lee, and A. 1998a. *Nowlan. Inadequate literacy is a barrier to asthma knowledge and self-care. Chest* 114(4): 1008–1015.

Williams, M. V., D. W. Baker, R. M. Parker, and J. R. Nurss. 1998b. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. *Archives Internal Medicine* 158(2): 166–172.

Wolf, M. S. 2006a. Literacy and mortality among medicare enrollees. *Journal General Internal Medicine* 21(S4): 81.

Wolf, M. S., S .J. Knight, E. A. Lyons, R. Durazo-Arvizu, S. A. Pickard, A. Arseven, A. Arozullah, K. Colella, P. Ray, and C. L. Bennett. 2006b. Literacy, race, and PSA level among low-income men newly diagnosed with prostate cancer. *Urology* 68(1): 89–93.

Woolf, S. H., E. C. Chan, R. Harris, S. L. Sheridan, C. H. Braddock, R. M. Kaplan, A. Krist, A. M. O'Connor, and S. Tunis. 2005. Promoting informed choice: transforming health care to dispense knowledge for decision making. *Annals of Internal Medicine* 143(4): 293–300.

Do Brief Measures of Readiness to Change Predict Alcohol Consumption and Consequences in Primary Care Patients With Unhealthy Alcohol Use?

Emily C. Williams, Nicholas J. Horton, Jeffrey H. Samet, and Richard Saitz

Background: Assessing readiness to change is recommended as part of brief interventions for patients with unhealthy alcohol use. However, the utility and predictive validity of readiness measures have not been well established.

Methods: In a prospective cohort study, we assessed primary care patients with unhealthy alcohol use (past-month drinking of risky amounts, or any amount and an affirmative response to CAGE alcohol screening questionnaire) and reassessed them 6 months later. At study entry, we assessed readiness to change using 1 multi-item measure of stage of change, and 5 single-item measures (readiness per se, importance of changing, confidence in ability to change, intention to cut down, intention to abstain). Outcomes included alcohol consumption and alcohol-related consequences. Multivariable regression models were fit for each measure of readiness and each outcome.

Results: Of 312 patients with unhealthy alcohol use, 228 (73%) were assessed at study entry and 6 months later and had complete data. Among readiness measures, only confidence and intention to abstain (1 point changes on single-item measures) were associated with consumption 6 months later: less heavy episodic drinking [adjusted odds ratio (AOR) 0.88, 95% CI 0.80–0.98 and AOR 0.79, 0.64–0.98, respectively], and less drinking of risky amounts (AOR 0.89, 0.79–1.00 and AOR 0.78, 0.62–0.98, respectively). Intention to abstain was also associated with more abstinence (AOR 1.43, 1.09–1.88). Single-item measures of readiness, importance, and intention to cut down were significantly associated with higher odds of alcohol consequences. Greater confidence (single item) was associated with a lower odds of any consequences (AOR 0.88, 0.79–0.98).

Conclusions: Greater readiness, as measured by several brief assessments, was associated with more consequences and was not predictive of consumption. However, assessing confidence in the ability to change one's alcohol use may have a role in predicting subsequent decreases in both consumption and consequences in primary care patients.

Key Words: Alcohol, Alcohol Use, Readiness to Change.

From the Health Services Research & Development, VA Puget Sound Health Care System, Seattle, Washington (ECW); the Center of Excellence in Substance Abuse Treatment and Education, VA Puget Sound Health Care System, Seattle, Washington (ECW); The Departments of Medicine and Health Services, the Graduate School of Public Health, University of Washington, Seattle, Washington (ECW); the Department of Mathematics and Statistics, Smith College, Northampton, Massachusetts (NJH); the Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, Massachusetts (JHS, RS); Youth Alcohol Prevention Center and Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts (RS); the Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts (JHS).

Received for publication January 17, 2006; accepted November 28, 2006.

The views expressed in this article are those of the authors and do not necessarily represent the views of their affiliated institutions.

The preliminary results of this study were presented in abstract form at the Society of General Internal Medicine annual meeting, Chicago, IL, May 14, 2004.

Dr. Saitz received support for this study from the Robert Wood Johnson Foundation as a Generalist Physician Faculty Scholar.

Ms. Williams received support for this publication from the University of Washington School of Public Health.

Reprint requests: Emily C. Williams, MPH, VA Puget Sound Health Care System, 1100 Olive Way, Suite 1400, Seattle, WA 98101; Fax: 206-764-2935; E-mail: emwilli@u.washington.edu

DOI: 10.1111/j.1530-0277.2006.00324.x

UNHEALTHY ALCOHOL USE is common and accounts for substantial morbidity and mortality (Saitz, 2005). Primary care visits are an opportune time to detect and prevent unhealthy alcohol use and related consequences (Barnes and Samet, 1997). Brief counseling interventions in primary care settings reduce drinking in patients with unhealthy alcohol use identified by screening (Ballesteros, et al., 2004; Bertholet et al., 2005; Bien et al., 1993; Kahan et al., 1995; Moyer et al., 2002; Poikolainen, 1999; Whitlock et al., 2004; Wilk et al., 1997). Therefore, practice guidelines recommend routine alcohol screening in primary care settings, followed by evidence-based brief alcohol counseling interventions for patients who screen positive (National Institute on Alcohol Abuse and Alcoholism, U.S. Department of Health and Human Services and National Institute of Health, 2005; U.S. Preventive Services Task Force, 2004).

Brief alcohol counseling interventions generally include assessment, feedback, advice, and goal setting (U.S. Preventive Services Task Force, 2004). Although their mechanisms of action are not entirely known, the effectiveness of these interventions relates to their patient centeredness (Miller and Rollnick, 2002; Miller et al.,

1993), or the ability of a clinician to negotiate with a patient in light of the patients' motivation to make his/her own decisions based on his/her readiness for behavioral change (Butler et al., 1996). Patient readiness can be conceptualized in several ways (Prochaska et al., 1997) and is often framed by Prochaska and DiClemente's transtheoretical model (TTM), which describes a patients' progression through 5 (sometimes modified to 3) discrete stages of change (Prochaska et al., 1992). A different facet of readiness is motivation for change, which includes confidence in the ability to succeed (often called self-efficacy; Bandura, 1990) and importance or value placed on changing (Butler et al., 1996; Miller and Rollnick, 2002).

To help tailor brief alcohol counseling interventions and predict subsequent drinking outcomes, use of measures of patients' readiness to change is recommended in clinical practice (Barnes and Samet, 1997; Dent et al., 1995; DiClemente et al., 2004; Prochaska et al., 1992; Rollnick et al., 1993; Samet and O'Connor, 1998; Samet et al., 1996). Many measures exist to assess patient readiness to change drinking. However, some are long, limiting their use in general healthcare settings where there are multiple preventive care agendas and limited time (Yarnall et al., 2003). An example is the SOCRATES, whose characteristics have been well described (Maisto et al., 1999; Miller and Tonigan, 1996), but whose 20-item scale may preclude its routine clinical use in these settings. An example of a shorter, but still fairly long, measure is the 12-item readiness to change questionnaire (RTCQ; Rollnick et al., 1992), which characterizes patients into stages of change based on the TTM (Prochaska and Norcross, 2001). Other shorter measures, if validated, might be used in clinical settings where time constraints are an issue. Examples include single-measure visual analogue scales (Miller and Rollnick, 2002), and a 3-item questionnaire designed to guide brief alcohol counseling interventions in primary care (Epler et al., 2005; Williams et al., 2006).

Whether aspects of readiness are associated with alcohol consumption or consequences for primary care patients with unhealthy alcohol use has not been well described (Riemsma et al., 2002). Self-efficacy is thought to be an important predictor of behavior change (Miller and Rollnick, 2002) and has been associated with better treatment outcomes (Demmel et al., 2004; Moos and Moos, 2006) and with decreased alcohol consumption and consequences (Blume et al., 2003) in patients who meet diagnostic criteria for alcohol use disorders. Further, the RTCQ has been predictive of subsequent alcohol consumption in general hospital inpatients (Heather et al., 1993). Based on these previous studies, we hypothesized that brief readiness measures predict alcohol consumption and consequences among primary care patients with unhealthy alcohol use identified by screening. This prospective cohort study assessed the concurrent validity and the ability of multiple measures of readiness to change drinking to predict alcohol consumption and consequences.

MATERIALS AND METHODS

Study Design

A prospective cohort of adult primary care patients with unhealthy alcohol use were enrolled in a randomized trial of providing patients' alcohol screening results to physicians (Saitz et al., 2003). Randomization for this trial occurred at the physician level, and analyses were adjusted for clustering of patients within physicians. This study represents a secondary analysis of data collected during this trial. The study was approved by the institutional review board of the Boston University Medical Center. Subjects provided informed consent to participate in the study and were told that their physician might be given the results of alcohol screening questions. Participants' privacy was further protected with a Certificate of Confidentiality obtained from the Federal government.

Participants

Patients were screened using a self-administered questionnaire (Aertgeerts et al., 2000) available in both Spanish and English, which included the CAGE alcohol screening questions (modified to refer to the past year rather than lifetime) and 3 questions about past-month quantity and frequency of drinking. Patients were eligible for the study if they screened positive for unhealthy alcohol use, either by answering yes to 1 or more of the CAGE questions and drinking any alcohol in the past month or by reporting drinking risky amounts in response to 3 questions about past-month quantity and frequency of consumption (National Institute on Alcohol Abuse and Alcoholism, U.S. Department of Health and Human services, and National Institute of Health, 2003). Risky amounts were defined in accordance with National Institute for Alcohol Abuse and Alcoholism recommendations (no more than 4 standard drinks per occasion or 14 per week for men, and no more than 3 per occasion or 7 per week for women; National Institute on Alcohol Abuse and Alcoholism, 1995). A more detailed description of study recruitment and the trial methods has been reported previously (Saitz et al., 2003). To be eligible for the current analysis, subjects had to have completed 6 months of follow-up and to have completed assessments on the variables of interest.

Assessments

Subjects were interviewed by a trained staff researcher in person at study entry both before and after meeting with their physician. Six months later, they were interviewed again, this time by phone. Demographic characteristics were assessed at study entry. Both assessments included readiness, alcohol consumption, and alcohol use consequence measures. Subjects were also asked about medical comorbidity (Charlson et al., 1987; Katz et al., 1996), psychiatric comorbidity (Spitzer et al., 1994; Whooley et al., 1997), and tobacco and other drug use. Assessments were completed in either Spanish or English. All questions not available in Spanish (Saitz et al., 1999) were translated from English, back-translated, and checked for accuracy.

Measures of readiness to change included 5 single questions and 1 validated scale. On the screening questionnaire, patients were asked to respond, on a scale of 0 to 10, to "how ready are you to change your drinking habits?" A response of 0 indicated "I don't drink; doesn't apply." Responses of 1 to 10 corresponded to not ready to ready. After subjects provided consent but before the physician visit, they were asked to report, on a scale of 1 to 10, the importance of changing and confidence in their ability to change (Miller and Rollnick, 2002). Response options for these questions

were scaled from 1 to 10, corresponding to not important or confident to very important or confident. At this time, subjects were also assessed with the validated RTCQ, which asks patients for scaled responses (5-point Likert scale, strongly disagree to strongly agree) to 12 statements. This instrument has demonstrated satisfactory test–retest reliability (Rollnick et al., 1992) and categorizes patients into precontemplation, contemplation, and action groups (Heather and Rollnick, 2000) based on the transtheoretical model (Prochaska and DiClemente, 1983). Finally, after the visit with the physician, subjects were asked to indicate their agreement with statements of intention to cut down and abstain using a 5-point categorized Likert Scale with response options ranging from "strongly disagree" to "strongly agree" and corresponding numerical scores from 1 to 5.

Alcohol consumption was assessed at the follow-up interview using the validated 30-day timeline followback (Sobell et al., 1996), which assesses the number of drinks consumed on each of the past 30 days. Alcohol consequences in the past 3 months were measured using the Short Inventory of Problems, Version 2 Revised (SIP-2R; Miller et al., 1995). This instrument includes 6 items that assess occurrence and frequency of consequences (response options include "never," "once or a few times," "once or twice a week," "daily or almost daily") and 9 items that assess the occurrence and extent of consequences (response options include "not at all," "a little," "somewhat," "very much"). Examples of alcohol-related consequences include (because of drinking) being unhappy, not eating properly, failing to do what is expected, feeling guilty or ashamed, harming physical health, damaging friendships or familiar relationships, or having an accident while drinking.

Predictor Variables—Readiness to Change

Readiness to change predictor variables were 5 single-item continuous measures and 1 multi-item categorical questionnaire. Single-item continuous measures included: (1) readiness to change (1–10), (2) importance of changing (1–10), (3) confidence in ability to change (1–10), (4) intention to cut down (1–5), and (5) intention to abstain (1–5). Additionally, stages of change (precontemplation, contemplation, or action) derived from the RTCQ were used as predictor variables (Heather and Rollnick, 2000).

Outcomes—Alcohol Consumption and Consequences

Three 30-day alcohol consumption dichotomous outcome variables were derived from the timeline followback measure. They included: (1) abstinence, (2) heavy episodic drinking, defined as >3 drinks on an occasion for women or participants over age 65 and >4 drinks per occasion for men, and (3) drinking risky amounts defined as either heavy episodic drinking (defined previously) or exceeding typical weekly limits (>7 drinks per week for women or those >65 years of age and >14 per week for men).

The alcohol consequences outcome measure was derived from the SIP-2R and was dichotomous (any of 15 consequences vs none) because 40% of subjects reported no consequences.

Analyses

Descriptive analyses were completed for all participating subjects. To assess concurrent validity, the mean scores on each single-item measure were obtained for patients in each stage of change as identified by the RTCQ, and differences in means were tested using a 2-degree of freedom test. Pearson's correlation coefficients were used to compare single-item measures of readiness.

To assess predictive validity, multivariable regression models were fit for each predictor and each outcome. No adjustment for multiple comparisons was undertaken. Models were adjusted for clustering within physician, randomization group, physician training (i.e., resident physician or not), patient sex, race, medical comorbidity, and education (high school completed). For dichotomous outcomes, generalized estimating equation regression models with working independence correlation structure and empirical variance estimators were fit (Zeger and Liang, 1986). Consequence models were adjusted for consequences at study entry; consumption models were adjusted for consumption at study entry. All analyses were carried out using SAS/STAT Version 8.2 (SAS/STAT® Software, 2001).

RESULTS

Among 4,143 patients who were approached to participate, 182 (4%) did not complete the self-administered screening questionnaire, and 565 of the remaining 3,961 (14%) screened positive for unhealthy alcohol use. Among those patients, 312 (55%) were enrolled in the study. As published previously, the subjects enrolled differed from those who were eligible but not enrolled in readiness to change drinking (single-item measure from screening, mean score 5.5 vs 4.9, respectively) and alcohol consumption (4.5 vs 3.4 drinks per drinking day, respectively; Saitz et al., 2003). Two hundred thirty-five enrolled subjects completed 6 months of follow-up, and 228/312 (73%) of these subjects had complete data on all variables of interest for analysis, with the exception of the single-item measure of readiness to change ($n = 207$). Among these 228 subjects, 128 (56%) saw physicians who had been randomized to receive the intervention (alcohol screening results). The baseline characteristics of subjects with complete data at 6 months compared with subjects enrolled but lost to follow-up or with incomplete data were not significantly different, with the exception of education. Graduation from high school was more common in subjects with complete data (68 vs 52%, $p = 0.01$).

The characteristics of the 228 subjects who were included in this analytic cohort are displayed in Table 1. Subjects were mostly male and African American, with a mean age of 44. Over two-thirds (68%) reported alcohol-related consequences, and subjects consumed 6 drinks per drinking day on average.

Most subjects (71%) indicated some readiness to change—118 (52%) were classified in the action and 43 (19%) in the contemplation stage of change by the RTCQ. Self-reported readiness on single-item readiness measures was generally high. On these measures with possible scores ranging from 1 to 10, the mean scores for readiness, importance, and confidence were 5.10 (SD 3.19), 6.00 (SD 3.61), and 8.00 (SD 2.56), respectively. For instruments ranging from 1 to 5, the mean scores for intention to cut down and intention to abstain were 3.43 (SD 1.05) and 2.62 (SD 1.09), respectively.

Mean confidence scores were lower for patients categorized in Contemplation than for those in other stages according to the RTCQ. Patients categorized by the RTCQ as being in precontemplation had lower mean scores on the other single-item readiness measures, and the mean scores tended to increase through the stages (from precontemplation to action), although the relation-

Table 1. Characteristics of Patients With Unhealthy Alcohol Use ($n = 228$)

Characteristic	% (*n*)
Female	39% (90)
Race/ethnicity	
African American	60% (136)
White	16% (37)
Latino	15% (34)
Other	9% (21)
High school education	68% (154)
No medical comorbidity	30% (69)
Patient's physician randomized to receive intervention (provision of screening results)	56% (128)
Any alcohol consequences	68% (155)
Males	75% (103)
Females	58% (52)
	Mean, SD (range)
Age	44, 13 (21–79)
Males	45, 14 (21–79)
Females	41, 10 (22–69)

	Study entry	Follow-up
	Mean, SD (range)	Mean, SD (range)
Drinks/drinking day	5.5, 5.0 (1–33.7)	6.7, 9.9 (1–95.2)
Males	13.3, 9.5 (1–33.7)	7.3, 10.6 (1–95.2)
Females	4.5, 3.8 (1–21.1)	5.7, 8.6 (1–63.2)
Alcohol-related consequences	7.9, 10.4 (0–45.0)	4.9, 8.2 (0–42.0)
Males	8.7, 10.7 (0–45.0)	5.9, 9.0 (0–39.0)
Females	6.6, 9.9 (0–44.0)	3.2, 6.6 (0–42.0)

ship was less consistent across contemplation and action (Table 2a). In addition, readiness, importance, and intention appeared to be positively correlated, but confidence was not highly correlated with readiness, importance, or intentions (Table 2b).

Both confidence (scale from 1 to 10) and intention to abstain (5-point Likert scale) were associated with reduced consumption: less heavy episodic drinking [adjusted odds ratio (AOR) for a 1-point increase 0.88, 95% CI 0.80–0.98 for confidence, AOR 0.79, 95% CI 0.64–0.98 for intention); and less drinking risky amounts (AOR 0.89, 95% CI 0.79–1.00 for confidence; AOR 0.78, 95% CI 0.62–0.98 for intention; Table 3). Intention to abstain was also associated with more abstinence (AOR 1.43, 95% CI 1.09–1.88). No other readiness measures were associated with alcohol consumption. All readiness measures, except intention to abstain and stage of change derived from the RTCQ, were significantly associated with later alcohol-related consequences. Readiness, importance, and intention to cut down were positively associated with the occurrence of consequences at follow-up despite controlling for the presence of consequences at study entry. More confidence, however, was associated with lower odds of any consequences (AOR 0.88, 95% CI 0.79, 0.98).

Table 2a. Mean Scores and Standard Deviations for Single-Item Readiness Measures Across the Stages of Change Identified by the Readiness to Change Questionnaire (RTCQ)

Single-item readiness measures	Stage of change as categorized by the RTCQ			
	Precontemplation	Contemplation	Action	*p* Value
Readiness to change	3.44 (2.77)	5.24 (2.64)	6.08 (3.23)	<0.0001
Importance of changing	3.30 (2.89)	7.47 (2.99)	7.01 (3.37)	<0.0001
Confidence in changing	8.37 (2.47)	6.88 (2.98)	8.19 (2.35)	0.0077
Intention to cut down	2.70 (1.09)	3.53 (0.93)	3.80 (0.85)	<0.0001
Intention to abstain	1.99 (0.83)	2.74 (1.00)	2.93 (1.12)	<0.0001

Table 2b. Pearson's Correlation Coefficients for Single-Item Readiness Measures[a]

	Readiness	Importance	Confidence	Intent to cut down
Importance	**0.59**			
Confidence	**0.17**	0.02		
Intent to cut down	**0.53**	**0.69**	0.02	
Intent to abstain	**0.45**	**0.55**	0.10	**0.52**

[a]Correlations significant at $p < 0.05$ are in bold above. *p* Values for the correlations were derived from a generalized estimating equation (GEE) linear regression model that accounts for clustering within physician by use of an independence working correlation matrix and empirical variance estimator.

DISCUSSION

In this study of adult primary care patients who screened positive for unhealthy alcohol use, patients' readiness to change drinking was generally high, regardless of the measure used. The measures exhibited a fair amount of concurrent validity, but notably, the measure of confidence was not highly correlated with the other measures. However, most measures of readiness were not associated with subsequent alcohol consumption. Confidence in ability to change drinking and intention to abstain were both associated with lower odds of heavy episodic drinking and drinking risky amounts. Additionally, a single-item measuring confidence to change was associated with lower odds of alcohol consequences, but other measures of readiness (single-item measures of readiness, importance of changing, and intention to cut down) were associated with more, not fewer, alcohol-related adverse consequences at follow-up.

It is well known that consequences and discrepancies between values and actions related to drinking often drive people toward change (Cunningham et al., 2005; Isenhart et al., 1998; Kahler, 2001; Sobell et al., 1993). As such, it is likely that the observed association between readiness and consequences reflects an effect cause relationship rather than cause effect, despite the use of a prospective analysis adjusting for consequences reported at study entry. In other words, people who experience more drinking consequences will report greater readiness (Kahler, 2001; Strecher et al., 1994; Williams et al., 2006). However, patients with long-standing alcohol problems and related consequences report low self-efficacy (Saxon et al., 2005). The strong relationship between readiness and consequences creates difficulties for the clinical utility or predictive ability of many measures of readiness, and it is here that

Table 3. Readiness to Change Drinking and Subsequent Alcohol Consumption and Consequences (Adjusted[a])

	Abstinence	Heavy episodic drinking	Drinking risky amounts	Any alcohol-related consequences
	AOR (95% CI)	AOR (95% CI)	AOR (95% CI)	AOR (95% CI)
Single-item readiness measures				
Readiness	0.99 (0.87–1.12)	1.04 (0.97–1.11)	1.04 (0.97–1.11)	**1.15 (1.04–1.27)**
Importance	1.02 (0.90–1.16)	1.05 (0.98–1.12)	1.04 (0.97–1.11)	**1.13 (1.01–1.27)**
Confidence	1.08 0.93–1.27	**0.88 (0.80–0.98)**	**0.89 (0.79–1.00)**	**0.88 (0.79–0.98)**
Intention to cut down	1.34 (0.93–1.93)	1.04 (0.83–1.30)	1.06 (0.84–1.34)	**1.58 (1.12–2.22)**
Intention to abstain	**1.43 (1.09–1.88)**	**0.79 (0.64–0.98)**	**0.78 (0.62–0.98)**	1.17 (0.85–1.61)
Multi-item Readiness to Change Questionnaire				
Stage of change (vs reference group precontemplation)				
Action	1.19 (0.47–2.97)	0.93 (0.49–1.76)	1.08 (0.56–2.05)	1.44 (0.78–2.64)
Contemplation	1.81 (0.48–6.85)	0.93 (0.39–2.20)	1.07 (0.46–2.47)	1.82 (0.56–5.97)

[a]Adjusted for clustering within physician, randomization, physician training, patient sex, race, medical comorbidity, and education. Consequence models were adjusted for consequences at study entry; consumption models were adjusted for consumption at study entry.

AOR, adjusted odds ratio.

Results in **bold** are statistically significant ($p \leq 0.05$).

the constructs of readiness and self-efficacy may diverge. Increasing consequences may lead patients to consider change more strongly, which would result in a higher report of readiness from a patient. However, this may not reflect a patient's self-perception of his/her ability to make change.

In this study, the measure of self-efficacy (confidence to change drinking) was not correlated with the other measures of readiness. Unlike the other readiness measures, confidence was associated with *fewer* consequences at follow-up. This finding is consistent with studies in other populations, which have shown that self-efficacy is associated with improved alcohol-related outcomes (Blume et al., 2003; Demmel et al., 2004; Ilgen et al., 2006; Moos and Moos, 2006) and suggests that this simple measure may be a useful tool for predicting alcohol consumption and future alcohol consequences in primary care patients with unhealthy alcohol use.

Also noteworthy among our findings was that the RTCQ did not predict subsequent alcohol consumption or consequences. Heather et al. (1993) evaluated the predictive validity of the RTCQ and found that stage of change was a significant predictor of reduced drinking at follow-up. That study was conducted among hospital inpatients, 18% of whom were in the hospital for conditions related to their drinking. To the extent that hospitalizations may be ''teachable moments,'' these patients could have experienced heightened readiness to change and increased motivation to change in the months following hospitalization (McBride et al., 2003). Whether the readiness measures evaluated in the current study would have similar results in hospitalized patients is unknown.

These divergent findings are also interesting in the context of increasing evidence that stage-of-change–based behavioral interventions delivered in medical care settings may not be more effective than those that do not specifically focus on stage of change (Riemsma et al., 2003; Van Sluijs et al., 2004). Stage-based interventions have been advocated for unhealthy alcohol use (Barnes and Samet, 1997; Burge and Schneider, 1999; Butler et al., 1999; Prochaska and DiClemente, 1983; Richmond and Mendelsohn, 1998; Rollnick et al., 1992; Samet and O'Connor, 1998; Samet et al., 1996). But because time is limited in busy clinical settings (Richmond and Mendelsohn, 1998), our finding that the RTCQ was not associated with subsequent alcohol-related outcomes suggests that identifying a patient's stage of change may not be essential, particularly if the purpose is to predict outcomes. However, there may still be clinical value to assessing readiness to change as part of a conversation or counseling strategy that emphasizes a patient centeredness when addressing unhealthy alcohol use (Butler et al., 1996; Miller et al., 1993). Given this study's finding that confidence is predictive of positive outcomes, interventions that have support of self-efficacy or confidence as primary or central components, such as motivational interviewing (Miller and Rollnick, 2002), might have greater utility than stage-based interventions for promoting behavior change.

Our study has several limitations. The patients in this cohort were in a clinical trial at an urban medical center, which may limit the generalizability of our findings. However, there is no reason to believe that rural or suburban patients with unhealthy alcohol use would have a different relationship between readiness to change and subsequent drinking outcomes. Further, participants reported greater readiness to change and tended to drink more than eligible patients who were not enrolled. It is possible that our findings would not apply to patients who are slightly less ready or drink slightly less, but it is not clear what importance these small differences in characteristics would have. Additionally, loss to follow-up may have resulted in bias. However, the analytic sample with complete data was similar to subjects lost to follow-up or with incomplete data.

Another limitation may be that, because patients were informed that screening results would be provided to their clinicians, they may not have been entirely honest in their responses after the initial readiness assessment that occurred during screening. While possible, this scenario seems unlikely because patients were informed about this sharing and that honest responses were desired. They then had a choice to participate in the study or not if they did not want results shared (or for any other reason). Because of the known limits of sensitivity of validated screening tools, we may have missed some patients with unhealthy alcohol use, but this would not have been because of anticipation of shared results as subjects were not informed of this possibility until after screening. Additionally, it is not clear that the relationship between readiness and consumption and consequences would have differed in those patients from the relationship seen in study subjects. Further, the fact that patients were assessed for most readiness measures before visits with clinicians but were assessed for their intentions to cut down and abstain after may have some impact on the predictive validity of these measures. However, we controlled for randomization group to address this possible effect. Finally, it is possible that the effects of readiness to change as a predictor of subsequent drinking behavior may be mediated or moderated by other factors such as self-efficacy or personal values. For example, if either readiness or confidence are lacking, then little behavior change might be expected, whereas the presence of readiness coupled with a high level of confidence might be expected to predict behavior change. The current analyses do not directly address this issue, which could be addressed with further research.

Despite these limitations, this study has noteworthy strengths. We prospectively evaluated multiple measures and facets of readiness to change and their associations with diverse drinking outcomes. Although other brief measures of readiness exist (Epler et al., 2005; LaBrie et al., 2005), their predictive validity is not known. Further, this study evaluates the predictive validity of the RTCQ in a primary care setting, where brief interventions for unhealthy alcohol use are recommended.

In summary, among primary care patients with unhealthy alcohol use, many brief measures of readiness to change may play an important role in discussions of a patient's alcohol use during brief intervention, especially given that physicians have exhibited noteworthy discomfort in alcohol-related discussions (McCormick et al., 2006). However, in this study, most measures did not appear to predict meaningfully later alcohol consumption or decreases in consequences. Tools to assess readiness to change drinking must be evaluated for their predictive ability as well as their clinical utility, as these measures do not all consistently predict change. Intention to abstain and confidence in ability to change (self-efficacy) both showed promise for predicting change as they were both associated with subsequent improvements in drinking outcomes. Confidence, as measured by a single item, was additionally associated with fewer alcohol-related consequences at follow-up. As such, assessing confidence in ability to change drinking may be useful as part of brief counseling interventions for unhealthy alcohol use in primary care settings as it holds the most promise for predicting decreases in both consumption and consequences.

REFERENCES

Aertgeerts B, Buntinx F, Fevery J, Ansoms S (2000) Is there a difference between CAGE interviews and written CAGE questionnaires? Alcohol Clin Exp Res 24:733–736.

Ballesteros J, Duffy JC, Querejeta I, Arino J, Gonzalez-Pinto A (2004) Efficacy of brief interventions for hazardous drinkers in primary care: systematic review and meta-analyses. Alcohol Clin Exp Res 28: 608–618.

Bandura A (1990) Self-regulation of motivation through anticipation and self-reactive mechanisms. Nebr Symp Motiv 38:69–164.

Barnes HN, Samet JH (1997) Brief interventions with substance-abusing patients. Medical Clin N Am 81:867–879.

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

Bien TH, Miller WR, Tonigan S (1993) Brief interventions for alcohol problems: a review. Addiction 88:315–336.

Blume AW, Lostutter TW, Schmaling KB, Marlatt GA (2003) Beliefs about drinking behavior predict drinking consequences. J Psychoactive Drugs 35:395–399.

Burge SK, Schneider FD (1999) Alcohol-related problems: recognition and intervention. Am Fam Physician 59:361–370, 372.

Butler C, Rollnick S, Stott N (1996) The practitioner, the patient and resistance to change: recent ideas on compliance. Can Med Assoc J 154:1357–1362.

Butler CC, Rollnick S, Cohen D, Bachmann M, Russell I, Scott N (1999) Motivational consulting versus brief advice for smokers in general practice: a randomized trial. Br J Gen Pract 49:611–616.

Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40:373–383.

Cunningham JA, Blomqvist J, Koski-Jannes A, Cordingley J (2005) Current heavy drinkers' reasons for considering change: results from a natural history general population survey. Addict Behav 30: 581–584.

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. Eur Addict Res 10:133–138.

Dent TH, Shepherd RM, Alexander GJ, London M (1995) Do cage scores predict readiness to reduce alcohol consumption in medical in-patients? Alcohol Alcohol 30:577–580.

DiClemente CC, Schlundt D, Gemmell L (2004) Readiness and stages of change in addiction treatment. Am J Addict 13:103–119.

Epler AJ, Kivlahan DR, Bush KR, Dobie DJ, Bradley KA (2005) A brief readiness to change drinking algorithm: concurrent validity in female VA primary care patients. Addict Behav 30:389–395.

Heather N, Rollnick S (2000) *Readiness to Change Questionnaire: User's manual (Revised Version)*. Centre for Alcohol & Drug Studies, University of Northumbria at Newcastle.

Heather N, Rollnick S, Bell A (1993) Predictive validity of the Readiness to Change Questionnaire. Addiction 88:1667–1677.

Ilgen M, Tiet Q, Finney J, Moos RH (2006) Self-efficacy, therapeutic alliance, and alcohol-use disorder treatment outcomes. J Stud Alcohol 67:465–472.

Isenhart CE, Van Krevelen S (1998) Relationship between readiness for and processes of change in a sample of alcohol dependent males. J Subst Abuse 10:175–184.

Kahan M, Wilson L, Becker L (1995) Effectiveness of physician-based interventions with problem drinkers. Can Med Assoc J 152: 851–859.

Kahler CW (2001) Generation and recall of alcohol-related information in excessive drinkers: relation to problem severity, outcome expectancies, and stage of change. Psychol Addict Behav 15:109–117.

Katz JN, Change LC, Sangha O, Fossel AH, Bates DW (1996) Can comorbidity be measured by questionnaire rather than medical record review? Med Care 34:73–84.

LaBrie JW, Quinlan T, Schiffman JE, Earleywine ME (2005) Performance of alcohol and safer sex change rulers compared with readiness to change questionnaires. Psychol Addict Behav 19:112–115.

Maisto SA, Conigliaro J, McNeil M, Kraemer K, O'Connor M, Kelley ME (1999) Factor structure of the SOCRATES in a sample of primary care patients. Addict Behav 24:879–892.

McBride CM, Emmons KM, Lipkus IM (2003) Understanding the potential of teachable moments: the case of smoking cessation. Health Educ Res 18:156–170.

McCormick KA, Cochran NE, Back AL, Merrill JO, Williams EC, Bradley KA (2006) How primary care providers talk to patients about alcohol: a qualitative study. J Gen Intern Med 21:966–972.

Miller WR, Benefield RG, Tonigan JS (1993) Enhancing motivation for change in problem drinking: a controlled comparison of two therapist styles. J Consult Clin Psychol 61:455–461.

Miller WR, Rollnick S (2002) *Motivational Interviewing: Preparing People for Change*, 2nd ed., The Guilford Press, New York, NY.

Miller WR, Tonigan JS (1996) Assessing drinkers' motivation for change: the stages of change readiness and treatment eagerness scale (SOCRATES). Psychol Addict Behav 10:81–89.

Miller WR, Tonigan J, Longabaugh R (1995) *The Drinker Inventory of Consequences (DrInC). NIH Pub. No. 94-3724*. National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD.

Moos RH, Moos BS (2006) Rates and predictors of relapse after natural and treated remission from alcohol use disorders. Addiction 101: 212–222.

Moyer A, Finney JW, Swearingen CE, Vergun P (2002) Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. Addiction 97:279–292.

National Institute on Alcohol Abuse and Alcoholism (1995) *The Physician's Guide to Helping Patients with Alcohol Problems. NIH Publication # 95–3769*. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, MD.

National Institute on Alcohol Abuse and Alcoholism, U.S. Department of Health and Human services, & National Institute of Health (2003) *Helping Patients with Alcohol Problems: A Practitioner's Guide. NIH Publication No. 03–3769*. National Institute on Alcohol Abuse and Alcoholism, U.S. Department of Health and Human Services, and National Institutes of Health, Bethesda, MD.

National Institute on Alcohol Abuse and Alcoholism, U.S. Department of Health and Human services, & National Institute of Health (2005) *Helping Patients Who Drink Too Much: A Clinician's Guide*. National Institute on Alcohol Abuse and Alcoholism, U.S. Department of Health and Human services, and National Institute of Health, Bethesda, MD.

Poikolainen K (1999) Effectiveness of brief interventions to reduce alcohol intake in primary health care populations: a meta-analysis. Prev Med 28:503–509.

Prochaska JO, DiClemente CC (1983) Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 51:390–395.

Prochaska JO, DiClemente CC, Norcross JC (1992) In search of how people change: applications to addictive behaviors. Am Psychol 47:1102–1114.

Prochaska JO, Norcross JC (2001) Stages of change. Psychotherapy 38:443–448.

Prochaska JO, Redding CA, Evers K (1997) The transitional model and stages of change, in *Health Behavior and Health Education* (Glanz K, Lewis FM, Rimber BK eds), pp 60–84. Jossey-Bass Inc, San Francisco, CA.

Richmond RL, Mendelsohn CP (1998) Physicians' views of programs incorporating stages of change to reduce smoking and excessive alcohol consumption. Am J Health Promot 12:254–257.

Riemsma RP, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, Walker A (2002) A systematic review of the effectiveness of interventions based on a stages-of-change approach to promote individual behavior change. Health Technol Assess 6, (No. 24, http://www.hta.nhsweb.nhs.uk/fullmono/mon624.pdf accessed July 8, 2005).

Riemsma RP, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, Walker A (2003) Systematic review of the effectiveness of stage based interventions to promote smoking cessation. BMJ 326:1175–1177.

Rollnick S, Heather N, Bell A (1992) Negotiating behaviour change in medical settings: the development of brief motivational interviewing. J Ment Health 1:25–37.

Rollnick S, Heather N, Gold R, Hall W (1992) Development of a short 'readiness to change' questionnaire for use in brief, opportunistic interventions among excessive drinkers. Br J Addict 87:743–754.

Rollnick S, Kinnersley P, Stott N (1993) Methods of helping patients with behaviour change. BMJ 307:188–190.

Saitz R (2005) Clinical practice. Unhealthy alcohol use. N Engl J Med 352:596–607.

Saitz R, Horton NJ, Sullivan LM, Moskowitz M, 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.

Saitz R, Lepore M F, Sullivan LM, Amaro H, Samet JH (1999) Alcohol abuse and dependence in Latinos living in the United States: validation of the CAGE (4M) questions. Arch Intern Med 159:718–724.

Samet JH, O'Connor PG (1998) Alcohol abusers in primary care: readiness to change behavior. Am J Med 105:302–306.

Samet JH, Rollnick S, Barnes H (1996) Beyond CAGE. A brief clinical approach after detection of substance abuse. Arch Intern Med 156:2287–2293.

SAS/STAT® Software. (2001) *Changes and Enhancements (Version Release 8.2)*. SAS Institute Inc, Cary, NC.

Saxon AJ, Kivlahan DR, Donovan DM (2005) Further validation of the Alcohol Dependence Scale as an index of severity. Alcohol Clin Exp Res 29S:31A.

Sobell LC, Brown J, Leo GI, Sobell MB (1996) The reliability of the Alcohol Timeline Followback when administered by telephone and by computer. Drug Alcohol Depend 42:49–54.

Sobell LC, Sobell MB, Toneatto T, Leo GI (1993) What triggers the resolution of alcohol problems without treatment. Alcohol Clin Exp Res 17:217–224.

Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV 3rd, Hahn SR, Brody D, Johnson JG (1994) Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. JAMA 272:1749–1756.

Strecher VJ, Kobrin SC, Kreuter MW, Roodhouse K, Farrell D (1994) Opportunities for alcohol screening and counseling in primary care. J Fam Pract 39:26–32.

U.S. Preventive Services Task Force (2004) Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. Ann Intern Med 140:554–556.

Van Sluijs EM, Van Poppel MN, Van Mechelen W (2004) Stage-based lifestyle interventions in primary care: are they effective? Am J Prev Med 26:330–343.

Whitlock EP, Polen MR, Green CA, Orleans T, Klein J (2004) 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 140: 557–568.

Whooley MA, Avins AL, Miranda J, Browner WS (1997) Case-finding instruments for depression. Two questions are as good as many. J Gen Intern Med 12:439–445.

Wilk AI, Jensen NM, Havighurst TC (1997) Meta-analysis of randomized control trials addressing brief interventions in heavy alcohol drinkers. J Gen Intern Med 12:274–283.

Williams EC, Kivlahan DR, Saitz R, Merrill JO, Achtmeyer CE, McCormick KA, Bradley KA (2006) Readiness to change in primary care patients who screened positive for alcohol misuse. Ann Fam Med 4:213–220.

Yarnall KSH, Pollak KI, Ostbye T, Krause KM, Micchener JL (2003) Primary care: is there enough time for prevention? Am J Public Health 93:635–641.

Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42:121–130.

Available online at www.sciencedirect.com
ScienceDirect
Drug and Alcohol Dependence 89 (2007) 161–169
DRUG and ALCOHOL DEPENDENCE
www.elsevier.com/locate/drugalcdep

Overdose after detoxification: A prospective study☆

James D. Wines Jr.[a,b,*], Richard Saitz[c,d], Nicholas J. Horton[e], Christine Lloyd-Travaglini[f], Jeffrey H. Samet[c,g]

[a] Alcohol and Drug Abuse Research Center (ADARC), McLean Hospital, 115 Mill Street, Belmont, MA 02478, USA
[b] Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA
[c] Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA 02118, USA
[d] Department of Epidemiology and the Youth Alcohol Prevention Center, Boston, University School of Public Health, Boston, MA 02118, USA
[e] Department of Mathematics and Statistics, Smith College, Northampton, MA 01063, USA
[f] Data Coordinating Center, Boston University School of Public Health, Boston, MA 02118, USA
[g] Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA 02118, USA

Received 25 April 2006; received in revised form 28 November 2006; accepted 14 December 2006

Abstract

Objective: The aim of this study was to determine predictors of non-fatal overdose (OD) among a cohort of 470 adults after detoxification from heroin, cocaine or alcohol.
Methods: We examined factors associated with time to OD during 2 years after discharge from an urban detoxification unit in Boston, MA, USA using multivariable regression analyses. Separate analyses were performed for both the total sample and a subgroup with problem opioid use.
Results: Lifetime prevalence for any OD was 30.9% (145/470) in the total sample and 42.3% (85/201) in patients with opioid problems. During the 2-year follow-up, OD was estimated to occur in 16.9% of the total sample and 26.7% of the opioid problem subgroup, with new-onset (incidence) OD estimated at 5.7% and 11.0%, respectively. Factors associated with an increased hazard of OD in both samples included white race, more depressive symptoms, and prior OD regardless of intent. Prior suicidal ideation or attempt was not associated with future OD.
Conclusions: Findings underscore both the high prevalence of non-fatal OD among detoxification patients especially opioid users, and the potency of prior OD as a risk factor for future OD. Depressive symptoms, a modifiable risk factor, may represent a potential intervention target to prevent OD, including some "unintentional" ODs.

Keywords: Overdose; Depressive symptoms; Opioids; Intent

1. Introduction

In 2003, there were 28,700 fatal poisonings and 817,797 non-fatal poisonings in the US (CDC, 2006). Poisoning, which include overdoses (ODs) on illicit drugs, alcohol, and medications, is the leading cause of injury death for individuals age 35–44 and the third leading cause of injury death overall, trailing motor vehicle accidents and firearm-related deaths. Poisoning fatalities in 11 states increased 56% from 1990 to 2001, and unintentional poisonings, especially opioid-related ODs, may be driving the increase (CDC, 2004).

Opioid-related ODs have increased at an alarming rate in portions of the US (Ballesteros et al., 2003; Landen et al., 2003; Sanford, 2002; Sorg and Greenwald, 2002) and other countries (Darke and Hall, 2003; Warner-Smith et al., 2001), and OD has surpassed HIV infection as the primary cause of death for heroin users in some areas (Frischer et al., 1993). Not surprisingly, heroin is frequently associated with opioid-related ODs, both as a single drug and in combination with other substances (CDC, 2004; Darke et al., 1996a; Paulozzi, 2006; Paulozzi et al., 2006; Powis et al., 1999; Tobin and Latkin, 2003). Much of the current scientific literature regarding opioid-related OD is based on studies focused exclusively on intravenous heroin users conducted in Australia and Europe (Darke and Hall, 2003). The increased availability, misuse, abuse of and dependence on

☆ Preliminary results were presented at the Complexities of Co-occurring Conditions Conference, June 23–25, 2004 in Washington, DC.

* Corresponding author at: McLean Hospital, 115 Mill Street, Belmont, MA 02478, USA. Tel.: +1 617 855 2769; fax: +1 617 855 2519.
E-mail address: jwines@hms.harvard.edu (J.D. Wines Jr.).

0376-8716/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.drugalcdep.2006.12.019

prescription opioids in the US (Compton and Volkow, 2006; Zacny et al., 2003) may limit the generalizability of these OD-related findings from other countries to the US population and to populations abusing prescription opioids. Indeed, fatal drug poisonings related to prescription opioids now exceed heroin- and cocaine-related poisoning deaths (Paulozzi, 2006; Paulozzi et al., 2006). Recently, leadership at the US National Institute on Drug Abuse (Compton and Volkow, 2006) stated, ". . .the relationship of prescription opioid drugs to death is of increasing concern (p. 104)."

Unintentional OD and suicide contribute considerably to the increased mortality exhibited by heroin users (Davoli et al., 1997; Harris and Barraclough, 1997; Hser et al., 2001; Hulse et al., 1999; Oppenheimer et al., 1994; Quaglio et al., 2001). Although existing research clearly supports the notion that most opioid-related ODs are considered unintentional, the reported proportion of unintentional ODs with opioids ranges widely (e.g., from 99% (Darke et al., 1996a) to 51% (Neale, 2000) in some studies). Controversy also exists as to whether or not some unintentional ODs represent a type of "hidden suicide" (Darke and Ross, 2002; Kjelsberg et al., 1995; Ohberg and Lonnqvist, 1998). Fueling this debate are contradictory findings from some studies examining the relationship between suicidal behavior and OD (Best et al., 2000; Darke and Ross, 2001; Darke et al., 2004; Kosten and Rounsaville, 1988; Murphy et al., 1983; Neale, 2000; Ravndal and Vaglum, 1999; Rossow and Lauritzen, 1999; Vingoe et al., 1999).

Several studies have also investigated the relationship between depressive symptoms and OD. Tobin and Latkin (2003) surveyed 729 opioid and cocaine users recruited from the community and found a strong association between elevated depressive symptoms and unintentional OD within the past year. In addition, researchers in Norway (Ravndal and Vaglum, 1999) interviewed 200 mixed drug users at entry into a therapeutic community and again after 5 years. Although opioids increased unintentional OD risk, no associations of this form of OD with psychopathology including depression were found. These results appear to contrast with the study (Tobin and Latkin, 2003) described above; however, the lack of an association between depression and unintentional OD may be explained by the use of different assessments for depression administered after a longer time interval (5 years) and perhaps by the samples selected (community versus treatment).

Prospectively obtained data on non-fatal OD among subjects with diverse drug preferences are limited, especially in the US, despite steadily increasing OD rates. In addition to determining the prevalence and incidence of non-fatal OD in a cohort of patients initially admitted for detoxification from heroin, cocaine or alcohol, the purpose of this prospective study was to (1) improve our understanding of OD risk by drug type, (2) better characterize the magnitude of the risk of prior OD for future OD, (3) examine the role of depressive symptoms and OD, and (4) explore the relationship of intent and OD. The study's first hypothesis was that OD risk varies across drug categories. The second and third hypotheses were that prior OD and depressive symptoms would be associated with OD during follow-up.

2. Methods

2.1. Design and subjects

Data for this prospective cohort study of OD were obtained from subjects enrolled in the HELP (Health Evaluation and Linkage to Primary care) study, a randomized clinical trial testing the effectiveness of a multidisciplinary health intervention to link individuals admitted for detoxification to primary medical care. The HELP study has been described in detail previously (Samet et al., 2003).

Subjects in the HELP study were recruited from an inpatient detoxification unit in Boston, MA, which primarily treats uninsured or Medicaid-eligible patients. Eligible subjects were at least 18 years old; identified alcohol, cocaine, or heroin as their first or second drug of choice; and lacked a primary care physician. Exclusion criteria for the HELP study were as follows: (1) not English or Spanish speaking; (2) being pregnant; (3) living outside the Boston Medical Center catchment area; (4) significant cognitive impairment; (5) unable to provide three contacts for tracking purposes; and (6) specific plans to leave the Boston area within the next year. Subjects were scheduled for follow-up every 6 months for 2 years after study entry, and 85% (400/468) were assessed at least once during follow-up (two subjects died prior to follow-up). All subjects provided written informed consent prior to enrolling in the study, which was approved by the Institutional Review Board at Boston University Medical Center; subjects were also protected by a Certificate of Confidentiality from the National Institutes of Health.

2.2. Assessments

Standardized interviews were administered by trained research personnel at study entry (after resolution of acute withdrawal symptoms) and during follow-up. Assessments for the current study included: age; gender; race/ethnicity; homelessness (defined as having lived on the street or in a homeless shelter at least one night during the previous 6 months); physical health status as measured by the Short Form Health Survey, Physical Component Summary (SF-36-PCS) (Ware, 1993); social support from friends and family assessed by the Perceived Social Support scale (PSS) (Procidano and Heller, 1983); sexual or physical abuse (Liebschutz et al., 2002); depressive symptoms as measured by the Center for Epidemiologic Studies Depression (CES-D) scale (Radloff, 1977); and the Addiction Severity Index (ASI) (McLellan et al., 1992). The ASI was used to construct indicator variables for suicidal ideation, suicide attempt, *Problem Drug* categories, and continuous variables for the total number of *Problem Drugs* and *Drug Use* as detailed below.

2.2.1. Lifetime suicidality assessed at detoxification (study entry). Lifetime suicidal ideation and lifetime suicidal attempt were assessed at study enrollment by two questions from the Addiction Severity Index (ASI): "Have you ever attempted suicide? (include actual suicidal gestures or attempts)" and "Have you ever experienced serious thoughts of suicide? (patient seriously considered a plan for taking his/her life)."

2.2.2. Problem drug categories assessed at detoxification (study entry). In the absence of formal research assessment of drug diagnoses (abuse or dependence), frequency of drug use was used as a marker for excessive or heavy use, very likely problematic for people who met admission criteria for inpatient detoxification. *Problem Drug* categories were constructed from the drugs listed in the ASI at the index admission for detoxification (Wines et al., 2004). *Problem Drug* for a particular substance was defined as use $\geq$3 times per week (for alcohol "use" means to intoxication or $\geq$3 drinks) for a year or more, or 5 or more days of use in the past 30 days (definitions based on the ASI and outcome definitions used in treatment studies) (McLellan et al., 1992; O'Malley et al., 1992; Volpicelli et al., 1992). Theoretical considerations were used to combine drugs with similar pharmacologic properties for analysis, creating four *Problem Drug* categories: alcohol; opioids (heroin, methadone and other opioids/analgesics); sedatives (barbiturates and sedatives/hypnotics/tranquilizers); and stimulants (cocaine and amphetamines). Indicator variables for each *Problem Drug* category were produced by combining the relevant drugs listed in the ASI. For example, if a subject met the *Problem Drug* definition for any of

the three opioid drugs (heroin, methadone and other opioids/analgesics), then the subject would be considered to have problem opioid use (Opioid *Problem Drug*), an indicator of frequent use and a proxy for opioid abuse/dependence. In addition, the total number of the 11 drugs listed on the ASI fulfilling the *Problem Drug* criteria was used to create a continuous variable.

2.2.3. Drug Use after detoxification (follow-up). *Drug Use* was defined as the number of days used in the past 30 days as recorded by the ASI (range, 0–30 days) during follow-up (Wines et al., 2004). For *Problem Drug* categories consisting of more than one drug (i.e., opioids, sedatives, stimulants), the single most commonly used drug by the cohort at study entry (i.e., heroin, benzodiazepines, cocaine) was utilized to create four *Drug Use* variables. This 'one drug per category' approach was taken in part to minimize ambiguity regarding the number of days used, as it cannot be determined from the ASI whether drugs used within a category are used on the same day or on different days (Wines et al., 2004). Specifically, Alcohol *Use* was defined as the number of days of any alcohol use; Heroin, Benzodiazepine, and Cocaine *Use* were defined as the number of days of each drug used. Also, as noted above, the total number of drugs used in the past 30 days was a continuous variable.

2.3. Outcome variables

The main outcome variables for this study were lifetime OD ("prior OD") assessed at the detoxification unit (study entry) and time to any OD during the 6-month intervals (6, 12, 18 and 24 months) after detoxification (follow-up). Lifetime OD ("prior OD") was ascertained at study entry by the following question, "Have you ever had a drug or alcohol OD requiring you to go to the emergency room (requiring medical attention right away)?" OD during follow-up was defined similarly by asking subjects if they had an OD since their last assessment. To explore the role of intent, lifetime prior OD was divided into two groups based on intent. The first group, "Prior Unintentional OD," was defined at study entry as a positive response to the Lifetime OD question above and a negative response to the suicide attempt question from the ASI. The second group, "Prior Other OD," was defined at study entry as a positive response to the Lifetime OD question above and a positive response on the attempted suicide question. It should be noted that the "No OD" group contained only subjects without any history of OD, the "Prior Unintentional OD" group contained only individuals with a history of unintentional OD, and the "Prior Other OD" group could contain subjects with an unintentional (i.e., if past suicide attempt was not by OD) or intentional OD, or both types of OD.

2.4. Statistical analyses

The extant scientific literature and clinical experience guided the selection of variables. Data from study entry were used to compare three groups according to OD history and intent (No OD, Prior Unintentional OD, and Prior Other OD) and to determine correlates of lifetime (prior) OD. Due to the increased lifetime (prior) OD risk from opioids, separate analyses were performed for both the total sample and the Opioid *Problem Drug* subgroup, which was constructed by limiting the sample to subjects meeting criteria for Opioid *Problem Drug*. For all analyses, age, SF-36-Physical Component Summary (PCS), Perceived Social Support (Family), Perceived Social Support (Friends), and depressive symptoms (CES-D) were treated as continuous variables. For the three group comparisons (No OD, Prior Unintentional OD, and Prior Other OD), continuous and categorical variables were analyzed using ANOVA and Chi-square, respectively (Tables 1 and 2). Post hoc tests (Duncan's method) were used to compare means of statistically significant continuous variables for the three group comparisons.

The first hypothesis was tested in logistic regression models predicting lifetime (prior) OD (Section 3.2). The models contained the *Problem Drug* categories and potential confounders: age, gender, race, homeless status, SF-36-Physical Component Summary (PCS), Perceived Social Support (Family), Perceived Social Support (Friends), history of physical or sexual abuse, and depressive symptoms (CES-D). Secondary analyses were conducted by replacing the *Problem Drug* categories with a single continuous variable representing the total number of drugs listed on the ASI fulfilling the *Problem Drug* criteria.

Table 1
Characteristics of all subjects admitted for detoxification (total sample)

Characteristic	% (*n*)/mean (S.D.)			
	No OD 69% (*n* = 325)	Prior Unintent OD 19% (*n* = 89)	Prior Other OD 12% (*n* = 56)	Total (*n* = 470)
Age**	35.2 a (7.7)	38.1 b (7.7)	35.1 a (7.7)	35.8 (7.8)
Male**	74.5 (242)	89.9 (80)	66.1 (37)	76.4 (359)
Race***				
Black	55.7 (181)	27.0 (24)	23.2 (13)	46.4 (218)
White	29.2 (95)	56.2 (50)	50.0 (28)	36.8 (173)
Hispanic	8.9 (29)	11.2 (10)	21.4 (12)	10.9 (51)
Other	6.2 (20)	5.6 (5)	5.4 (3)	6.0 (28)
Homelessness***	39.4 (128)	61.8 (55)	64.3 (36)	46.6 (219)
Physical health status***	49.3 a (10.6)	45.8 b (10.6)	44.3 b (10.6)	48.1 (10.8)
Social support (family)*	7.3 a (4.7)	6.6 a,b (4.7)	5.5 b (4.7)	7.0 (4.7)
Social support (friends)	6.8 (4.1)	6.8 (3.7)	5.8 (3.6)	6.7 (4.0)
Depressive symptoms***	30.7 a (11.9)	35.4 b (11.9)	41.5 c (11.9)	32.9 (12.5)
Abuse history***				
No abuse	32.8 (106)	20.7 (18)	10.9 (6)	28.0 (130)
Physical abuse only	43.7 (141)	49.4 (43)	30.9 (17)	43.2 (201)
Sexual abuse	23.5 (76)	29.9 (26)	58.2 (32)	28.8 (134)
Problem drug				
Alcohol†	83.4 (271)	92.1 (82)	91.1 (51)	86.0 (404)
Opioids***	35.7 (116)	60.7 (54)	55.4 (31)	42.8 (201)
Sedatives***	15.4 (50)	36.0 (32)	50.0 (28)	23.4 (110)
Stimulants*	80.0 (260)	65.2 (58)	78.6 (44)	77.0 (362)

OD = overdose. Mean (S.D.) for continuous variables age, physical health status, social support, and depressive symptoms. Means with a different letter (a, b, c) are significantly different in pairwise comparisons. As an example, the mean age of the No OD and Prior Other OD do not significantly differ from each other (since they both are listed with letter a), but the mean Prior Unintentional OD is significantly higher than either of the two other groups. $^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 2
Characteristics of subjects with opioid problems admitted for detoxification (opioid subgroup)

Characteristic	% (*n*)/mean (S.D.)			
	No OD 57.7 (*n* = 116)	Prior Unintent OD 26.9 (*n* = 54)	Prior Other OD 15.4 (*n* = 31)	Total (*n* = 201)
Age*	34.3 a (7.8)	37.9 b (7.8)	33.5 a (7.8)	35.1 (7.9)
Male	72.4 (84)	85.2 (46)	67.7 (21)	75.1 (151)
Race†				
Black	35.3 (41)	24.1 (13)	16.1 (5)	29.4 (59)
White	44.8 (52)	59.3 (32)	48.4 (15)	49.3 (99)
Hispanic	13.8 (16)	11.1 (6)	32.3 (10)	15.9 (32)
Other	6.0 (7)	5.6 (3)	3.2 (1)	5.5 (11)
Homelessness**	34.5 (40)	59.3 (32)	58.1 (18)	44.8 (90)
Physical health status	46.5 (10.8)	45.7 (9.4)	45.0 (11.3)	46.1 (10.5)
Social support (family)	7.1 (4.5)	6.6 (4.8)	5.8 (4.5)	6.8 (4.6)
Social support (friends)	6.9 (4.0)	6.4 (3.8)	5.6 (3.8)	6.6 (4.0)
Depressive symptoms***	32.4 a (11.3)	35.6 a (11.3)	42.1 b (11.3)	34.8 (11.8)
Abuse history*				
No abuse	35.7 (41)	28.9 (15)	12.9 (4)	30.3 (60)
Physical abuse only	40.0 (46)	40.4 (21)	32.3 (10)	38.9 (77)
Sexual abuse	24.4 (28)	30.8 (16)	54.8 (17)	30.8 (61)
Problem drug				
Alcohol*	72.4 (84)	90.7 (49)	83.9 (26)	79.1 (159)
Opioids	100 (116)	100 (54)	100 (31)	100 (201)
Sedatives***	31.0 (36)	55.6 (30)	74.2 (23)	44.3 (89)
Stimulants	81.0 (94)	74.1 (40)	87.1 (27)	80.1 (161)

OD = overdose. Mean (S.D.) for continuous variables age, physical health status, social support, and depressive symptoms. Means with a different letter (a, b, c) are significantly different in pairwise comparisons. As an example, the mean age of the No OD and Prior Other OD do not significantly differ from each other (since they both are listed with letter a), but the mean Prior Unintentional OD is significantly higher than either of the two other groups. †$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Time-to-event (survival) models were constructed from statistically or clinically significant variables used in the cross-sectional, logistic regression analyses. OD rates during follow-up were based on Kaplan–Meier estimates. The second and third hypotheses were tested using proportional hazards regression models predicting time to first new report of any OD during follow-up (Table 3). The longitudinal models contained Prior OD (or Prior Unintentional OD and Prior Other OD); two time-varying covariates, recent (past 7 days) depressive symptoms and recent (past 30 days) *Drug Use*; and potential confounders: age, gender, race, homeless status, and physical or sexual abuse. The reports of OD were interval censored (possible values were 6 months, 12 months, 18 months, 24 months, or censored at last observed timepoint). Secondary analyses were conducted by replacing the *Drug Use* variables with a single, continuous, time-varying, variable representing the total number of drugs with any use in the past 30 days. In additional secondary analyses, lifetime (prior) suicidal ideation and lifetime (prior) suicide attempt were added to the cross-sectional and longitudinal models. Finally, in order to assess the robustness of our findings related to depressive symptoms, several sensitivity analyses were performed: (1) time-varying variables were time lagged by one visit in order to assure the temporal ordering of depressive symptoms and OD; and (2) a binary variable for depressive symptoms using a CES-D cutoff score of ≥23 (Golub et al., 2004) was substituted for the continuous CES-D variable. Two-sided $p < 0.05$ was considered statistically significant and SAS/STAT software version 8.2 (Cary, NC) was used to perform all analyses.

3. Results

3.1. Lifetime overdose by intent assessed at detoxification (study entry)

Characteristics of the total sample ($n = 470$) are depicted in Table 1. The three groups (No OD, Prior Unintentional OD, and Prior Other OD) significantly differed by age, gender, race, homeless status, physical health status, family social support, sexual/physical abuse, *Problem Drug*, and depressive symptoms (Table 1). Depressive symptoms were significantly elevated in both the Prior Other OD group and the Prior Unintentional OD group compared to the No OD group.

Characteristics of the Opioid *Problem Drug* subgroup ($n = 201$) can be found in Table 2. Within the Opioid *Problem Drug* subgroup, the three groups also significantly differed by age, homeless status, sexual/physical abuse history, *Problem Drug*, and depressive symptoms (Table 2). Although the results were not statistically significant, gender, race, physical health status, and social support-family differences were also similar to those seen in the entire cohort. Depressive symptoms were elevated in both the Prior Other OD group and the Prior Unintentional OD group compared to the No OD group; however, the difference between the Prior Unintentional OD group and the No OD group was not statistically significant.

3.2. Lifetime overdose: prevalence and risk factors

Lifetime prevalence for any OD ("prior OD") was 30.9% (145/470) in the total sample and 42.3% (85/201) for the Opioid *Problem Drug* subgroup. In logistic regression analysis, factors associated with lifetime (prior) OD in the total sample included: male (OR: 2.99, 95% CI: 1.52–5.90); white versus black race (OR: 1.97, 95% CI: 1.13–3.46); Hispanic ethnicity versus black race (OR: 2.82, 95% CI: 1.29–6.16); homelessness (OR: 2.21, 95% CI: 1.36–3.61); sexual abuse versus no abuse (OR: 3.95,

Table 3
Factors associated with overdose during prospective follow-up after detoxification[a]

Characteristic	HR (95% CI)			
	Total sample ($n = 397$)		Opioid subgroup ($n = 169$)	
	W/ prior OD	W/ prior OD by intent	W/ prior OD	W/ prior OD by intent
Prior OD	6.19 (3.23–11.88)***	NA	5.94 (2.67–13.18)***	NA
Prior Unintentional OD	NA	7.07 (3.50–14.29)***	NA	7.11 (3.06–16.50)***
Prior Other OD	NA	5.13 (2.37–11.08)***	NA	4.05 (1.45–11.32)**
Age	0.98 (0.95–1.02)	0.98 (0.95–1.02)	0.99 (0.94–1.03)	0.98 (0.94–1.03)
Male	1.52 (0.71–3.24)	1.44 (0.67–3.10)	1.72 (0.67–4.42)	1.57 (0.60–4.07)
Race				
White vs. black	3.89 (1.78–8.53)***	4.04 (1.84–8.87)***	2.97 (1.14–7.71)*	3.13 (1.19–8.26)*
Hispanic vs. black	2.37 (0.85–6.64)	2.45 (0.88–6.87)†	1.69 (0.47–6.07)	2.01 (0.54–7.41)
Other vs. black	0.95 (0.20–4.54)	0.94 (0.19–4.51)	0.48 (0.05–4.30)	0.45 (0.05–4.08)
Homelessness	0.85 (0.46–1.59)	0.86 (0.47–1.60)	0.68 (0.32–1.44)	0.68 (0.33–1.41)
Depressive symptoms[b]	1.05 (1.03–1.08)***	1.06 (1.03–1.08)***	1.04 (1.01–1.07)*	1.04 (1.01–1.07)**
Physical abuse only vs. no abuse	1.34 (0.59–3.04)	1.31 (0.58–2.98)	1.06 (0.42–2.68)	1.07 (0.42–2.73)
Sexual abuse vs. no abuse	2.63 (1.14–6.09)*	2.77 (1.20–6.42)*	1.66 (0.62–4.39)	1.95 (0.72–5.28)
Drug Use[b]				
Alcohol	1.03 (1.00–1.06)*	1.03 (1.00–1.06)†	1.02 (0.98–1.05)	1.01 (0.98–1.05)
Heroin	1.03 (1.00–1.06)†	1.02 (1.00–1.05)†	1.03 (1.00–1.06)†	1.03 (0.99–1.06)
Benzodiazepines	1.02 (0.98–1.06)	1.02 (0.98–1.06)	1.02 (0.97–1.07)	1.02 (0.97–1.07)
Cocaine	0.93 (0.88–0.99)*	0.94 (0.88–0.99)*	0.95 (0.89–1.01)	0.95 (0.89–1.02)

HR = hazard ratio; CI = confidence interval. †$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

[a] Proportional hazards models control for all covariates above.

[b] Time-varying covariates (depressive symptoms: past 7 days and *Drug Use*: # days past 30).

95% CI: 1.96–7.98); physical abuse only versus no abuse (OR: 2.02, 95% CI: 1.10–3.72); Stimulant *Problem Drug* (OR: 0.56, 95% CI: 0.32–1.00); Opioid *Problem Drug* (OR: 2.21, 95% CI: 1.27–3.85); and more depressive symptoms (OR: 1.04, 95% CI: 1.02–1.06). Here the OR represents the change for each point increase in the CES-D score. In a secondary analysis, prior suicide attempt (OR: 3.20, 95% CI: 1.76–5.81), but not prior suicidal ideation (OR: 1.35, 95% CI: 0.78–2.33), was associated with OD when added to the model. The total number of drugs listed on the ASI that fulfilled the *Problem Drug* criteria was also associated with OD (OR: 1.15, 95% CI: 1.03–1.29).

In the Opioid *Problem Drug* subgroup, more depressive symptoms (OR: 1.03, 95% CI: 1.00–1.07) and Sedative *Problem Drug* (OR: 2.27, 95% CI: 1.13–4.54) were associated with lifetime (prior) OD. As in the total sample, secondary analysis showed that prior suicide attempt (OR: 2.80, 95% CI: 1.21–6.47), not prior suicidal ideation (OR: 1.37, 95% CI: 0.64–2.93), was associated with prior OD. The total number of *Problem Drug* categories was also associated with prior OD (OR: 1.20, 95% CI: 1.01–1.44).

3.3. Overdose after detoxification (follow-up)

During the 2-year follow-up, Kaplan–Meier method estimated that 16.9% (95% CI: 13.0–20.9) of the total sample ($n = 397$, 60 individuals with at least 1 OD) and 26.7% (95% CI: 19.5–33.9) of the Opioid *Problem Drug* subgroup ($n = 169$, 40 individuals with at least 1 OD) reported an OD. Among individuals with a prior OD, the KM estimate was that 43.5% (95% CI: 33.7–53.3) of subjects in the total sample ($n = 120$) and 52.4% (95% CI: 38.6–66.1) of the Opioid *Problem Drug* subgroup ($n = 70$), OD'd again within 2 years after detoxification. Among individuals without a prior OD, the estimated 24-month incidence in the total sample ($n = 277$) and the 24-month incidence in the Opioid *Problem Drug* subgroup ($n = 99$) was 5.7% (95% CI: 2.8–8.7) and 11.0% (95% CI: 4.5–17.4), respectively.

In the total sample, predictors of OD during the 2-year follow-up in the proportional hazards regression analysis included white race, past sexual abuse, less frequent cocaine use, more frequent alcohol use, recent depressive symptoms, and prior OD regardless of intent (Table 3). More frequent heroin use was marginally significant ($p < 0.10$). Interestingly, Prior Unintentional OD was the strongest predictor of future OD (HR: 7.07, 95% CI: 3.50–14.29). Recent depressive symptoms were significantly associated with OD (HR 1.05, 95% CI: 1.03–1.08 for each point increase in the CES-D score). When entered into the model, depressive symptoms ≥ 23 were associated with increased risk of OD (HR: 3.94, 95% CI: 1.88–8.24). Time lagged analyses revealed similar results for depressive symptoms both as a continuous (HR: 1.07, 95% CI: 1.01–1.13) and dichotomous variable (HR: 3.79, 95% CI: 0.62–23.26), although limited power may have resulted in loss of statistical significance. A secondary analysis showed that neither prior suicide attempt (HR: 1.27, 95% CI: 0.68–2.35) nor prior suicidal ideation (HR: 1.07, 95% CI: 0.62–1.87) was significantly associated with future OD. Additional secondary analyses demonstrated that the

total number of drugs with any use in the past 30 days was not significantly associated with OD (HR: 1.08, 95% CI: 0.91–1.29).

In the Opioid *Problem Drug* subgroup, white race, recent depressive symptoms, and prior OD regardless of intent (Table 3) predicted OD during follow-up. Although there were no statistically significant associations between frequency of recent *Drug Use* (heroin use was marginally significant) and OD, odds ratios were comparable to the *Drug Use* variables from the entire cohort. Prior Unintentional OD was again the strongest predictor of future OD (HR: 7.11, 95% CI: 3.06–16.50). Recent depressive symptoms were significantly associated with OD (HR: 1.04, 95% CI: 1.01–1.07). When entered into the model, depressive symptoms ≥ 23 were associated with increased risk of OD (HR: 2.33, 95% CI: 1.02–5.34). Notwithstanding loss of statistical significance due to power limitations, time lagged analyses again demonstrated comparable findings for depressive symptoms both as a continuous (HR: 1.06, 95% CI: 0.98–1.14) and binary predictor (HR: 4.89, 95% CI: 0.42–57.76). Like the total sample, neither prior suicide attempt (HR: 0.89, 95% CI: 0.38–2.06) nor prior suicidal ideation (HR: 0.89, 95% CI: 0.41–1.92) was significantly associated with future OD. Finally, the total number of drugs with any use in the past 30 days was not significantly associated with OD (HR: 1.15, 95% CI: 0.93–1.41).

4. Discussion

Findings about these detoxification unit patients provide insight regarding OD as a leading cause of death among those with alcohol and other drug problems. Self-reports revealed the high prevalence of OD in detoxification patients (31%), especially problem opioid users (42%). Risk factors for OD were similar for the total sample and the Opioid *Problem Drug* subgroup, suggesting that some OD risk-identification efforts (e.g., screening for depressive symptoms) could apply universally to patients admitted for detoxification. This study extends previous research by demonstrating that prior OD among detoxification patients is a strong predictor of future OD. It is also noteworthy that the association of elevated depressive symptoms and OD, noted previously in a retrospective study of a community sample (Tobin and Latkin, 2003), was observed prospectively in our treatment population. Overall, these data suggest that some OD considered "accidental" may not be completely unintentional.

OD is a common occurrence among detoxification patients in general (Ravndal and Vaglum, 1999) and opioid users in particular (Darke et al., 1996a; Gossop et al., 1996; Neale and Robertson, 2005; Ravndal and Vaglum, 1999; Seal et al., 2001). Our definition of OD, which can be applied to any drug with an OD risk, has face validity (Powis et al., 1999) and by requiring emergency medical treatment suggests that non-fatal OD results in considerable service utilization and healthcare cost. However, direct comparison to studies using other OD definitions (e.g., more drug-specific definitions (Darke et al., 1996a)) may be difficult. OD rates after detoxification in our study are similar to the study by Ravndal and Vaglum (1999), which also sampled patients with diverse drug preferences (17% over 2 years versus 29% over 5 years). Additionally, OD rates after discharge for the Opioid *Problem Drug* subgroup (27% over 2 years) are comparable to the estimated annual rates for non-fatal OD among heroin users in Australia, which range from 19% to 30% (Darke et al., 2003). While it should be emphasized that the degree to which risk factors for non-fatal and fatal OD overlap is currently unknown (Powis et al., 1999), the increased risk of OD from opioids is consistent with some toxicology studies of fatal OD (Coffin et al., 2003; Gossop et al., 2002). Opioids' ability to depress the respiratory center in the brain is well known and most likely represents the mechanisms of action for opioid-related OD (White and Irvine, 1999).

Prior suicide attempt is one of the strongest predictors of both future suicide attempt (Preuss et al., 2003; Wines et al., 2004) and completed suicide (Harris and Barraclough, 1997); however, less is known regarding quantifying the risk of prior non-fatal OD (elevated six-fold in our study) on future non-fatal OD. Results are consistent with a retrospective study of 312 injection drug users that found that each OD increased the risk of having another OD (Powis et al., 1999). Our finding of Prior Unintentional OD as the strongest predictor of OD was somewhat unexpected, but important, as it appears to reinforce both the preventable nature of OD and the need for effective interventions in high-risk individuals with a history of OD (Powis et al., 1999). The frequency of multiple drug use in our cohort and results from both the *Problem Drug* and *Drug Use* variables are generally consistent with the polydrug theory (Darke and Zador, 1996) of opioid-related OD and recent findings from a case-crossover study by Dietze et al. (2005). That stimulants may be associated with less risk should be interpreted with caution as stimulants like cocaine have been associated with both non-fatal (Kaye and Darke, 2004) and fatal OD (Coffin et al., 2003), and the effect measure (odds ratio) for stimulants in our study represents a relative comparison to other drugs, not a comparison to a no drug control.

Our findings regarding the association of depressive symptoms and OD are very similar to the study by Tobin and Latkin (2003). Although depressive symptoms, a modifiable risk factor, may represent a potential intervention target to prevent OD, their exact role remains uncertain, as several explanations may account for the association between depressive symptoms and OD. Specifically, depressive symptoms may increase in the aftermath of an OD; however, several studies have shown that these symptoms actually improve after an intentional OD (Jack and Williams, 1991; Sarfati et al., 2003). Chronically elevated depressive symptoms may also form a trait-risk for OD, yet depressive symptoms were not increased among individuals with an OD that occurred more than a year previous (Tobin and Latkin, 2003). In addition, individuals may indulge in more risky drug use as an attempt to self-medicate dysphoric mood states (Chilcoat and Breslau, 1998; Khantzian, 1985; Wines et al., 2004), resulting from a mood disorder or other source (e.g., medical illness) (DSM-IV, 1994; Raimo and Schuckit, 1998). Conversely, drug use may cause the depressive symptoms (substance-induced depression) (Brown and Schuckit, 1988; Ries et al., 2001) leading to the OD (Preuss et al., 2003; Wines et al., 2004). Individuals relapsing after extended periods of abstinence may experience extreme guilt and hopelessness (abstinence violation effect) (Marlatt, 1985) that when coupled

with decreased drug tolerance (Strang et al., 2003) may substantially enhance the risk of OD. Finally, acute life stressors (e.g., relationship disruptions) that have been associated with OD (Neale and Robertson, 2005) could conceivably be mediated by depressive symptoms (Hammen et al., 1986). Regardless of their cause, abrupt increases in depressive symptoms occurring just prior to the OD strongly argues that worsening dysphoria may lie on a causal pathway for OD and as such represent a prime intervention target. Determining whether these transient mood states can trigger an OD or if they are related to some other factor outside the causal pathway requires further research, perhaps using a case-crossover design which is capable of quantifying the impact of rapid mood shifts (Mittleman et al., 1995).

In light of our study results, it is noteworthy that some investigators have argued that depression can be an implicit marker of intent for suicide (Rosenberg et al., 1988). In addition, several researchers (Farrell et al., 1996; Neale, 2000; Rossow and Lauritzen, 1999; Zador, 2005) have suggested that intent for OD may lie on a continuum, and we found some indirect support for this theory. In the Prior Unintentional OD group, the proportion of patients with a sexual abuse history, a well-known risk factor for suicidal behavior (Molnar et al., 2001), was higher than the No OD group, but lower than the Prior Other OD group, which almost certainly contained some subjects with intentional ODs. Depressive symptoms and perceived family support were similarly distributed. Notably, a definition of a suicide attempt commonly used in the US requires that the self-injurious behavior occur with some "nonzero" level of intent to kill oneself (O'Carroll et al., 1996). Given the importance of the accurate determination of intent, rigorously designed studies of individuals with severe opioid-related OD, using standardized suicide-related assessments are needed to further characterize drug-related OD, develop more sensitive intent scales, and explore the notion of a continuum of intent.

Despite several strengths of our study including the longitudinal design utilizing a relatively large, diverse cohort and time-to-event analyses, there are some important limitations. First, our definition of OD has both disadvantages and advantages. It may have underestimated the true prevalence and incidence of OD as a substantial percentage of ODs (approximately 50% or more in some studies) do not result in emergency medical treatment (Darke et al., 1996b; Davidson et al., 2002; Powis et al., 1999). Furthermore, our findings may have been influenced by selecting individuals most likely to utilize medical services. At the same time as our OD definition likely sampled more severe ODs, it probably yielded more accurate effect estimates via improved specificity of the outcome. In order to more fully address these important issues, further methodological development and psychometric testing of differing outcome definitions used in OD research is warranted. Second, there is potential misclassification between the Prior Unintentional OD and Prior Other OD groups, in that some subjects in the Prior Other OD group may have had only an unintentional OD. Given the exploratory nature of the analyses pertaining to intent and suicidal behavior, results should be considered preliminary and in need of replication. Third, some standardized assessments used refer to overlapping but not identical time periods during follow-up (e.g., CES-D: past 1 week, ASI: past 1 month), making the exact sequencing of depressive symptoms and OD difficult to ascertain. Nonetheless, time lagged sensitivity analyses revealed comparable findings. Fourth, formal diagnostic instruments were not used to make specific substance use disorder and major depressive disorder diagnoses; however, subjects in our study almost certainly met substance dependence or abuse criteria given scores on addiction measurements like the ASI coupled with admission criteria for the detoxification unit. Fifth, not all subjects were observed at all follow-up timepoints, and we were only able to estimate time to new OD in 6-month intervals. Finally, results from our study of detoxification patients may not generalize to other populations (e.g., drug users not in treatment), yet the magnitude of OD risk from elevated depressive symptoms in our study was very similar (same adjusted HR and OR) to data from a community sample of drug users (Tobin and Latkin, 2003).

In summary, the current study illustrates the high prevalence of OD in urban detoxification patients, the potency of prior OD as a risk factor for future OD, and the potential role of worsening depression in the etiology of OD. We also found preliminary evidence that some "accidental" ODs may not be completely unintentional. Regardless of whether or not an OD event is considered intentional, our data suggest that the effective treatment of depressive symptoms may hold potential for the prevention of OD. There is a striking lack of OD risk-recognition, prevention, and treatment efforts (Cook et al., 1998; Pollini et al., 2006). Addiction specialists, especially those in detoxification settings, could consider screening patients with a simple OD question, "Have you ever overdosed?" The increasing rates of OD in the US (CDC, 2004; Paulozzi, 2006; Paulozzi et al., 2006) highlight the urgent need for research focused on this growing public health problem, including nationally representative, longitudinal studies containing valid and reliable outcome measures for unintentional OD and suicidal behavior.

Acknowledgements

This work was supported by grants from the National Institute on Drug Abuse. "Drug-Related Suicidal and/or Homicidal Behavior," K23-DA00455 (JDW), "Enhanced Linkage of Drug Abusers to Primary Medical Care," R01-DA10019 and the National Institute on Alcohol Abuse and Alcoholism, "Enhanced Linkage of Alcohol Abusers to Primary Care," R01-AA10870. This work was also supported in part by a General Clinical Research Center grant MO1-RR00533 from the National Center for Research Resources.

References

Ballesteros, M.F., Budnitz, D.S., Sanford, C.P., Gilchrist, J., Agyekum, G.A., Butts, J., 2003. Increase in deaths due to methadone in North Carolina. JAMA 290, 40.

Best, D., Gossop, M., Man, L., Finch, E., Greenwood, J., Strang, J., 2000. Accidental and deliberate overdose among opiate addicts in methadone maintenance treatment: are deliberate overdoses systematically different? Drug Alcohol Rev. 19, 213–216.

Brown, S.A., Schuckit, M.A., 1988. Changes in depression among abstinent alcoholics. J. Stud. Alcohol 49, 412–417.

Centers for Disease Control Prevention, 2004. Unintentional and undetermined poisoning deaths—11 states, 1990–2001. Morb. Mortal. Wkly Rep. 53, 233–238.

Centers for Disease Control and Prevention, 2006. Web-based Injury Statistics Query and Reporting System (WISQARS). Available at: www.cdc.gov/ncipc/wisqars/default.htm (accessed April 4, 2006).

Chilcoat, H.D., Breslau, N., 1998. Posttraumatic Stress Disorder and drug disorders: testing causal pathways. Arch. Gen. Psychiatry 55, 913–917.

Coffin, P.O., Galea, S., Ahern, J., Leon, A.C., Vlahov, D., Tardiff, K., 2003. Opiates, cocaine and alcohol combinations in accidental drug overdose deaths in New York City, 1990–98. Addiction 98, 739–747.

Compton, W.M., Volkow, N.D., 2006. Major increases in opioid analgesic abuse in the United States: concerns and strategies. Drug Alcohol Depend. 81, 103–107.

Cook, S., Moeschler, O., Michaud, K., Yersin, B., 1998. Acute opiate overdose: characteristics of 190 consecutive cases. Addiction 93, 1559–1565.

Darke, S., Hall, W., 2003. Heroin overdose: research and evidence-based intervention. J. Urban Health 80, 189–200.

Darke, S., Mattick, R.P., Degenhardt, L., 2003. The ratio of non-fatal to fatal heroin overdose. Addiction 98, 1169–1171.

Darke, S., Ross, J., 2001. The relationship between suicide and heroin overdose among methadone maintenance patients in Sydney, Australia. Addiction 96, 1443–1453.

Darke, S., Ross, J., 2002. Suicide among heroin users: rates, risk factors and methods. Addiction 97, 1383–1394.

Darke, S., Ross, J., Hall, W., 1996a. Overdose among heroin users in Sydney, Australia. I. Prevalence and correlates of non-fatal overdose. Addiction 91, 405–411.

Darke, S., Ross, J., Hall, W., 1996b. Overdose among heroin users in Sydney, Australia. II. Responses to overdose. Addiction 91, 413–417.

Darke, S., Ross, J., Lynskey, M., Teesson, M., 2004. Attempted suicide among entrants to three treatment modalities for heroin dependence in the Australian Treatment Outcome Study (ATOS): prevalence and risk factors. Drug Alcohol Depend. 73, 1–10.

Darke, S., Zador, D., 1996. Fatal heroin 'overdose': a review. Addiction 91, 1765–1772.

Davidson, P.J., Ochoa, K.C., Hahn, J.A., Evans, J.L., Moss, A.R., 2002. Witnessing heroin-related overdoses: the experiences of young injectors in San Francisco. Addiction 97, 1511–1516.

Davoli, M., Perucci, C.A., Rapiti, E., Bargagli, A.M., D'Ippoliti, D., Forastiere, F., Abeni, D., 1997. A persistent rise in mortality among injection drug users in Rome, 1980 through 1992. Am. J. Public Health 87, 851–853.

Dietze, P., Jolley, D., Fry, C., Bammer, G., 2005. Transient changes in behaviour lead to heroin overdose: results from a case-crossover study of non-fatal overdose. Addiction 100, 636–642.

DSM-IV, 1994. Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association, Washington, DC.

Farrell, M., Neeleman, J., Griffiths, P., Strang, J., 1996. Suicide and overdose among opiate addicts. Addiction 91, 321–323.

Frischer, M., Bloor, M., Goldberg, D., Clark, J., Green, S., McKeganey, N., 1993. Mortality among injecting drug users: a critical reappraisal. J. Epidemiol. Community Health 47, 59–63.

Golub, E.T., Latka, M., Hagan, H., Havens, J.R., Hudson, S.M., Kapadia, F., Campbell, J.V., Garfein, R.S., Thomas, D.L., Strathdee, S.A., 2004. Screening for depressive symptoms among HCV-infected injection drug users: examination of the utility of the CES-D and the Beck Depression Inventory. J. Urban Health 81, 278–290.

Gossop, M., Griffiths, P., Powis, B., Williamson, S., Strang, J., 1996. Frequency of non-fatal heroin overdose: survey of heroin users recruited in non-clinical settings. BMJ 313, 402.

Gossop, M., Stewart, D., Treacy, S., Marsden, J., 2002. A prospective study of mortality among drug misusers during a 4-year period after seeking treatment. Addiction 97, 39–47.

Hammen, C., Mayol, A., deMayo, R., Marks, T., 1986. Initial symptom levels and the life-event-depression relationship. J. Abnorm. Psychol. 95, 114–122.

Harris, E.C., Barraclough, B., 1997. Suicide as an outcome for mental disorders. A meta-analysis. Br. J. Psychiatry 170, 205–228.

Hser, Y.I., Hoffman, V., Grella, C.E., Anglin, M.D., 2001. A 33-year follow-up of narcotics addicts. Arch. Gen. Psychiatry 58, 503–508.

Hulse, G.K., English, D.R., Milne, E., Holman, C.D., 1999. The quantification of mortality resulting from the regular use of illicit opiates. Addiction 94, 221–229.

Jack, R.L., Williams, J.M., 1991. The role of attributions in self-poisoning. Br. J. Clin. Psychol. 30 (Pt 1), 25–35.

Kaye, S., Darke, S., 2004. Non-fatal cocaine overdose among injecting and non-injecting cocaine users in Sydney, Australia. Addiction 99, 1315–1322.

Khantzian, E.J., 1985. The self-medication hypothesis of addictive disorders: focus on heroin and cocaine dependence. Am. J. Psychiatry 142, 1259–1264.

Kjelsberg, E., Winther, M., Dahl, A.A., 1995. Overdose deaths in young substance abusers: accidents or hidden suicides? Acta Psychiatr. Scand. 91, 236–242.

Kosten, T.R., Rounsaville, B.J., 1988. Suicidality among opioid addicts: 2.5 year follow-up. Am. J. Drug Alcohol Abuse 14, 357–369.

Landen, M.G., Castle, S., Nolte, K.B., Gonzales, M., Escobedo, L.G., Chatterjee, B.F., Johnson, K., Sewell, C.M., 2003. Methodological issues in the surveillance of poisoning, illicit drug overdose, and heroin overdose deaths in New Mexico. Am. J. Epidemiol. 157, 273–278.

Liebschutz, J., Savetsky, J.B., Saitz, R., Horton, N.J., Lloyd-Travaglini, C., Samet, J.H., 2002. The relationship between sexual and physical abuse and substance abuse consequences. J. Subst. Abuse Treat. 22, 121–128.

Marlatt, G.A., 1985. Relapse prevention: theoretical rationale and overview of the model. In: Marlatt, G.A., Gordon, J.R. (Eds.), Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. Guilford Press, New York, pp. 3–70.

McLellan, A.T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., Argeriou, M., 1992. The fifth edition of the Addiction Severity Index. J. Subst. Abuse Treat. 9, 199–213.

Mittleman, M.A., Maclure, M., Sherwood, J.B., Mulry, R.P., Tofler, G.H., Jacobs, S.C., Friedman, R., Benson, H., Muller, J.E., 1995. Triggering of acute myocardial infarction onset by episodes of anger. Determinants of Myocardial Infarction Onset Study Investigators. Circulation 92, 1720–1725.

Molnar, B.E., Berkman, L.F., Buka, S.L., 2001. Psychopathology, childhood sexual abuse and other childhood adversities: relative links to subsequent suicidal behaviour in the US. Psychol. Med. 31, 965–977.

Murphy, S.L., Rounsaville, B.J., Eyre, S., Kleber, H.D., 1983. Suicide attempts in treated opiate addicts. Compr. Psychiatry 24, 79–89.

Neale, J., 2000. Suicidal intent in non-fatal illicit drug overdose. Addiction 95, 85–93.

Neale, J., Robertson, M., 2005. Recent life problems and non-fatal overdose among heroin users entering treatment. Addiction 100, 168–175.

O'Carroll, P.W., Berman, A.L., Maris, R.W., Moscicki, E.K., Tanney, B.L., Silverman, M.M., 1996. Beyond the Tower of Babel: a nomenclature for suicidology. Suicide Life Threat. Behav. 26, 237–252.

Ohberg, A., Lonnqvist, J., 1998. Suicides hidden among undetermined deaths. Acta Psychiatr. Scand. 98, 214–218.

O'Malley, S.S., Jaffe, A.J., Chang, G., Schottenfeld, R.S., Meyer, R.E., Rounsaville, B., 1992. Naltrexone and coping skills therapy for alcohol dependence. A controlled study. Arch. Gen. Psychiatry 49, 881–887.

Oppenheimer, E., Tobutt, C., Taylor, C., Andrew, T., 1994. Death and survival in a cohort of heroin addicts from London clinics: a 22-year follow-up study. Addiction 89, 1299–1308.

Paulozzi, L.J., 2006. Opioid analgesic involvement in drug abuse deaths in American metropolitan areas. Am. J. Public Health 96, 1755–1757.

Paulozzi, L.J., Budnitz, D.S., Xi, Y., 2006. Increasing deaths from opioid analgesics in the United States. Pharmacoepidemiol. Drug Saf. 15, 618–627.

Pollini, R.A., McCall, L., Mehta, S.H., Vlahov, D., Strathdee, S.A., 2006. Nonfatal overdose and subsequent drug treatment among injection drug users. Drug Alcohol Depend. 83, 104–110.

Powis, B., Strang, J., Griffiths, P., Taylor, C., Williamson, S., Fountain, J., Gossop, M., 1999. Self-reported overdose among injecting drug users in London: extent and nature of the problem. Addiction 94, 471–478.

Preuss, U.W., Schuckit, M.A., Smith, T.L., Danko, G.P., Bucholz, K.K., Hesselbrock, M.N., Hesselbrock, V., Kramer, J.R., 2003. Predictors and correlates of suicide attempts over 5 years in 1,237 alcohol-dependent men and women. Am. J. Psychiatry 160, 56–63.

Procidano, M.E., Heller, K., 1983. Measures of perceived social support from friends and from family: three validation studies. Am. J. Community Psychol. 11, 1–24.

Quaglio, G., Talamini, G., Lechi, A., Venturini, L., Lugoboni, F., Mezzelani, P., 2001. Study of 2708 heroin-related deaths in north-eastern Italy 1985–98 to establish the main causes of death. Addiction 96, 1127–1137.

Radloff, L.S., 1977. The CES-D scale: a self-report depression scale for research in the general population. App. Psych. Meas. 1, 385–401.

Raimo, E.B., Schuckit, M.A., 1998. Alcohol dependence and mood disorders. Addict. Behav. 23, 933–946.

Ravndal, E., Vaglum, P., 1999. Overdoses and suicide attempts: different relations to psychopathology and substance abuse? A 5-year prospective study of drug abusers. Eur. Addict. Res. 5, 63–70.

Ries, R.K., Demirsoy, A., Russo, J.E., Barrett, J., Roy-Byrne, P.P., 2001. Reliability and clinical utility of DSM-IV substance-induced psychiatric disorders in acute psychiatric inpatients. Am. J. Addict. 10, 308–318.

Rosenberg, M.L., Davidson, L.E., Smith, J.C., Berman, A.L., Buzbee, H., Gantner, G., Gay, G.A., Moore-Lewis, B., Mills, D.H., Murray, D., O'Carroll, P.W., Jobes, D., 1988. Operational criteria for the determination of suicide. J. Forensic Sci. 33, 1445–1456.

Rossow, I., Lauritzen, G., 1999. Balancing on the edge of death: suicide attempts and life-threatening overdoses among drug addicts. Addiction 94, 209–219.

Samet, J.H., Larson, M.J., Horton, N.J., Doyle, K., Winter, M., Saitz, R., 2003. Linking alcohol- and drug-dependent adults to primary medical care: a randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. Addiction 98, 509–516.

Sanford, C., 2002. Deaths from Unintentional Drug Overdoses in North Carolina, 1997-2001: A DHHS Investigation into Unintentional Poisoning-related Deaths. North Carolina Department of Health and Human Services, Raleigh, Available at: www.communityhealth.dhhs.state.nc.us/Injury/unintentional%20poisonings%20report-9-02-final.pdf (accessed April 4, 2006).

Sarfati, Y., Bouchaud, B., Hardy-Bayle, M.C., 2003. Cathartic effect of suicide attempts not limited to depression: a short-term prospective study after deliberate self-poisoning. Crisis 24, 73–78.

Seal, K.H., Kral, A.H., Gee, L., Moore, L.D., Bluthenthal, R.N., Lorvick, J., Edlin, B.R., 2001. Predictors and prevention of nonfatal overdose among street-recruited injection heroin users in the San Francisco Bay Area, 1998–1999. Am. J. Public Health 91, 1842–1846.

Sorg, M.H, Greenwald, M., 2002. Maine Drug-related Mortality Patterns: 1997–2002. Office of the Maine Attorney General, Augusta, Available at: www.mainegov-images.informe.org/ag/pr/drugreport.pdf (accessed April 4, 2006).

Strang, J., McCambridge, J., Best, D., Beswick, T., Bearn, J., Rees, S., Gossop, M., 2003. Loss of tolerance and overdose mortality after inpatient opiate detoxification: follow up study. BMJ 326, 959–960.

Tobin, K.E., Latkin, C.A., 2003. The relationship between depressive symptoms and nonfatal overdose among a sample of drug users in Baltimore, Maryland. J. Urban Health 80, 220–229.

Vingoe, L., Welch, S., Farrell, M., Strang, J., 1999. Heroin overdose among a treatment sample of injecting drug misusers: accident or suicidal behaviour? J. Substance Use 4, 88–91.

Volpicelli, J.R., Alterman, A.I., Hayashida, M., O'Brien, C.P., 1992. Naltrexone in the treatment of alcohol dependence. Arch. Gen. Psychiatry 49, 876–880.

Ware, J.E., 1993. SF-36 Health Survey Manual and Interpretation Guide. The Health Institute, Boston, MA.

Warner-Smith, M., Darke, S., Lynskey, M., Hall, W., 2001. Heroin overdose: causes and consequences. Addiction 96, 1113–1125.

White, J.M., Irvine, R.J., 1999. Mechanisms of fatal opioid overdose. Addiction 94, 961–972.

Wines Jr., J.D., Saitz, R., Horton, N.J., Lloyd-Travaglini, C., Samet, J.H., 2004. Suicidal behavior, drug use and depressive symptoms after detoxification: a 2-year prospective study. Drug Alcohol Depend. 76 (Suppl.), S21–S29.

Zacny, J., Bigelow, G., Compton, P., Foley, K., Iguchi, M., Sannerud, C., 2003. College on Problems of Drug Dependence taskforce on prescription opioid non-medical use and abuse: position statement. Drug Alcohol Depend. 69, 215–232.

Zador, D., 2005. Reducing non-fatal overdoses among heroin users as an all-of-society challenge. Addiction 100, 141–142.