General Internal Medicine
Boston University School of Medicine
2005 Publications

1. Aagaard EM, Julian K, **Dedier J**, Soloman I, Tillisch J, Pérez-Stable J. Factors affecting medical students' selection of an internal medicine residency program. J Natl Med Assoc. 2005;97:1264-1270.
2. **Aggarwal A**, **Carr P**. Work-up of women with anemia. Women's health in primary care. 2005; 8:307-314.
3. Barnett RC, Gareis KC, **Carr PL**. Career satisfaction and retention of a sample of women physicians who work reduced hours. J Women's Health. 2005;14:146-53.
4. **Berlowitz DR**, **Ash AS**, Glickman M, **Friedman RH**, Pogach LM, Nelson AL, Wong AT. Developing a quality measure for clinical inertia in diabetes care. Health Services Review. 2005;40:1836-53.
5. Boehmer U, **Freund KM**, 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. 2005;59:307-314.
6. Boehmer U, Linde R., **Freund KM**. Sexual minority women's coping and psychological adjustment after a diagnosis of breast cancer. J Women's Health. 2005;14:214-24
7. **Carr PL**, Rothberg MB, **Friedman RH**, Felsenstein D, Pliskin JS. "Shotgun" versus sequential testing. Cost-effectiveness of diagnostic strategies for vaginistis. J Gen Intern Med. 2005;72:182-6.
8. **Chuang CH**, **Freund KM**. Emergency contraception knowledge among women in a Boston community. Contraception 2005;71:157-160.
9. **Chuang CH**, **Freund KM**. Emergency contraception: an intervention on primary care providers. Contraception 2005;72:182-186.
10. **David MM**, **Ko L**, Prudent N, **Green EH**, **Posner MA**, **Freund KM**. Mammography Use. J Natl Med Assoc. 2005;67:253-260.
11. **De Alba I**, Hubbell FA, McMullin JM, Sweningson JM, **Saitz R**. Impact of U.S. citizenship status on cancer screening among immigrant women. J Gen Intern Med. 2005;20:290-6.
12. Dworetzky BA, Strahonja-Packard A, **Shanahan CW**, Paz J, Schauble B, Bromfield EB. Characteristics of male veterans with psychogenic nonepileptic seizures. Epilepsia. 2005;46:1418-22.
13. Fang MC, Singer DE, Chang Y, **Hylek EM**, **Henault LE**, Jensvold NG, Go AS. Gender differences in the risk of ischemic stroke and peripheral embolism in atrial fibrillation. Circulation. 2005;112:1687-1691.
14. **Farzanfar R**, **Frishkopf S**, **Migneault J**, **Friedman R**. Telephone-linked care for physical activity: a qualitative evaluation of the use patterns of an information technology program for patients. J Biomed Inform. 2005;38:220-228.
15. **Freiberg MS**, **Samet JH**. Alcohol and coronary heart disease: the answer awaits a randomized controlled trial. Circulation. 2005;112:1379-1381.

16. Garcia D, Regan S, Crowther M, Hughes RA, **Hylek EM**. Warfarin maintenance dosing patterns in clinical practice: implications for safer anticoagulation in the elderly population. Chest. 2005;127:2049-2056.

17. **Green EH**, **Freund KM**, **Posner MA**, **David MM**. Pap smear rates among Haitian immigrant women in eastern Massachusetts. Public Health Rep. 2005;120:133-139.

18. Holmboe ES, Bowen JL, Green M, Gregg J, DiFrancesco L, Reynolds E, Alguire P, **Battinelli D**, Lucey C, Duffy D. Reforming internal medicine residency training: a report from the Society of General Internal Medicine's task force for residency reform. J Gen Intern Med. 2005;20:1-8.

19. Isaacson JH, Hopper JA, **Alford DP**, Parran T. Prescription drug use and abuse: risk factors, red flags and prevention strategies. Postgraduate Medicine. 2005;118:119-26.

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

21. Kraemer KL, Roberts MS, **Horton NJ**, Palfai T, **Samet JH**, Freedner N, Tibbetts N, **Saitz R**. Health utility ratings for a spectrum of alcohol-related health states. Med Care. 2005;43:541-50.

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

23. **Liebschutz, JM**, Geier JL, **Horton NJ**, **Chuang CH**, **Samet JH**. Physical and sexual violence and health care utilization in HIV-infected persons with alcohol problems. AIDS Care. 2005;17:566-578.

24. Linden JA, Oldeg P, Mehta SD, McCabe KK, **LaBelle C**. HIV postexposure prophylaxis in sexual assault: current practice and patient adherence to treatment recommendations in a large urban teaching hospital. Acad Emerg Med. 2005;12:640-646.

25. McHugo GJ, Caspi Y, Kammerer N, Mazelis R, Jackson EW, Russell L, Clark C, **Liebschutz J**, Kimerling R. The assessment of trauma history in women with co-occurring substance abuse and mental health disorders and a history of interpersonal violence. J Behav Health Serv Res. 2005;32:113-27.

26. **Murabito JM**, Evans JC, D'Agostino RB, Wilson PWF, Kannel WB. Temporal trends in the incidence of intermittent claudication from 1950 to 1999. Am J Epidemiol. 2005;162:1-8.

27. **Murabito JM**, Pencina MJ, Nam B, D'Agostino RB, Wang TJ, Lloyd-Jones D, Wilson PWF; O'Donnell CJ. Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. JAMA. 2005;294:3117-3123.

28. Norredam M, **Crosby S**, Munarriz R, Piwowarczyk L, Grodin M. Urologic complications of sexual trauma among male survivors of torture. Urology. 2005;65:28-32.

29. Olsen Y, **Alford DP, Horton NJ**, **Saitz R**. Addressing smoking cessation in methadone programs. J Addict Dis. 2005;24:33-48.

30. **Paasche-Orlow MK**. "The challenges of informed consent for low literate populations" in Schwartzenberg J, Wang C. Understanding Health Literacy: Implications for Medicine and Public Health. Chicago:AMA Press, 2005

31. **Paasche-Orlow MK**, Brancati FL. Assessment of medical school institutional review board policies regarding compensation of subjects for research-related injury. Am J Med. 2005;118:177-180.

32. **Paasche-Orlow MK**, Clarke JG, Hebert MR, Ray MK, Stein MD. Educational attainment but not literacy is associated with HIV risk behavior among incarcerated women. J Women's Health. 2005;14:852-859.

33. **Paasche-Orlow MK,** Jacob DM Powell JN. Notices of privacy practices: a survey of the health insurance portability and accountability act of 1996 documents presented to patients at U.S. hospitals. Med Care. 2005;43:558-564.

34. **Paasche-Orlow MK**, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med. 2005;20:175-184.

35. **Paasche-Orlow MK**, Riekert KA, Bilderback A, Chanmugam A, Hill P, Rand CS, Brancati FL, Krishnan JA. Tailored education may reduce health literacy disparities in asthma self-management. Am J Respir Crit Care Med. 2005;172:980-986.

36. Palepu A, **Horton NJ**, Tibbetts N, **Meli S**, **Samet JH.** Substance abuse treatment and hospitalization among a cohort of HIV-infected individuals with alcohol problems. Alcohol Clin Exp Res. 2005;29:389-394.

37. Palepu A, Raj A, **Horton NJ**, Tibbetts N, **Meli S**, **Samet JH**. Substance abuse treatment and risk behaviors among HIV-infected persons with alcohol problems. J Subst Abuse Treat. 2005;28:3-9.

38. Richter KP, Choi WS, **Alford DP**. Smoking policies in U.S. outpatient drug treatment facilities. Nicotine Tob Res. 2005;7:1-6.

39. **Rubin A**, **Migneault, JP**, **Marks L**, Goldstein E, **Ludena K**, **Friedman RH**. Automated telephone-screeening for problem drinking. J Stud Alcohol. 2006;67:454-457.

40. **Saitz R**. Unhealthy alcohol use. N Engl J Med. 2005;352:596-607.

41. **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.

42. **Samet JH**, Galanter M, **Bridden C**, Lewis DC. Association for medical education and research in substance abuse. Addiction. 2005;101:10-15.

43. **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. Antivir Ther. 2005;10:83-93.

44. **Shanahan CW**, Lincoln A, **Horton NJ**, **Saitz R**, Winter M, **Samet JH**. Relationship of depressive symptoms and mental health functioning to repeat detoxification. J Subst Abuse Treat. 2005;29:117-23.

45. Shwartz M, Peköz EA, **Ash AS**, **Posner MA**, Restuccia JD, Iezzoni LI. Do variations in disease prevalence limit the usefulness of population-based hospitalization rates for studying variations in hospital admissions? Med Care. 2005;43:4-11.

46. **Walley AY**, White MC, Kushel MB, Song YS, Tulsky JP. Knowledge of and interest in hepatitis C treatment at a methadone clinic. J of Sub Abuse Treatment. 2005;28:181-87.

47. Zhao Y, **Ash AS**, Ellis RP, Ayanian JZ, Pope GC, Bowen B, Weyuker L. Predicting pharmacy costs and other medical costs using diagnoses and drug claims. Med Care. 2005;43:34-43.

Factors Affecting Medical Students' Selection of an Internal Medicine Residency Program

Eva M. Aagaard, MD; Katherine Julian, MD; Julien Dedier, MD; Ira Soloman, MD; Jan Tillisch, MD; and Eliseo J. Pérez-Stable, MD

San Francisco and Los Angeles, California; Boston, Massachusetts; and New York, New York

These results presented in part at the Society of General Internal Medicine Annual Meeting, May 2001, San Diego, CA.

Objective: To determine factors that influence medical student selection of internal medicine residency programs by ethnicity and gender.

Design/Setting: A cross-sectional mailed survey of graduating medical students applying to four residency programs in 1999.

Measurements: A five-point (5=most important) Likert scale was used to evaluate factors and included 14 items on location characteristics, 20 on program features, six on recruitment, three on future plans and three on advising.

Results: Of 2,820 surveys, 1,005 were completed (36%). The most important factors to applicants were house staff morale (mean ± SD, 4.5 ± 0.7), academic reputation (4.5 ± 0.8), and positive interview experience (4.1 ± 1.0). Women rated gender diversity of faculty (3.3 vs. 2.3, p=0.0001) and house staff (3.3 vs. 2.5, p=0.0001), location of residency program near spouse (4.2 vs. 3.9, p=0.0001) or spouse's job (3.8 vs. 3.5, p=0.0002) and emphasis on primary care (2.9 vs. 2.4, p=0.0001) more highly than men. Minority applicants were more likely than whites to identify the following factors as more important: ethnic diversity of patients (3.8 vs. 3.4, p=0.008), house staff (3.3 vs. 2.4, p<0.0001) and faculty (3.1 vs. 2.3, p<0.0001); service to the medically indigent (3.8 vs. 3.3, p=0.004); feeling of being wanted (3.8 vs. 3.4, p=0.002); and an academic environment supportive of ethnic minorities (3.5 vs. 2.3, p<0.0001).

Conclusions: Location and program factors are most important in influencing decisions to choose a residency program. However, women and minority applicants also place significant importance on family and diversity factors. Programs need to consider differential factors in recruitment of diverse students.

Key words: ethnic minority ■ residency selection ■ medical education ■ medical students ■ women ■ internal medicine

 From Division of General Internal Medicine, Department of Medicine, Medical Effectiveness Research Center for Diverse Populations, The University of California at San Francisco, San Francisco, CA (Aagaard, Julian, Pérez-Stable); Section of General Internal Medicine Research Unit, Department of Medicine, Boston Medical Center, Boston, MA (Dedier); Division of General Internal Medicine, Department of Medicine, Columbia University, New York, NY (Soloman); Department of Medicine, University of California, Los Angeles, CA (Tillisch). Send correspondence and reprint requests for *J Natl Med Assoc.* 2005;97:1264–1270 to: Eliseo J. Pérez-Stable, MD, 400 Parnassus Ave., Box 0320, Room A-405, San Francisco, CA 94143-0320; phone: (415) 476-5369; fax: (415) 476-7964; e-mail: eliseops@medicine.ucsf.edu

INTRODUCTION

Selection and recruitment of qualified medical school graduates is a major concern of internal medicine residency directors and faculty at teaching institutions. The process of residency program selection is highly competitive, particularly with regard to top-ranked applicants.[1,2] Highly qualified underrepresented minority (URM) applicants represent an ever-smaller proportion of candidates for residency programs. This is true despite the fact that minority populations are the fastest-growing segment of the U.S. population.[3] According to the U.S. Census Bureau's projections, African Americans, Latinos, American Indians/Alaskan Natives and Asians, which made up nearly 30% of the U.S. population in 2000, will comprise 47.2% by the year 2050.[4] Despite the overall demographic changes in the United States, a decrease in URM applicants of 7.1% was noted from 1996 to 1997, such that minority applicants represented only 11% of the entire applicant pool to medical school.[5,6] Much of this decrease may be attributed to changes in political climate, including limits or elimination on affirmative action policy in key states.[7] On the other hand, the number of women enrolling in medical school has steadily increased over the last 25 years, with women enrollees making up 44.6% of all medical school enrollments in 2000.[8]

While it is clear that the most important aspect of recruitment involves increasing the number of qualified medical school applicants, it remains critical to internal medicine residency programs to be able to recruit excellent URM applicants and maintain an

increased proportion of women. Although studies have examined the effect of race and ethnicity on specialty choice,[9] to our knowledge, few studies have examined how or why medical students select a particular residency program,[1,10-15] and no studies have focused specifically on internal medicine residency programs. Furthermore, no studies have examined the residency program choices of URM applicants specifically or how the factors that determine applicant selection of a particular residency program differ between minority and majority candidates. The purpose of this study was to determine what factors most influence medical student selection of particular internal medicine residency programs. In addition, we sought to determine whether these factors differ by ethnicity and gender.

METHODS

Participants and Study Design

We performed a cross-sectional survey of fourth-year medical students applying during the 1999 National Residency Matching Plan (NRMP) in internal medicine. Potential participants were chosen from the lists of all medical students applying for placement in one or more of four internal medicine residency programs, including the University of California–San Francisco, University of California–Los Angeles, Massachusetts General Hospital,and New York Presbyterian Hospital. These four institutions were selected because of similar quality, and thus, would be expected to attract a similar pool of applicants.

Lists of applicants were provided by each institution and were cross-referenced for duplicate applications. One institution (UCLA) mailed their own surveys to assure complete anonymity of their applicant pool. NRMP numbers were provided by this institution to allow cross-references for duplicate applications. Only one survey was mailed to each student, and follow-up surveys were not sent to nonresponders in an attempt to limit response bias from students who did not match at their desired location. All applicants with mailing addresses outside of the United States and Canada were excluded from the study in an attempt to limit international medical graduates (IMG) from responding. We limited IMG for several reasons: 1) these applicants tend to be less competitive in the NRMP at these four institutions,[16] 2) because of extremely low acceptance and interview rates of these students, we felt that their inclusion could result in significant response bias, and 3) they may have significantly different motivating factors that influence their residency choice. Because of the sampling procedure described above, the exact number of international medical graduates excluded is unknown. The remaining 2,908 students were mailed a questionnaire. Only one mailing was possible in the one month between receiving the students' addresses and the results of the match. This timeframe was adhered to in order to decrease possible bias introduced by the students' match results. The institutional review board of University of California–San Francisco approved this study.

Questionnaire

The questionnaire obtained demographic information, including age, gender, ethnicity, relationships with significant others, medical school and class standing. Ethnicity was described as white/Caucasian, African-American, Latin American/Hispanic, Native American, Pacific Islander, Asian or other. URM were classified as African-American, Latin American/Hispanic, Native American and Pacific Islander. Class standing was determined by appli-

Table 1. Characteristics of fourth year medical students applying to the 1999 NRMP in internal medicine who responded to the survey

Characteristics	Men (%) (n=546)	Women (%) (n=451)
Age (years)	*27.9 ± 5.1	*28.1 ± 5.2
Ethnicity		
Caucasian/white	334 (61)	258 (57)
Latino/Hispanic	38 (7)	18 (4)
African American/black	10 (2)	13 (3)
American Indian/Native American	3 (<1)	1 (<1)
Pacific Islander	4 (1)	5 (1)
Asian/Asian American	135 (25)	131 (29)
Mixed	18 (3)	16 (4)
Unknown	4 (1)	9 (2)
Marital Status		
Married	170 (31)	127 (28)
Single with significant other	208 (38)	190 (42)
Single	168 (31)	133 (30)
Perceived Class Rank		
Top 25%	320 (59)	260 (58)
Middle 50%	189 (35)	152 (34)
Bottom 25%	18 (3)	17 (4)
Unknown	19 (3)	22 (5)

* Mean ± standard deviation; eight respondents were excluded because gender was not reported

cant's self-ranking in the top quarter, middle half or bottom quarter of the class, and whether or not they anticipated nomination into their local chapter of Alpha Omega Alpha (AOA), if applicable.

Information on the importance of 46 factors in choosing a particular residency program was obtained. Items chosen for evaluation were derived from prior studies,[2,11,12,14,15,17,18] discussion with program directors in internal medicine and focus group discussions with current residents. Focus group discussions were used to generate evaluation items but did not address the cultural validity of the survey. The surveys were then pretested for completeness and clarity. Factors assessed included 14 on location characteristics, 20 on program features, six on recruitment strategies, three on future plans and three reflecting advice received. Location characteristics included questions on gender and ethnic diversity of residents, faculty, and patients and geographic aspects of area. Questions also addressed programmatic factors, such as academic reputation, program size, program description, benefits/financial incentives and emphasis on primary care as well as interview techniques and process opportunities for future training or jobs in the area and advice received from a role model, friend or dean. Data were gathered using a five-point Likert scale, with 1 indicating that the factor was not important and 5 indicating that the factor was very important in the applicant's selection of a particular residency program. A factor was described as important if it received a mean response of 3 or greater.

Data Analysis

Data were analyzed using SAS version 8.2.[19] Descriptive statistics, including means, standard deviations and percents, were computed for each of the variables by gender and ethnicity (URM, Caucasian and Asian). Initial exploratory analyses for race and gender differences included Chi-squared tests for categorical variables and Student's t test for continuous variables. We further developed logistic regression models to explore the association of ethnicity (URM vs. non-URM) with the demographic measures and the factors in selecting a residency program. Our primary measures of outcomes were the factors in choosing a residency program. The major hypotheses were evaluated using analysis of covariance models (ANCOVA) to examine any gender or ethnic/racial group differences. For each outcome variable, we investigated the interactions among ethnicity, gender, age, marital status and class rank; we also compared Akaike's information criterion (AIC) and the residual log likelihood of the models with and without the interaction terms.[20] Since the main effects models tended to be more parsimonious with smaller AIC values, we decided to base our results on the main effects models. Estimates of adjusted means and standard errors (SE) were obtained by gender and by race, controlling for ethnicity, age, marital status and class rank. A significance level of 0.05 was used for all statistical tests.

RESULTS

Of the 2,908 students to whom a survey was sent, 88 surveys were returned for insufficient address, resulting in a total of 2,820 surveys presumably received by students. A total of 1,043 students returned completed questionnaires, for a response rate of 37%. Of these respondents, 38 were IMG and were excluded from all analyses. Another eight respondents did not indicate their gender and were not included in the analysis. The demographic characteristics of the respondents are presented in Table 1. No information was available on the demographic characteristics of nonrespondents due to the confidentiality of the NRMP list.

The factors described as important to all residency applicants in choosing a residency program are presented in Table 2. The factors felt to be most important were good house staff morale, the academic reputation of the program, a positive interview experience, the variety of clinical experiences and location near spouse or significant other.

Regression analysis demonstrated no significant interaction among gender, ethnicity, age and marital status. However, class ranking in the top third of the class was significantly negatively associated with URM status [OR 0.09 (95% CI 0.04–0.02)]. All variables were adjusted for in the ANCOVA results described below and in Tables 2 and 3. Results presented depict adjusted means, SE and corresponding p values.

The importance placed on factors determining residency choice differed by applicant gender (Table 2). The greatest differences in importance were seen regarding issues of gender diversity, where women rated the gender diversity of the faculty and house staff as important factors in their choice of residency, whereas men did not. Although having a same-gender interviewer was more important to women than men, such an experience was not particularly important to applicants of either gender. While both men and women felt that a positive interview experience was important, it was somewhat more important to women. Women placed more importance than men on familial issues, such as the location of the residency program being near their spouse/significant other (SO) or spouse's job, program support of applicants with children, and the presence of maternity and paternity leave policies. Men placed more importance on issues, such as the location of the res-

idency program being near their spouse's family and future fellowship opportunities.

Women placed a higher importance than men on issues of ethnic diversity in location and program characteristics when choosing a residency (Table 2). Women rated the following characteristics as more important than men: politics of the area being supportive of ethnic minorities; program academics supportive of ethnic minorities; serving the medically indigent; and the ethnic diversity of the city, patient population, house staff and faculty.

Several factors differed in importance among URM, Caucasian and Asian applicants (Table 3). Of the factors that differed among these groups, those that were most important to minority applicants were serving the medically underserved; amount of minority recruitment; feeling of being wanted; and ethnic diversity of the city, patients, house staff and faculty. An academic environment supportive of ethnic minorities and a political climate supportive of minorities were also noted as more important to URM applicants.

DISCUSSION

Our study reveals that the most important factors in choosing a particular internal medicine residency program are similar for all applicants and reflect the program location and characteristics of the program, such as good house staff morale, academic reputation and variety of clinical experiences provided. Although location characteristics are fixed, our study suggests there are several factors that a program may improve or emphasize to make their residency more attractive to applicants. The most important of these appears to be creating a positive environment that reflects good house staff morale and a positive, rather than competitive, interview experience. In addition, applicants are looking for a program with a diversity of clinical experiences. Program directors should emphasize the variety of hospitals, clinics and specialty rotations available to their residency applicants and expand such opportunities if not already available. Although these findings may seem intuitive, there are no previously published empiric data from applicants that support these observations. An earlier survey of family medicine program directors also found high-quality residents and faculty as well as having residents with "good attitudes" to be markers of success.[21]

Several other modifiable factors were of particular importance to women applicants. For example, women place significant emphasis on issues of gender diversity among the faculty and house staff. Thus, increasing the number and visibility of women faculty and residents may improve recruitment of highly qualified female applicants. Women also place a high value on family-friendly program characteristics and location characteristics that are favorable for their spouse/SO. Simple measures, such as clearly stating maternity and paternity leave policies and demonstrating support of residents with children, may also improve recruitment. A survey of matriculating and graduating medical students in 1993–1994 showed that compared to men, women rated specific curricular areas as having had inadequate instruction and that women were more likely to select a generalist specialty. However, data on factors influencing choice of program have not been reported.[10]

Both women and ethnic minorities place more importance on the ethnic diversity of the faculty and house staff. This suggests that increasing the number and visibility of faculty and house staff from diverse ethnic backgrounds is important for recruitment of these groups. Ethnic minorities also value a feeling of being wanted by the program, implying the importance of developing and implementing outreach efforts to such individuals.

Attracting minority physicians is of key importance to caring for all patients in the United States. Studies have demonstrated that minority physicians tend to serve members of their own racial or ethnic population group significantly more than they serve members of other groups, even after accounting for socioeconomic differences of area.[22,23] Minority physicians are also more likely to serve in a health workforce shortage area[24] and are more likely to care for patients with Medicaid and with no insurance.[22] In addition to providing a disproportionate amount of the care to ethnic minority groups and medically indigent patients, there is also evidence to suggest that language-concordant physicians provide better quality of care to monolingual Spanish-speaking Latino patients.[23] Furthermore, minority physicians can help to increase cultural awareness and reduce the language and cultural barriers that limit access to care for many minority patients.[3] Given these studies, recruitment of highly qualified minority applicants needs to be a priority among residency programs, particularly those serving ethnically diverse and medically underserved populations.

In our study, ethnic minorities placed significant importance on the ethnic diversity of patients and serving the medically indigent; this is consistent with data demonstrating the likelihood of minority physicians to care for medically underserved.[22,24] In addition, our data suggest that women may be more likely to have an interest in serving these populations as well. These results are supported by the other surveys of graduating medical students showing that women viewed caring for the medically indigent more positively than men.[10,25]

The most significant limitation to our study is the low response rate of 36% that introduces the possibility of bias. The challenge of distributing a survey

Table 2. Importance of factors in choosing an Internal Medicine Residency Program as rated by 1,005 applicants to the 1999 NRMP in Internal Medicine and the difference in the importance placed on those factors by gender*

Factor	Men Mean ± SE	Women Mean ± SE	P Value†
Location Characteristics			
Location near spouse/SO**	3.9 ± 0.11	4.3 ± 0.12	0.0001
Job opportunities for spouse/SO	3.4 ± 0.11	3.9 ± 0.13	0.0002
Cultural activities in area	3.4 ± 0.07	3.4 ± 0.08	0.4
Location near family	3.3 ± 0.09	3.5 ± 0.10	0.05
Ethnic diversity of city	3.3± 0.08	3.6 ± 0.09	0.001
Educational opportunities for spouse/SO	3.0 ± 0.13	3.2 ± 0.15	0.3
Location near friends	2.7 ± 0.09	2.7 ± 0.09	0.7
Common political values in area	2.1 ± 0.08	2.3 ± 0.09	0.02
Good environment for children	2.8 ± 0.11	2.8 ± 0.13	0.9
Recreational activities in area	3.2 ± 0.07	3.2± 0.08	0.4
Weather in area	2.9 ± 0.08	2.8 ± 0.09	0.04
Cost of living	2.6 ±0.08	2.6 ± 0.09	1.0
Location near spouse's/SO's family	2.6 ± 0.10	2.3 ± 0.12	0.01
Politics supportive of minorities	2.3 ± 0.08	2.7 ± 0.09	0.0001
Program Characteristics			
Academic reputation of program	4.4 ± 0.05	4.4 ± 0.06	0.7
Good house staff morale	4.5 ± 0.05	4.6 ± 0.05	0.1
Variety of clinical experiences offered	4.0 ± 0.06	4.1 ± 0.07	0.07
Ethnic diversity of patients	3.4 ± 0.08	3.7 ± 0.09	0.0002
Good on-call schedule	3.3 ± 0.07	3.5 ± 0.08	0.04
Number of hospitals rotated through	3.0 ± 0.07	3.2 ± 0.08	0.07
Research opportunities	3.0 ± 0.09	2.8 ± 0.09	0.6
Gender diversity of house staff	2.3 ± 0.08	3.2 ± 0.09	0.0001
Gender diversity of faculty	2.2 ± 0.08	3.2 ± 0.09	0.0001
Number of residents in the program	2.7 ± 0.08	2.9 ± 0.08	0.04
Emphasis on primary care	2.5 ± 0.09	2.0 ± 0.10	0.001
Ethnic diversity of house staff	1.7 ± 0.08	2.1 ± 0.09	0.0001
Academics supportive of minorities	2.7 ± 0.09	2.1 ± 0.10	0.0001
Amount of vacation	2.6 ± 0.08	2.7 ± 0.08	0.5
Good salary	2.8 ± 0.07	2.7 ± 0.08	0.06
Supportive of applicants with children	2.3 ± 0.12	2.7 ± 0.13	0.0004
Ethnic diversity of faculty	2.5 ± 0.08	2.9 ± 0.09	0.0001
Maternity/paternity leave policy	2.0 ± 0.08	2.5 ± 0.09	0.0001
Other financial incentives	2.0 ± 0.07	1.8 ± 0.08	0.007
Serving medically indigent	3.3 ± 0.08	3.7 ± 0.09	0.0002
Recruitment			
Positive interview experience	4.1 ± 0.07	4.3 ± 0.07	0.009
Feeling of being wanted/recruited	3.5 ± 0.08	3.7 ± 0.09	0.1
Prior experience at the program	2.8 ± 0.11	2.8 ± 0.13	0.4
Amount of minority recruitment	2.2 ± 0.07	2.4 ± 0.08	0.02
Same gender of interviewer	1.2 ± 0.05	1.5 ± 0.05	0.0001
Same ethnicity interviewer	1.3 ± 0.04	1.4 ± 0.04	0.07
Plans after Residency			
Fellowship opportunities in area	3.6 ± 0.09	3.4 ± 0.10	0.02
Desired location to live after residency	3.4 ± 0.09	3.3 ± 0.10	0.04
Job opportunities in area	3.2 ± 0.09	3.2 ± 0.10	0.8
Advising			
Advice of a role model	3.2 ± 0.08	3.2 ± 0.09	0.8
Advice of dean	2.7 ± 0.08	2.7 ± 0.09	1.0
Advice of friend	2.7 ± 0.08	2.7 ± 0.08	0.6

* Indicates adjusted mean response on a five-point Likert scale where 1=not important, 3=somewhat important, and 5=very important; † P value represents the result of the ANCOVA comparing responses by gender and controlling for ethnicity, marital status, age and class rank; **SO: significant other

to students from across the country at a time when many may be traveling and with only one-month window of time to complete follow-up was daunting. It is unclear how this low response rate may bias the results, if at all. Because we could not obtain significant demographic information on nonresponders, it is difficult to assess how this may have influenced our results. Importantly, the gender and ethnicity characteristics of the respondents were similar to those applying to internal medicine residencies nationally,[26] indicating less likelihood of significant response bias based on those variables. Our study did have a greater proportion of Asian respondents and fewer African-American respondents in comparison to the national statistics, but this is likely secondary to the geography of the institutions surveyed. Furthermore, although response rate was low, we were able to survey more than 1,000 students, a large sample size for such a study. We were also able to survey applicants applying to multiple institutions, and representing a broad geographic range. It is important to note that the institutions studied represented the east and west coasts, and not the midwest and south, and therefore may not be generalizable to applicants primarily interested in those geographic areas of the country.

Another limitation is that the class standing of our applicants represents the top of the applicant pool. Although this information is limited by self-report, it is likely correct, as the four institutions participating in the study tend to be among the most competitive internal medicine programs nationally. Our study was specifically designed to obtain information from these top applicants and because the study does not attempt to obtain information from the entire applicant pool, it may limit generalizability to other programs.

The survey nature of the study may limit the completeness of the information obtained. It is possible that there are other factors that applicants value highly that were not asked in the survey. We attempted to limit the chance of this with a thorough review of the literature, discussions with program directors and pretests of the survey with residents to assess for completeness. Finally, we recognize that many of the numerical differences seen between applicant groups, although statistically significant, are small and may not represent meaningful differences. Furthermore, because we tested for multiple associations, it is possible that some of the differences seen are simply a matter of chance. We attempted to limit this possibility by adjusting for multiple comparisons using the Scheffe method. Despite these limitations, to our knowledge, we are the first to report on the relative importance of various factors in choosing a residency program among minority applicants.

In summary, we conclude that internal medicine residency applicants value programmatic and location factors most highly in choosing a particular internal medicine residency program. Creating a positive experience for applicants and improving house staff morale

Table 3. Factors that differed in importance between URM* and Caucasian, and Asian and Caucasian applicants to the 1999 NRMP in Internal Medicine**

Factor	White (n=597)	URM (n=92)	P Value† (n=266)	Asian	P Value‡
Emphasis on primary care	2.6 ± 0.06	2.9 + 0.15	0.05	2.8 ± 0.09	0.04
Serving medically indigent	3.3 ± 0.06	3.7 + 0.14	0.004	3.3 ± 0.08	0.96
Good salary	2.6 ± 0.05	2.9 + 0.13	0.02	2.8 ± 0.08	0.02
Other financial incentives	1.7 ± 0.05	2.0 + 0.12	0.02	2.0 ± 0.07	0.0001
Amount of minority recruitment	1.6 ± 0.05	2.9 + 0.12	<0.0001	2.0 ± 0.07	<0.0001
Feeling of being wanted	3.4 ± 0.06	3.8 + 0.14	0.01	3.6 ± 0.08	0.09
Same ethnicity of interviewer	1.2 ± 0.03	1.5 + 0.07	<0.0001	1.4 ± 0.04	<0.0001
Ethnic diversity of faculty	2.2 ± 0.06	3.1 + 0.14	<0.0001	2.7 ± 0.08	<0.0001
Ethnic diversity of house staff	2.3 ± 0.06	3.3 + 0.14	<0.0001	2.8 ± 0.08	<0.0001
Ethnic diversity of patients	3.3 ± 0.06	3.7 + 0.14	0.01	3.3 ± 0.08	0.86
Ethnic diversity of city	3.3 ± 0.06	3.6 + 0.14	0.04	3.4 ± 0.08	0.07
Academic environment supportive of ethnic minorities	2.3 ± 0.06	3.4 + 0.14	<0.0001	2.9 ± 0.09	<0.0001
Political climate supportive of minorities	2.0 ± 0.06	3.0 + 0.14	<0.0001	2.2 ± 0.08	0.005

* URM: underrepresented minorities; ** Responses are reported as mean ± SE; † P value represents the result of the ANCOVA comparing minority to Caucasian applicants and controlling for gender, marital status, age and class rank; ‡ P value represents the result of the ANCOVA comparing Asian to Caucasian applicants and controlling for gender, marital status, age and class rank; 50 respondents did not answer the ethnicity question and are excluded from this analysis.

may enhance applicant interest in an internal medicine residency program, as these factors appear to be key in applicants' choice of residency program. Diversity of clinical experiences should also be emphasized and enhanced. For those programs with particular interest in recruiting women and ethnic minorities, emphasis might be placed on expanding gender and ethnic diversity among residents and faculty in the long-term. In the short-term, program directors can pay particular attention to voicing an interest in these issues and in the case of ethnic minorities, reaching out to them to make them feel more wanted. These interventions should be studied further to evaluate their effect on the ability of residency programs to recruit highly qualified female and minority applicants.

ACKNOWLEDGEMENT

We appreciate the collaboration of the staff at the National Residency Matching Program in supporting this study, Ginny Gildengorin, PhD for data management and analyses, Steve Gregorich, PhD for statistical consultation, and Sylvia Correro for administrative support. This work was supported by grant #P30 AG15272 awarded by the National Institute on Aging, the National Institute of Nursing Research, and the Office of Research on Minority Health to the University of California–San Francisco and HRSA Faculty Development Grant D08 PE50109.

REFERENCES

1. Sheldon GF. Recruitment and selection of the "best and brightest". Ann Thorac Surg. 1993;55:1340-1344.

2. Boex JR, Kirson SM, Keyes-Welch M, et al. Financial incentives in residency recruiting for primary care: scope, characteristics and students' perceptions. Acad Med. 1994;69:903-906.

3. Education CoGM. The 12th report of the council on graduate medical education. Trends in minority participation in medicine. 1998.

4. Census Bot. [U.S. Department of Commerce]. Current Population Reports, Population Projections of the United States by Age, Race and Hispanic Origin: 1995-2050.

5. Bigby JR. Declining enrollment of minorities in medical school. SGIM Forum. 1998. January:2,8.

6. U.S. Department of Health and Human Services, Health Resources and Services Administration. Council on Graduate Medical Education 12th Report: Minorities in Medicine. May 1998.

7. Carlisle D, Gardner J, Liu H. The entry of underrepresented minority students into U.S. medical schools: an evaluation of recent trends. Am J Public Health. 1998;88:1314-1318.

8. Bickel J, Clark V, Lawson R (AAMC). Women applicants, enrollees, and graduates—Selected Years, 1949–1950 through 2000–2001, AAMC; 2001.

9. Leverette J, Massabki A, Peterson H. Factors affecting medical students' selection of Canadian psychiatric residency programs: Part I—a comparison with American peers. Can J Psychiatry. 1996;41:578-581.

10. Leverette J, Massabki A, Peterson H. Factors affecting medical students' selection of Canadian psychiatric residency programs: Part II-Some contemporary issues. Can J Psychiatry. 1996;41:582-586.

11. Sacks MH, Karasu S, Cooper AM, et al. The medical student's perspective of psychiatry residency selection procedures. Am J Psych. 1983;140: 781-783.

12. Wayland M, Minnick S. Medical students' choices of teaching institutions for residency training. Acad Med. 1995;70:542.

13. Sledge WH, Leaf PJ, Sacks MH. Applicants' choice of residency training program. Am J Psych. 1987;144:501-503.

14. Bickel J, Ruffin A. Gender-associated differences in matriculating and graduating medical students. Acad Med. 1995;70:552-559.

15. Linzer M, Slavin T, Mutha S, et al. Admission, recruitment and retention: finding and keeping the generalist-oriented student. SGIM Task Force on Career Choice in Primary Care and Internal Medicine. J Gen Intern Med. 1994;9[4 Suppl 1]:S14-S23.

16. Rosenthal MP, Marquette PA, Diamond JJ. Trends along the debt-income axis: implications for medical students' selections of family practice careers. Acad Med. 1996;71:675-677.

17. Kikano GE, Galazka SS, Flocke SA, et al. Markers of successful recruitment of students to family practice residency programs. Fam Med. 1994;26:492-496.

18. Keith SN, Bell RM, Swanson AG, et al. Effects of affirmative action in medical school. N Eng J Med. 1985;313:1519-1524.

19. Komaromy M, Grumbach K, Drake M, et al. The role of black and Hispanic physicians in providing health care for underserved populations [see comments]. N Engl J Med. 1996;334:1305-1310.

20. Pérez-Stable EJ, Nápoles-Springer A, Miramontes JM. The effects of ethnicity and language on medical outcomes of patients with hypertension or diabetes. Med Care. 1997;35[12]:1212-1219.

21. Crandall S, Volk R, Loemker V. Medical students' attitudes toward providing care for the underserved. Are we training socially responsible physicians? JAMA. 1993;269:2519-2523.

22. Brotherton S, Simon F, Tomany S. U.S. Graduate Medical Education, 1999–2000. JAMA. 2000;284:1121-1126. ■

Work-Up of Women With Anemia

ARPITA AGGARWAL, MD, and PHYLLIS L. CARR, MD

ABSTRACT: Anemia results from a decrease in the normal amount of circulating hemoglobin and is commonly seen in women. Symptoms may include fatigue, tachycardia, palpitations, and dyspnea on exertion. Identifying the precise etiology is crucial to successful management. A pathophysiologic approach to diagnosis, which classifies anemias based on either decreased production or increased destruction of red blood cells, is most common. Laboratory testing involves measurement of hemoglobin and hematocrit in venous blood. An alternative approach, using mean corpuscular volume and red-cell distribution width, facilitates the classification of anemia. Tests include examination of a Wright's-stained peripheral blood smear and determination of red-cell indices.

Anemia is a common medical problem with multiple and varied causes that must be clearly understood to provide an adequate differential diagnosis and facilitate management. Appropriate treatment of anemia can significantly improve quality of life.

Anemia is defined as a disease resulting from a decrease in the normal amount of circulating hemoglobin, which leads to a reduction of the oxygen-carrying capacity of blood. The normal hemoglobin range can be different in individuals, depending upon age, sex, and race.[1] For example, in a neonate the hemoglobin level is between 18 and 22 g/dL. By the age of 3 months, it decreases to 1 to 17 g/dL, and in women of childbearing age, the normal range is from 12 to 14 g/dL. In general, normal hemoglobin levels are 1 to 2 g/dL lower in women and in African-American men than in white men. Also, some studies have found that the healthy elderly have lower hemoglobin than do younger adults.[2-4] Overall, the World Health Organization (WHO) recommends a hemoglobin level above 12.0 g/dL for women and 13.0 g/dL for men.[5]

EPIDEMIOLOGY

In the Tromso Study,[6] the highest prevalence of anemia in women was among the elderly; 16.5% were anemic according to the WHO standard, with a hemoglobin level below 12.0 g/dL. The lowest prevalence of anemia was among those ages 55 to 64, where the prevalence was 2.2%. The overall prevalence of anemia was 2.5%.

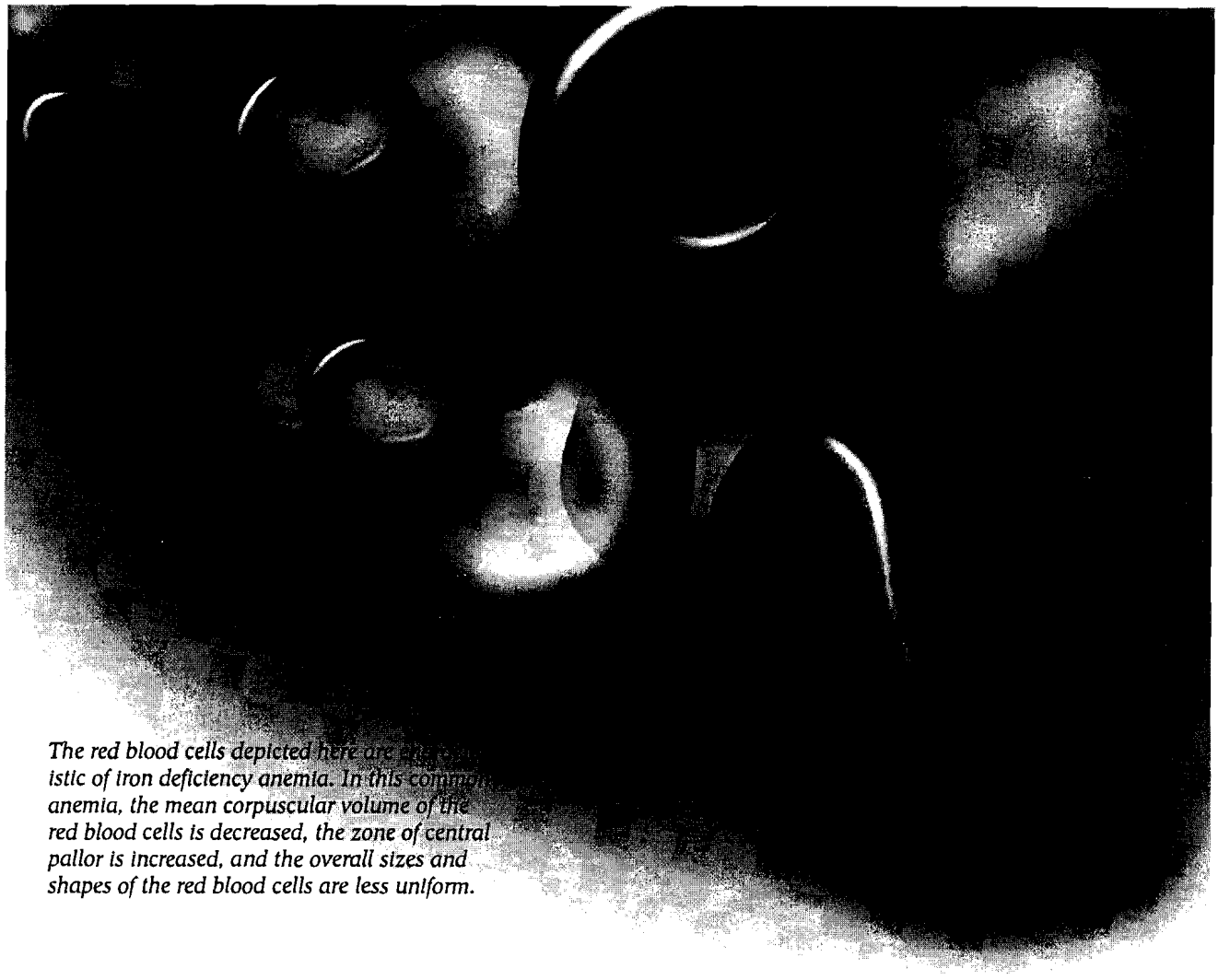
Iron deficiency is by far the most common cause of anemia among women of childbearing age because of loss of iron during menstruation. Anemia is less likely to occur in women taking birth control pills, due to decreased blood flow during menstruation, and more likely to occur in women with certain kinds of intrauterine devices, which can cause menorrhagia.[1,7,8] Perimenopausal women often have menorrhagia, which can result from fibroids and fluctuating hormonal levels. Therapy needs to be determined by the cause.

Pernicious anemia is most prevalent in persons of Northern European descent, occurring in 0.1%.[9,10] Thalassemia minor is a common cause of mild anemia in patients from the Mediterranean or the Far East.[11] Sickle cell disease affects one in 375 African-Americans; hence, screening for anemia in such populations is important.

CLINICAL PRESENTATION

Women with anemia most frequently present with fatigue, tachycardia, palpitations, and dyspnea on exertion. The history needs to include a detailed calendar of

Dr. Carr is Associate Professor of Medicine and Associate Dean of Students at Boston University School of Medicine. Dr. Aggarwal is a second-year fellow in general internal medicine/women's health at Boston University School of Medicine.



The red blood cells depicted here are characteristic of iron deficiency anemia. In this common anemia, the mean corpuscular volume of the red blood cells is decreased, the zone of central pallor is increased, and the overall sizes and shapes of the red blood cells are less uniform.

menstruation, with frequency, duration of flow, and the presence or absence of clots noted. Gastrointestinal symptoms associated with etiologics of blood loss, such as peptic ulcer disease (abdominal pain one to two hours after eating), gastritis (heartburn), prolonged vomiting, and other forms of blood loss in the gut, such as gastroesophageal reflux disease (GERD), as well as the presence of dark tarry stools or bright red blood in stools (evidencing lower-gastrointestinal bleeding) should also be noted as a routine part of the history.

INITIAL WORK-UP

There are numerous ways to approach the diagnosis and classification of anemia. The most common approach is on a pathophysiologic basis, ie, whether the anemia is related to diminished production of or loss of red blood cells (RBCs).[12,13] Table 1 shows an anemia classification based on pathophysiology. This can be determined by evaluating the reticulocyte count with the complete blood count. If the reticulocyte count is high, blood loss is the etiology, although early on, reticulocytes may not be present. If reticulocytes are absent or low, the anemia results from reduced production of RBCs.

Another diagnostic approach categorizes anemia based on mean corpuscular volume (MCV) into the subtypes microcytic, normocytic, or macrocytic, as outlined in Table 2.[12,13] Either approach can be used, or they can be combined.

Initial testing of the anemic patient should include a "complete" blood count (CBC). This routinely includes hemoglobin (Hb) and hematocrit levels, RBC count, RBC indices, and the white blood cell count. Anemia is defined in terms of red cell mass and hemoglobin concentration. In a new presentation of anemia, before further evaluation, the blood test should be repeated to avoid a false-positive diagnosis. This could occur when the plasma volume is expanded, such as in pregnancy, or alternatively, anemia could be masked if the patient is dehydrated with a decreased plasma volume.

Diagnosing Anemia in Women

Anemia is a common medical problem in women, occurring more frequently than in men, due to blood loss from menstruation and childbirth and as a result of certain problems that occur more often in women, such as collagen vascular disease. Finding the precise etiology is necessary to appropriately manage the anemia.

Women with anemia typically present with fatigue, tachycardia, palpitations, and dyspnea on exertion. History taking should include a detailed menstrual calendar that records frequency, duration of flow, and the presence or absence of clots. Gastrointestinal symptoms like heartburn or dark tarry stools should also be noted.

Confirming the diagnosis of anemia is an essential first step. The most common approach is to measure hemoglobin and hematocrit in venous blood with a complete blood count and classify the anemia according to the diminished production or increased destruction of red blood cells.

An alternative approach to diagnosing anemia is to measure mean corpuscular volume (MCV) and categorize the anemia into microcytic, normocytic, or macrocytic subtypes. Iron deficiency anemia, thalassemias, and anemia of chronic disease are associated with low MCV (less than 80 fL), while pernicious anemia is associated with a high MCV (above 100 fL).

Appropriate treatment can significantly improve quality of life, but the multiple and varied causes of anemia must be clearly understood to provide an adequate differential diagnosis and facilitate management.

Confirming the anemia is the essential first step.

MCV, mean corpuscular hemoglobin concentration, and red cell distribution width (RDW) help in the assessment of erythrocyte size, hemoglobin concentration, and shape. The combined use of MCV and RDW can facilitate the classification of anemia. If the MCV is normal but the RDW is high, there is variation of the red cell size, which can indicate an iron deficiency anemia, with large reticulocytes released prematurely from the bone marrow mixed with the small cells normally seen with decreased hemoglobin. Examination of a Wright's-stained peripheral blood smear will confirm this etiology.

CLASSIFICATION

Microcytic anemia

Low MCV (less than 80 fL) is associated with microcytic anemia. Although it is most commonly caused by iron deficiency, other common causes (eg, thalassemia and anemia of chronic disease) should also be considered. These syndromes can be differentiated by several simple blood tests (Table 3), such as an iron level and total iron-binding capacity (TIBC), or a ferritin level for iron deficiency anemia. Another blood test, a hemoglobin electrophoresis, can detect a variant hemoglobin, such as in sickle cell disease or thalassemia. A more difficult test, the bone marrow biopsy, can reveal absent iron stores in iron deficiency anemia or increased iron in the anemia of chronic disease. This test is rarely needed for these etiologies but is useful in anemias of bone marrow infiltration and the sideroblastic anemias.

Iron deficiency anemia

Iron stored in the body is critically balanced between dietary intake and the amount required by the body. The average American diet contains an intake of 10 to 15 mg of iron per day, but only 10% of that is absorbed. Women lose significant amounts of iron during menstruation; these quantities can be highly variable.[14] On average, women need 3 to 4 mg per day of absorbed

iron in their diet to keep up with the menstrual loss. Also, during pregnancy and lactation, dietary iron requirements increase. Another cause of iron deficiency in women is gastrointestinal bleeding (Table 4).

Clinical presentation: Mild anemia generally presents with easy fatigability, tachycardia, palpitations, and tachypnea on exertion. Severe anemia leads to skin and mucosal changes like glossitis, angular cheilosis, and koilonychias. Other symptoms like dysphasia (Plummer-Vinson syndrome) may also occur. Pica is often found among patients with iron deficiency anemia.

Diagnostic test results: Microcytic, hypochromic erythrocytes with occasional "pencil" and target forms are pathognomonic for iron deficiency anemia. Simple laboratory blood tests such as the CBC, peripheral blood smears, and iron indices show low iron and ferritin levels, high TIBC, and high RDW (Table 3).

Treatment: Oral iron supplements like ferrous gluconate or polysaccharide-iron preparations are well tolerated and have few gastrointestinal side effects. Ferrous sulfate is a less well-tolerated iron supplement that also can be used. It is recommended to take iron with a citrus juice (for vitamin C) on an empty stomach for better absorption and fewer side effects. Oral iron replacements can take up to eight or nine months to restore iron reserves, depending on the severity of the anemia. Total body iron can be rapidly replaced by an intravenous administration of iron, but this is rarely needed and is reserved only for patients refractory to oral iron replacement.[15]

Table 1. Traditional classification of anemia based on pathophysiology

Decreased production of red blood cells	
Etiology	**Type of anemia**
Hemoglobin synthesis	Iron deficiency, thalassemia, anemia of chronic disease
DNA synthesis	Megaloblastic anemia
Stem cell	Aplastic anemia, myeloproliferative anemia
Bone marrow infiltration	Malignant, infection, granulomatous disease, pure red cell aplasia
Increased destruction of red blood cells	
Etiology	**Type of anemia**
Blood loss	.....
Hemolysis (**intrinsic**)	
Membrane	Hereditary spherocytosis, elliptocytosis
Hemoglobin	Sickle cell anemia, unstable hemoglobin
Glycolysis	Pyruvate kinase deficiency
Oxidation	G6PD deficiency
Hemolysis (**extrinsic**)	
Immune	Warm antibody, cold antibody
Microangiopathic	Thrombotic thrombocytopenic purpura, hemolytic uremic syndrome, mechanical cardiac valve
Infection	Disseminated intravascular coagulation
Hypersplenism	.....

Adapted from Hoffman et al, eds. *Hematology: Basic Principles and Practice.* 1995[12]; Kumar et al, eds. *Robbins and Cotran's Pathologic Basis of Disease.* 2005.[13]

Anemia of chronic disease

Anemia of chronic disease is the second most prevalent type of anemia after iron deficiency anemia.[16] It is found primarily among patients with chronic inflammation, infection, and autoimmune disease.[17,18] Patients present with similar symptoms of anemia (fatigue, tachycardia, palpitations, and dyspnea on exertion). Severity of symptoms depends upon the level of anemia, which can range from mild (hemoglobin levels of 10 to 11 g/dL) to severe anemia (hemoglobin < 8 g/dL; 20% of cases).[19] In general, anemia of chronic disease is a normochromic, hypoproliferative anemia, but when severe, it can be hypochromic and microcytic. Measurement of total iron, TIBC, and ferritin help to distinguish the anemia of chronic disease from iron deficiency anemia and thalassemia (Table 3).[20] Table 5 outlines the possible causes of anemia of chronic disease.

Thalassemias

The thalassemias are hereditary disorders characterized by reduction in the synthesis of globin chains, either alpha (α) or beta (β), leading to decreased production of hemoglobin (Tables 6 and 7). Normal hemoglobin chain synthesis is α^2/β^2. Thalassemia trait is present when one gene is affected, as opposed to both genes; the latter is known as thalassemia major. This is most commonly found in persons of Asian or African ethnicity and in those of Mediterranean extraction.[21,22]

Clinical presentation for α-thalassemia is variable, depending upon the number of alleles affected (Table 6).[13,16,22] Patients with only one allele are asymptomatic with a normal blood smear. (If all four alleles are affected, fetal demise occurs.) Patients with thalassemia trait and thalassemia minor do not require any specific treatment. Thalassemia major requires appropriate monitoring of RBC indices and management depending on the level of hemoglobin. Low hemoglobin (< 8 g/dL) may require transfusion. Patients

Table 2. Differential diagnosis of anemia based on MCV

Low MCV (microcytic anemia: MCV < 80 fL)
Iron deficiency anemia (most common)
Thalassemic disorders
Anemia of chronic disease (especially in elderly women)
Sideroblastic anemia
Congenital
Lead
Alcohol
Drugs
Copper deficiency
Zinc poisoning (rare)
Normal MCV (normocytic anemia: MCV, 80 – 100 fL)
Acute blood loss (most common)
Early iron deficiency anemia
Anemia of chronic disease
Infection, HIV
Inflammation
Malignancy
Bone marrow suppression
Bone marrow invasion
Acquired pure red cell aplasia
Aplastic anemia/myelofibrosis
Autoimmune hemolytic anemia
Erythroblastosis fetalis
Transfusion reaction
Collagen vascular disease
Hemolytic uremic syndrome/ thrombocytopenic purpura
Chronic renal disease (decreased erythropoietin)
Endocrine disorder
Hypothyroidism
Hypopituitarism
Spherocytosis (hereditary anomaly of the red cell membrane)
Paroxysmal nocturnal hemoglobinuria
High MCV (macrocytic anemia: MCV > 100 fL)
Ethanol abuse
Folic acid/vitamin B_{12} deficiency
Myelodysplastic syndromes
Acute myeloid leukemia
Reticulocytosis
Hemolytic anemia
Response to blood loss
Response to appropriate therapy
Drug-induced anemia
Liver disease/severe hypothyroidism

MCV, mean corpuscular volume.
Adapted from Hoffman et al, eds. *Hematology: Basic Principles and Practice*. 1995.[12]; Kumar et al, eds. *Robbins and Cotran's Pathologic Basis of Disease*. 2005.[13]

with hemoglobin H (thalassemia major) have a tendency for oxidative injury in the presence of infections and need to be monitored for transfusion.[23]

The majority of patients with β-thalassemia minor (Table 7) are asymptomatic but have a mild hypochromic microcytic anemia similar to iron deficiency anemia. They have microcytosis (MCV above 75 fL) even when the hemoglobin is above 10 g/dL, and they have a normal RDW. Twenty percent of these patients develop splenomegaly. Women with β-thalassemia develop more profound anemia of pregnancy, but outcomes are generally favorable.[24]

Patients with thalassemia major first manifest symptoms at age 6 months with pallor, irritability, growth retardation, hepatosplenomegaly, and jaundice (hemolysis).[24] Later these patients develop symptoms associated with ineffective erythropoiesis, eg, bony abnormalities and abnormal skeletal development. Eighty percent of untreated children die within the first five years of life, due to severe anemia, high-output congestive heart failure, and infections.[25]

β-thalassemia minor requires constant monitoring and is treated with occasional transfusions as needed. Treatment for β-thalassemia major is more complicated and requires repeated transfusions. Management of transfusion complications is considered to be a regular part of therapy. Patients with β-thalassemia major need splenectomy and iron chelation with deferoxamine.[26] There are other therapies under investigation, eg, the use of allogeneic bone marrow transplant and gene therapy.[12]

Macrocytic anemia

Pernicious anemia, dietary deficiency of vitamin B_{12} (rare), and folic acid deficiency (more common in the elderly) can lead to a macrocytic anemia (MCV above 100 fL). Alcohol abuse and medications can also cause macrocytic anemia.

Pernicious anemia

Although more common in older adults of Northern European descent, pernicious anemia has also been found among younger black females. It is a hereditary disorder which presents only after age 35. Pernicious anemia is caused by defective absorption of vitamin B_{12} (cobalamin). Cobalamin binds with intrinsic factor in the stomach to prevent enzymatic degradation and is absorbed in the terminal ileum. Antibodies against intrinsic factor decrease the cobalamin absorption leading to anemia. Gastric surgery, removal of the terminal ileum, or bacterial overgrowth in the ileum can also cause a decrease in absorption of cobalamin. Absence of cobalamin leads to the formation of abnormal blood cells causing ineffective erythropoiesis, or formation of blood cells.

Clinical presentation: A megaloblastic state causes changes in the

Table 3. Lab results typical of microcytic anemia

Lab result	Iron deficiency anemia	Anemia of chronic disease	Thalassemia trait
Iron	Low	Low	Normal/high
Ferritin	Low	High	High
TIBC	High	Low	Normal
RDW	High	Normal	Normal
Marrow iron	Absent	Increased	Increased

TIBC, total iron-binding capacity.
RDW, red cell distribution width.

mucosal cells, such as glossitis and diarrhea. It also causes severe abnormalities in the peripheral nervous system leading to paresthesias, difficulty in balance, neuropsychiatric changes, and even dementia.

Diagnostic test results: Generally the MCV is above 110 fL, but sometimes there is a normal MCV with vitamin B_{12} deficiency in the presence of a combined microcytic anemia. The peripheral blood smear shows anisocytosis, poikilocytosis, hypersegmented neutrophils (greater than 4 lobes), and a low reticulocyte count. Since vitamin B_{12} deficiency causes abnormal hematopoiesis, it also leads to low production of white blood cells and platelets.

The vitamin B_{12} assay as part of the Schilling test helps to establish the diagnosis of pernicious anemia. Some recommend testing for increased serum methylmalonic acid and homocysteine levels, which are more sensitive in the early stages of B_{12} deficiency, in making the diagnosis.[27]

Treatment: Replacement of vitamin B_{12} is initiated with intramuscular injections of 1,000 mg weekly for four weeks and then monthly for life. Oral supplements of B_{12} can be used after the first month of injections, if absorption is not a problem.[28]

CONCLUSION

The evaluation and diagnosis of anemia, a common medical problem, requires finding the precise etiology in order to appropriately manage the underlying cause. Anemia occurs more frequently in women, due to blood loss during menstruation and childbirth and also as a result of certain problems that occur more frequently in women, such as collagen vascular disease. A full understanding of the pathophysiology of anemia facilitates a rational approach to its management.

Table 4. Causes of iron deficiency

Dietary deficiency (uncommon in the United States)
Decreased iron absorption (pernicious anemia, gastric surgery, or removal of the terminal ileum)
Pregnancy
Lactation
Blood loss (gastrointestinal loss, menstruation)
Iron sequestration (pulmonary hemosiderosis)

Adapted from Hoffman et al, eds. *Hematology: Basic Principles and Practice.* 1995[12]; Kumar et al, eds. *Robbins and Cotran's Pathologic Basis of Disease.* 2005[13]; Hillman and Ault. *Hematology in Clinical Practice.* 2002.[20]

Table 5. Anemia of chronic disease

Caused by sequestration of iron in reticuloendothelial system
Chronic infection (granulomatous disease)
Inflammation (collagen vascular disease and other autoimmune diseases)
Malignancy
Liver disease
Chronic renal disease (decreased erythropoietin)

Adapted from Hoffman et al, eds. *Hematology: Basic Principles and Practice.* 1995[12]; Kumar et al, eds. *Robbins and Cotran's Pathologic Basis of Disease.* 2005.[13]

Table 6. α-Thalassemia

No. of alleles affected	Characteristics	Blood smear
1. Carrier state	No anemia, asymptomatic	No abnormalities
2. Thalassemia (trait)	Usually clinically silent	Mild microcytic anemia
3. Hemoglobin H disease	Precipitation of β-chain tetramers	Intraerythrocytic inclusions
4. Hydrops fetalis	Fetal demise, total body edema	Bart's β4 Hb precipitations

Hb, hemoglobin.
Adapted from Kumar et al, eds. *Robbins and Cotran's Pathologic Basis of Disease.* 2005[13]; Hillman and Ault. *Hematology in Clinical Practice.* 2002[20]; Bunn et al. *Major Probl Intern Med.* 1977.[22]

Table 7. Thalassemia major and minor

Thalassemia major (β-/β-)
Anemia develops at age 6 months when adult Hb replaces fetal Hb, and clinically presents with splenomegaly and frontal bossing due to extramedullary hematopoiesis
Electrophoresis shows severely low Hb A, high Hb A2, and Hb F
Thalassemia minor (β+/β-)
Typically asymptomatic carriers who have the trait
Electrophoresis shows low Hb A, high Hb A2, and normal to high Hb F

Hb, hemoglobin.
Adapted from Hoffman et al, eds. *Hematology: Basic Principles and Practice.* 1995[12]; Kumar et al, eds. *Robbins and Cotran's Pathologic Basis of Disease.* 2005.[13]

REFERENCES

1. Babior BM, Stossel TP. *Hematology, a Pathophysiological Approach*. New York, NY: Churchill Livingstone; 1984:167.
2. Baldwin JG Jr. Hematopoietic function in the elderly. *Arch Intern Med*. 1988; 148:2544-2546.
3. Balducci L. Epidemiology of anemia in the elderly: information on diagnostic evaluation. *J Am Geriatr Soc*. 2003; 51(suppl 3):S2-S9.
4. Guralnik JM, Eisenstaedt RS, Ferrucci L, et al. Prevalence of anemia in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. *Blood*. 2004;104:2263-2268.
5. World Health Organization. *Nutritional anaemias: report of a WHO scientific group*. Geneva, Switzerland: World Health Organization; 1968.
6. Skjelbakken T, Langbakk B, Dahl IM, Lochen ML. Haemoglobin and anaemia in a gender perspective: the Tromso Study. *Eur J Haematol*. 2005;74:381-388.
7. Bermejo B, Olona M, Serra M, et al. Prevalence of iron deficiency in the female working population in the reproductive age [in Spanish]. *Rev Clin Esp*. 1996;196:446-450.
8. Kitai E, Kaplan B, Raick Y, et al. Community screening to reveal iron deficiency in healthy menstruating women in Israelian suburbs. *Eur J Obstet Gynecol Reprod Biol*. 1996;67:21-25.
9. Carmel R. Prevalence of undiagnosed pernicious anemia in the elderly. *Arch Intern Med*. 1996;156:1097-1100.
10. Chanarin I. *The Megaloblastic Anaemias*. 2nd ed. Boston, Mass: Blackwell Scientific Publications, Inc; 1979.
11. Park KW. Sickle cell disease and other hemoglobinopathies. *Int Anesthesiol Clin*. 2004;42:77-93.
12. Hoffman R, et al, eds. *Hematology: Basic Principles and Practice*. 2nd ed. New York: Churchill Livingstone; 1995:586-610.
13. Kumar V, Abbas AK, Fausto N, eds. *Robbins and Cotran's Pathologic Basis of Disease*. 7th ed. Philadelphia, Pa: Elsevier Saunders; 2005:622-649.
14. Ballin A, Berar M, Rubinstein U, et al. Iron state in female adolescents. *Am J Dis Child*. 1992;146:803-805.
15. Semba RD, Guralnik, JM Chaves P, et al. Iron status and anemia in a population-based study of women with and without disability living in the community: the Women's Health and Aging Studies. *Haematologica*. 2004;89:357-358.
16. Matzner Y, Levy S, Grossowicz N, et al. Prevalence and causes of anemia in elderly hospitalized patients. *Gerontology*. 1979;25:113-119.
17. Weiss G, Goodnough LT. Anemia of chronic disease. *N Engl J Med*. 2005; 352:1011-1023.
18. Weiss G. Pathogenesis and treatment of anaemia of chronic disease. *Blood Rev*. 2002;16:87-96.
19. Cash JM, Sears DA. The anemia of chronic disease: spectrum of associated diseases in a series of unselected hospitalized patients. *Am J Med*. 1989; 87:638-644.
20. Hillman R, Ault KA. Clinical approach to anemia. In: *Hematology in Clinical Practice*. 3rd ed. New York, NY: McGraw-Hill; 2001.
21. Benz EJ Jr, Forget BG. Pathogenesis of the thalassemia syndromes. *Pathobiol Annu*. 1980;10:1-33.
22. Bunn HF, Forget BG, Ranney HM. Hemoglobinopathies. *Major Probl Intern Med*. 1977;12:1-291.
23. Cheng ML, Ho HY, Tseng HC, et al. Antioxidant deficit and enhanced susceptibility to oxidative damage in individuals with different forms of alpha-thalassaemia. *Br J Haematol*. 2005; 128:119-127.
24. Sheiner E, Levy A, Yerushalmi R, Katz M. Beta-thalassemia minor during pregnancy. *Obstet Gynecol*. 2004;103:1273-1277.
25. Weatherall DJ, Clegg JB. *The Thalassemia Syndromes*. 3rd ed. Oxford, England: Blackwell Scientific Publications; 1981.
26. Olivieri NF, Brittenham GM. Iron-chelating therapy and the treatment of thalassemia. *Blood*. 1997;89:739-761.
27. Sumner AE, Chin MM, Abrahm JL, et al. Elevated methylmalonic acid and total homocysteine levels show high prevalence of vitamin B_{12} deficiency after gastric surgery. *Ann Intern Med*. 1996;124:469-476.
28. Hathcock JN, Troendle GJ. Oral cobalamin for treatment of pernicious anemia? *JAMA*. 1991;265:96-97.

Zhao Y, Ash AS, Ellis RP, Ayanian JZ, Pope GC, Bowen B, Weyuker L. Predicting Pharmacy Costs and Other Medical Costs Using Diagnoses and Drug Claims. Medical Care. 2005 Jan;43: 34-43.

Shwartz M, Pekoz EA, Ash AS, Posner MS, Restuccia JD, Iezzoni LI. Do Variations in Disease Prevalence Limit the Usefulness of Population-based Hospitalization Rates for Studying Variations in Hospital Admissions? Medical Care. 2005 Jan;43:4-11.

Chuang C, Freund KM. Emergency Contraception Knowledge among Women in a Boston Community. Contraception 2005 Feb;71:157-160.

David MM, Ko L, Prudent N, Green EH, Posner MA, Freund KM. Mammography Use, Journal of National Medical Association 2005 Feb;67:253-260.

Barnett RC, Gareis KC, Carr PL. Career Satisfaction and Retention of a Sample of Women Physicians who Work Reduced Hours. Journal of Women's Health. 2005 Mar;14(2):146-53.

Green EH, Freund KM, Posner MA, David MM. Pap Smear Rates Among Immigrant Haitian Women in Eastern Massachusetts. Public Health Report. 2005 Mar-Apr;120(2):133-139.

Boehmer U, Linde R., Freund KM. Sexual Minority Women's Coping and Psychological Adjustment after a Diagnosis of Breast Cancer. Journal of Women's Health. 2005 Apr ;14(3):214-24

Aggarwal A, Carr P. Work-Up of Women With Anemia. Women's Health in Primary Care. 2005 Aug; 8(7):307-314.

Chuang C, Freund KM. Emergency Contraception: an Intervention on Primary Care Providers. Contraception 2005 Sep;72:182-186.

Carr, PL, Rothberg MC, Friedman RH, Felsentein D, Pliskin JS. "Shotgun" Versus Sequential Testing. Cost-effectiveness of Diagnostic Strategies for Vaginistis. Journal of General Internal Medicine. 2005 Sep; 72(3):182-6.

Boehmer U, Freund KM, Linde R. Support Providers of Sexual Minority Women With Breast Cancer Who They Are and How They Impact The Breast Cancer Experience. Journal of Psychosomatic Research. 2005 Nov;59:307-314.

Berlowitz DR, Ash AS, Glickman M, Friedman RH, Pogach LM, Nelson AL, Wong AT. Developing a Quality Measure for Clinical Inertia in Diabetes Care. Health Services Review. 2005 Dec;40:1836-53.

JOURNAL OF WOMEN'S HEALTH
Volume 14, Number 2, 2005

Career Satisfaction and Retention of a Sample of Women Physicians Who Work Reduced Hours

ROSALIND C. BARNETT, Ph.D.,[1] KAREN C. GAREIS, Ph.D.,[1] and PHYLLIS L. CARR, M.D.[2]

ABSTRACT

Objective: **To better understand the career satisfaction and factors related to retention of women physicians who work reduced hours and are in dual-earner couples in comparison to their full-time counterparts.**

Methods: **Survey of a random sample of female physicians between 25 and 50 years of age working within 25 miles of Boston, whose names were obtained from the Board of Registration in Medicine in Massachusetts. Interviewers conducted a 60-minute face-to-face closed-ended interview after interviewees completed a 20-minute mailed questionnaire.**

Results: **Fifty-one full-time physicians and 47 reduced hours physicians completed the study; the completion rate was 49.5%. The two groups were similar in age, years as a physician, mean household income, number of children, and presence of an infant in the home. Reduced hours physicians in this sample had a different relationship to experiences in the family than full-time physicians. (1) When reduced hours physicians had low marital role quality, there was an associated lower career satisfaction; full-time physicians report high career satisfaction regardless of their marital role quality. (2) When reduced hours physicians had low marital role or parental role quality, there was an associated higher intention to leave their jobs than for full-time physicians; when marital role or parental role quality was high, there was an associated lower intention to leave their jobs than for full-time physicians. (3) When reduced hours physicians perceived that work interfering with family was high, there was an associated greater intention to leave their jobs that was not apparent for full-time physicians.**

Conclusions: **Women physicians in this sample who worked reduced hours had stronger relationships between family experiences (marital and parental role quality and work interference with family) and professional outcomes than had their full-time counterparts. Both career satisfaction and intention to leave their employment are correlated with the quality of home life for reduced hours physicians.**

INTRODUCTION

With the advent of more dual-earner couples, increasing numbers of women physicians are seeking part-time or reduced hours employment, largely to allow greater time for family.[1] The percentage of part-time pediatricians increased from 11% to 15% between 1993 and 2000.[1] Previous work has shown that there is little difference in a number of outcomes for reduced hours physicians compared with their full-time counterparts, including job role quality and career satis-

[1]Women's Studies Research Center, Brandeis University, Waltham, Massachusetts.
[2]Department of Medicine, Boston University School of Medicine, Boston, Massachusetts.
This work was funded by the National Institute for Occupational Safety and Health (grant OH 03848).

faction.[2] Working the preferred number of hours, for both full-time and reduced hours physicians, had the greatest impact on job role quality, burnout, marital role quality, and life satisfaction.[2] The Women Physicians' Health Study similarly found a strong association between work control and career satisfaction,[3] with lack of control being a strong predictor of burnout in women, but not in men.[3,4]

Understanding important factors in the career satisfaction and retention of part-time physicians has become increasingly valuable, as part-time physicians in clinical practices have been shown to have higher productivity than full-time physicians[5–7] and equal to or higher quality performance, with similar patient satisfaction and ambulatory costs.[8] The only area in which reduced hours physicians did not always have comparable outcomes to full-time physicians was in visit-based continuity of care.[9] A study from Brigham and Women's Hospital found that part-time physicians scored higher on the Health Plan Employer Data and Information Set (HEDIS) quality measure, including access and availability of care, effectiveness of care, satisfaction with the experience of care, cost of care, and informed healthcare choices.[10] Reduced hours employees in other professions are also more productive than their full-time counterparts.[11,12] Given these data, it becomes increasingly prudent and cost-effective to maximize the retention of part-time physicians.

Most women physicians who work reduced hours likely do so to accommodate family matters.[13] Little is known, however, about the effects of family on career satisfaction for full-time and part-time physicians or of the effect of family on the reported likelihood of retention of women physicians in full-time and part-time work roles. We conducted a study of full-time and reduced hours female physicians in dual-earner couples to better understand how career satisfaction and likelihood of leaving the job were related to family indicators, such as marital role quality, parental role quality, and a work interfering with family scale.

MATERIALS AND METHODS

Sample

The sample was randomly drawn from the Board of Registration in Medicine, which licenses all physicians practicing in the Commonwealth of Massachusetts.[2] First, we identified a subsample of female physicians between 25 and 50 years of age who worked within a 25-mile radius of Boston and sent them introductory letters describing the study. Next, trained screeners contacted each physician to determine if she met the eligibility criteria for inclusion in the study. A woman physician was eligible if she worked full-time or reduced hours as defined by her employer. All reduced hours physicians worked at least 20 hours per week, and all participants were in dual-earner couples, with husbands working at least 20 hours per week, and all couples had at least one child who was under high school age. Our goal was to interview 50 full-time and 50 reduced hours physicians to have adequate statistical power to test our hypothesis. However, reduced hours physicians were harder to locate than were their full-time counterparts. Therefore, we expanded our sampling strategy by asking participants to nominate eligible physicians and by asking Partners, an umbrella organization representing physicians from a number of major Boston hospitals, to send out a letter to member physicians asking eligible staff to contact us if they were interested in participating. These efforts resulted in 6 additional participants.

Data were collected between September 1999 and March 2001. The final sample consisted of 51 full-time doctors and 47 reduced hours doctors ($n = 98$). For purposes of this study, a doctor was categorized as full-time or reduced hours if her employer considered her to be so. We also queried participants, both full-time and reduced hours, to see if they were working their preferred number of hours. Among the 92 doctors obtained via random sampling, the completion or cooperation rate was 49.5% (calculated as participants divided by participants plus refusers). An additional 6 respondents were volunteers or were nominated by other doctors as eligible.

Procedures

Screeners passed on the names of eligible and willing participants to trained interviewers. These interviewers sent a recruitment package to each potential participant indicating that she would be called shortly to set up an interview. The letter described the time commitment and remuneration that participation in the study would entail. The package also included endorsements from the Massachusetts Medical Society and the

American Medical Women's Association, along with two papers describing our previous project with reduced hours physicians. Trained interviewers conducted 60-minute face-to-face closed-ended interviews with each participant at a time and place convenient to the participant. Each participant also completed a 20-minute mailed questionnaire in advance and returned it at the time of the interview. The interview and mailed survey covered various objective and subjective aspects of participants' jobs (e.g., salary, number of hours worked, career satisfaction, schedule fit), the quality of their major social roles (partner, parent, employee), and various quality of life indicators (e.g., psychological distress, life satisfaction, physical symptoms). Each physician received $25 for her participation.

Measures

The outcomes for this analysis were (1) intention to leave the job within a year and (2) career satisfaction. Intention to leave the job was a single item question asking participants to use a 7-point scale to indicate the likelihood that they would voluntarily terminate their present employment within the next 12 months. We measured career satisfaction using a scale developed for a study of academic physicians (for a description of the sample and methods, see ref. 14). Participants used a 7-point scale to rate six aspects of career satisfaction: (1) their current work setting, (2) their potential to achieve their professional goals, (3) their overall professional practice, (4) the extent to which this practice met their expectations, (5) their overall professional research, and (6) the extent to which this research met their expectations. Cronbach's alpha was 0.86 in the present sample.

The predictors of career satisfaction and likelihood of leaving the job in 1 year were parental role quality, marital role quality, and work-family interference. We used a measure of parental role quality by Barnett et al.[15] Respondents indicated on a 4-point scale the degree to which each of a list of 44 parenting items was currently either rewarding or of concern. Concern items were negatively weighted, and reward items were positively weighted in constructing the role quality score, which was the weighted average of the individual item scores. In the present sample, Cronbach's alpha was 0.92 for the rewards and 0.83 for the concerns. We measured marital role quality using a 15-item brief form[16] of the marital role quality scale.[17] Respondents indicated on a 4-point scale the degree to which each of a list of relationship items was currently either rewarding or of concern. Internal consistency was excellent, with Cronbach's alphas of 0.91 for rewards and 0.89 for concerns in the present sample. We assessed work-family interference using items from a scale developed by MacDermid et al.[18] We selected one item each addressing the energy, strain, and behavioral components of work-family interference along with a fourth, more global item assessing the overall severity of work-family interference. Cronbach's alpha was 0.73 in the present sample.

The moderator, which is a variable that can influence the strength of the relationship between two other variables, full-time vs. reduced-hours status as defined by the employer, was scored as a dichotomous variable (1 = full-time, 2 = reduced hours). Covariates for all regression analyses included the presence of at least one preschool child at home, years as a physician, household income per capita, and negative affectivity. For moderator analyses involving marital role quality, years married was included as an additional covariate; for moderator analyses involving work-family interference, husband's work hours was included as an additional covariate.

Years as a physician and years married are self-explanatory. We assessed husband's work hours by asking respondents to estimate the number of hours their husbands worked in an average work week. Presence of a preschool child was coded as a dummy variable (1 = at least one child under 5 years of age in the home, 0 = else). We calculated household income per capita by dividing each respondent's report of yearly household income by the number of people living in the household. Because the distribution of this variable is highly skewed, we used the natural log of per capita income.

We assessed negative affectivity using the Trait Anxiety Scale,[19] on which respondents indicated on a 4-point scale the degree to which they were characterized by 10 specific traits. Internal consistency was high, with a Cronbach's alpha of 0.90 in the present sample. We controlled for negative affectivity, an individual trait predisposing to a negative view of the world, because it is thought to account for spuriously high correlations between self-report measures of predictor and outcome variables, especially in cross-sectional analyses.[20,21]

Analysis

Data were analyzed using SPSS software (Chicago, IL). We generated frequency distributions and descriptive statistics (means and standard deviations [SD]) for comparisons of responses by work status. Independent sample *t* tests were computed to test the significance of differences between full-time and reduced hours physicians.

We hypothesized that the quality of family life—which we assessed with measures of parental role quality, marital role quality, and perceptions of work-family interference—would be a stronger predictor of career satisfaction and intention to leave their jobs among reduced hours physicians than among their full-time counterparts.

To test this hypothesis, we first estimated a series of six main effects simultaneous regression models predicting two outcomes: career satisfaction and intention to leave one's job. In all models, the predictors included one of the variables assessing the quality of family life (i.e., parental role quality, marital role quality, or work-family interference) along with the hypothesized moderator, full-time vs. reduced hours status. As discussed, all models included the following covariates: presence of a preschool child, years as a physician, household income, and negative affectivity. Additional covariates included length of marriage in the analyses involving marital role quality and husband's work hours in the analyses involving work-family interference.

To test whether there was a significant moderating effect of working full-time or reduced hours, we added an interaction term of the form PREDICTOR × MODERATOR to each regression model. In this way, we could determine whether the addition of the interaction term explained a significant proportion of the variance over and above that explained by the simpler model. A significant increase in proportion of variance explained would support the moderation hypothesis; that is, that the relationship linking quality of family life to a particular outcome was different for full-time physicians than for reduced hours physicians.

RESULTS

Our study included 51 full-time physicians. Reduced hours physicians (47) worked a mean of 32.1 hours per week (range 20–60 hours per week), and full-time physicians worked 48.7 hours per week (range 35–90 hours per week), a difference of 16.5 hours ($p = 0.000$). There was no difference in the mean age of the physicians in the two groups (40.5 years vs. 39.9 years, $p = 0.563$) or in the number of years as a physician (10.8 years vs. 9.6 years, $p = 0.354$). Full-time physicians, however, had been working full-time for a longer period than reduced hours physicians had been working reduced hours (85.6 months vs. 48.2 months, $t(93.4) = 3.33$, $p = 0.001$). There was no difference in the mean household income ($244,421 vs. $229,889; $p = 0.625$), the proportion of physicians practicing in an academic setting (3 full-time physicians vs. 2 reduced hours physicians), the number of children (2.3 vs. 2.1, $p = 0.468$), or the presence of an infant in the home (19.6% vs. 17.0%; $p = 0.744$). In addition, there was no difference between the two groups in the likelihood of having a preschool (60.8% vs. 74.5%, $p = 0.150$), a school age (59.6% vs. 40.4%, $p = 0.106$), or a teenage child (27.5% vs. 14.9%, $p = 0.129$) at home.

In our sample, there was a difference between full-time and reduced hours physicians in the relationship between marital role quality and career satisfaction ($p = 0.017$). Specifically, as shown in Figure 1, the career satisfaction of reduced hours women physicians was more strongly associated with their marital role quality than was the career satisfaction of full-time women physicians. Low marital role quality was associated with low career satisfaction among the reduced hours physicians, whereas marital role quality appeared to be unrelated to career satisfaction among the full-time physicians.

There was also a significant interaction effect of full-time vs. reduced hours status on the relationship between marital role quality and intention to leave one's job ($p = 0.004$) in our sample. As shown in Figure 2, when marital role quality was low, reduced hours physicians expressed a greater intention to leave their job within a year than did full-time physicians. When marital role quality was high, however, reduced hours physicians expressed less intention to leave their jobs than did their full-time counterparts.

Full-time vs. reduced hours status in our sample was also a significant moderator of the relationship between parental role quality and intention to leave one's job ($p = 0.006$). As shown in Figure 3, when there was high parental role quality, there was no difference in intention to leave the job between the two groups. However, when

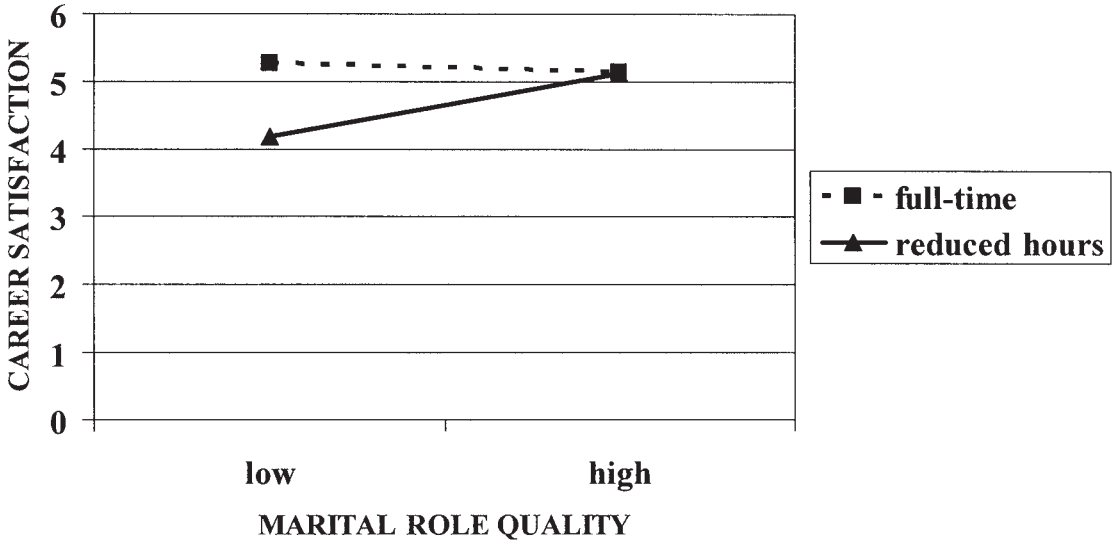


FIG. 1. Effect of marital role quality on career satisfaction, moderated by work schedule.

there was low parental role quality, reduced hours physicians were more likely than their full-time counterparts to express a greater intention to leave their jobs within 1 year.

Finally, full-time and reduced hours physicians in our sample differed in the relationship between work-family interference and intention to leave their jobs ($p = 0.038$). As shown in Figure 4, for full-time physicians, there appears to be little relationship between work-family interference and intention to leave the job. In contrast, high work-family interference for reduced hours physicians was associated with a higher intention to leave their jobs within 1 year.

DISCUSSION

In this sample, family experiences were more strongly associated with professional outcomes among women physicians who worked reduced hours than among their full-time counterparts. Career satisfaction and reported likelihood of leaving their job within a year for reduced hours women physicians were strongly related to the quality of their marital and parental roles, as well as their sense of work interfering with family. For full-time women physicians, there was much less association between family life and professional outcomes. Interestingly, when marital role qual-

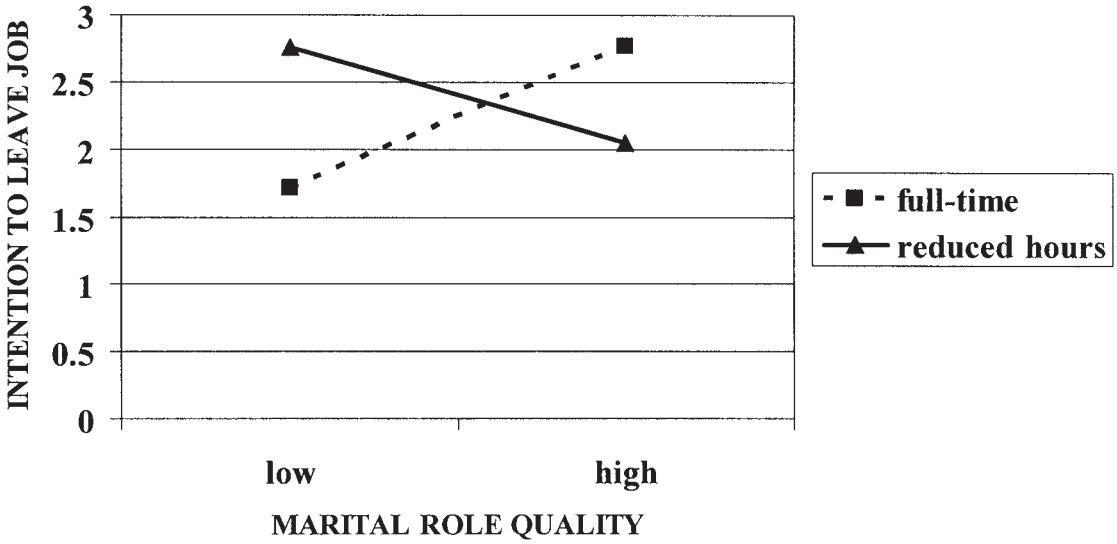


FIG. 2. Effect of marital role quality on intention to leave job, moderated by work schedule.

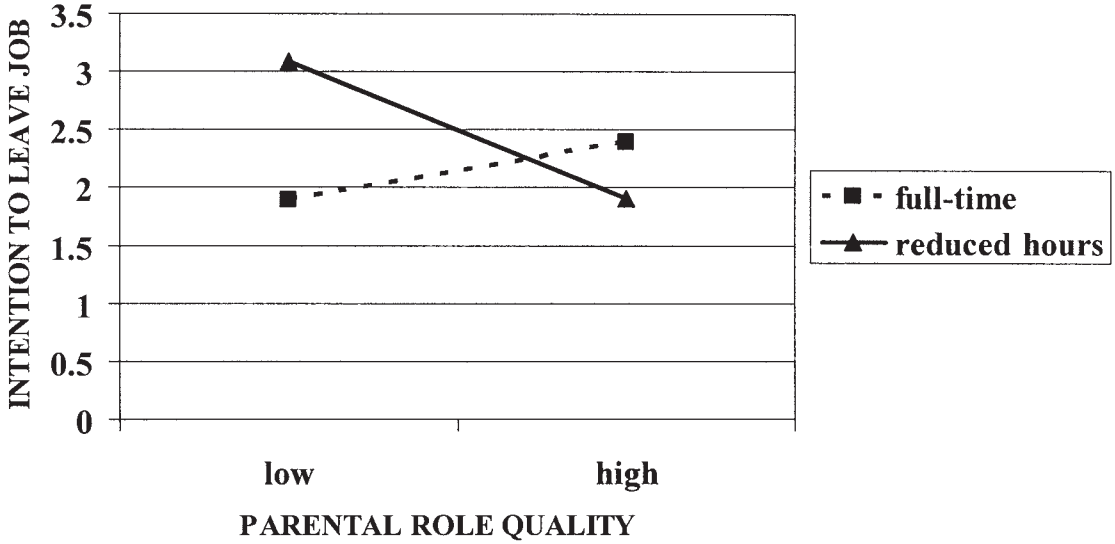


FIG. 3. Effect of parental role quality on intention to leave job, moderated by work schedule.

ity was high, full-time physicians expressed a greater intention to leave their job than did reduced hours physicians. It may be more difficult to spend greater hours away from home in these circumstances for full-time physicians, whereas part-time physicians may have better balanced the time they have to spend with their families in their work schedules. As this is a cross-sectional study, we have no way of knowing if the high marital role quality is the predictor or the outcome of the higher intention to leave the job.

Equal to or higher productivity and job performance of part-time women physicians compared with their full-time counterparts have been found in a number of studies,[5,10] comparing favorably to findings in other professions.[11,12] Women physicians who choose reduced hours work may be more concerned about the ways in which they meet their obligations both at work and at home and may find that such schedules improve their ability to meet both standards. More detailed research into the reasons that women choose reduced hours, other than the overarching reason of accommodating family, is required to answer these questions.

There is a general trend in studies of various populations for both men and women to report that they want jobs that allow them to be invested

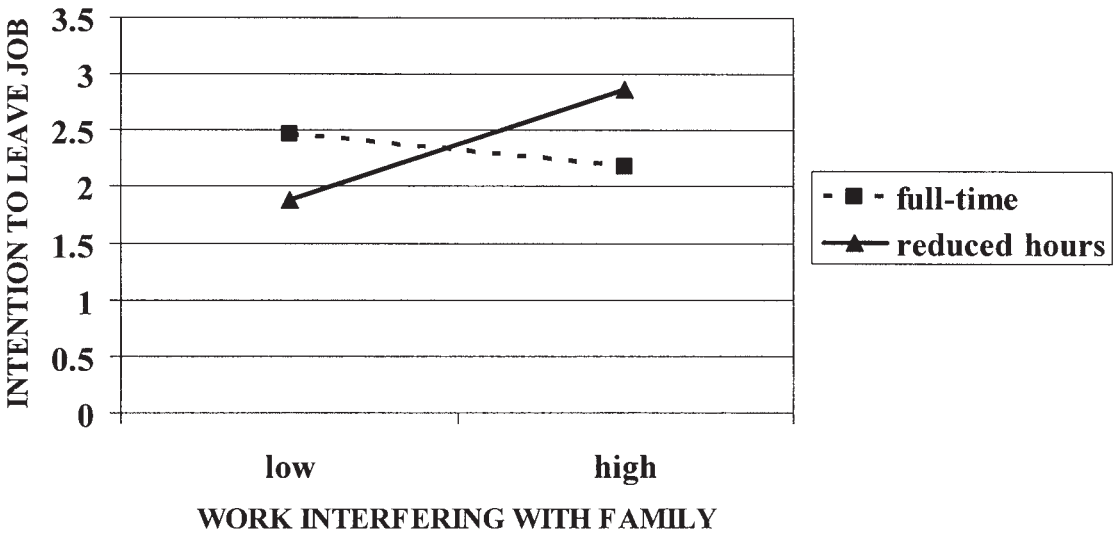


FIG. 4. Effect of work-family interference on intention to leave job, moderated by work schedule.

both in their careers and in their families. For example, in a survey of young men and women, both reported that family is at least as important to them as work,[22] a study of lawyers found that men and women alike want to be both challenged at work and engaged at home,[23] and an international study of high-level male and female executives found that a sizeable subgroup (32%) were what Galinsky has termed "dual-centric," placing the same priority on their personal and family lives as on work.[24] With the increase in dual-earner couples, both men and women physicians appear to be seeking careers in medicine that permit reasonable lifestyles and time for family. The most sought after residencies include radiology, anesthesiology, dermatology, emergency medicine, and other specialties with predictable hours and high hourly earnings.[25] Primary care, including general medicine, family practice, and, to a lesser extent, pediatrics, are much less sought after, with only 52.5% of residency positions filling with U.S. graduates in internal medicine, 56.2% in family medicine, and 74.5% in pediatrics in the 2004 match.[25] It has even been suggested that predictions of physician shortages in some areas could accompany the increase in women physicians because of increasing demands for reduced hours positions.[1] This scenario makes the retention of reduced hours physicians assume even greater importance.

There are a number of limitations to our study. The time-consuming and in-depth quality of this work does not permit large sample sizes, which may restrict the generalizability of our findings. Similarly, our data are from one geographic region, which may vary from other areas in terms of the percent of practice in managed care contracts, other styles of practice, and unmeasured regional variations.

There are a number of strengths to our study. We have examined the relationship of marital role quality, parental role quality, and work interfering with family with intention to leave the job and career satisfaction for women physicians working reduced hours and full-time. Such a study has not previously been done. Dual-earner couples increase the complexity of job retention for employers, with factors relating to both jobs influencing the ability of retaining an employee. Better understanding of the relationship of these parameters to satisfaction and retention can enlighten employers to promote wise policies for part-time women physicians, potentially decreasing the economic costs involved in early loss of physicians from the workforce.

CONCLUSIONS

Part-time women physicians appear to be more sensitive than their full-time peers to family experiences in terms of their career satisfaction and their reported intention to leave their jobs. With higher productivity, similar patient satisfaction and ambulatory costs, as well as equal to higher-quality performance, employers would be wise to encourage programs for reduced hours physicians, improve conditions, and permit flexible work patterns for physicians who desire part-time or reduced hours employment.

REFERENCES

1. Cull WL, Mulvey HJ, O'Connor KG, Sowell DR, Berkowitz CD, Britton CV. Pediatricians working part-time: Past, present, future. Pediatrics 2002;109:1015.
2. Carr PL, Gareis KC, Barnett RC. Characteristics and outcomes for women physicians who work reduced hours. J Wom Health 2003;12:399.
3. Frank E, McMurray JE, Linzer M, Elon L. Career satisfaction of U.S. women physicians: Results from the Women Physician's Health Study. Arch Intern Med 1999;159:1417.
4. Robinson GE. Career satisfaction in female physicians. JAMA 2004;291:635.
5. Warde C. Time is of the essence. J Gen Intern Med. 2001;16:712.
6. Fairchild DG, McLoughlin KS, Gharib S, et al. Productivity, quality, and patient satisfaction: Comparison of part-time and full-time primary care physicians. J Gen Intern Med. 2001;16:663.
7. Hartwell JK, Barnett RC, Borgatti S. When medical managers' beliefs about reduced-hour physicians. J Health Organization and Management 2004;18:262.
8. Parkerton PH, Wagner EH, Smith DG, Straley HL. Effect of part-time practice on patient outcomes. J Gen Intern Med. 2003;18:717.
9. Murray A, Safran DG, Roger WH, Inui T, Chang H, Montgomery J. Part-time physicians: Physician workload and patient-based assessments of primary care performance. Arch Fam Med. 2000;9:327.
10. Committee on Part-time Career Tracks in Clinical and Investigative Medicine. Interim Report, April 2000 (unpublished report). Boston, MA: Brigham and Women's Hospital, 2000.
11. Olmsted B, Smith S. Creating a flexible workplace: How to select and manage alternative work options. New York: AMACOM, 1994.
12. Schor JB. The overworked American: The unexpected decline of leisure. New York: Basic Books, 1991.

13. Levinson W, Kaufman K, Bickel J. Part-time faculty in academic medicine: Present status and future challenges. Ann Intern Med 1993;119:220.
14. Carr PL, Ash AS, Friedman RH, et al. Relation of family responsibilities and gender to the productivity and career satisfaction of medical faculty. Ann Intern Med 1998;129:532.
15. Barnett RC, Brennan RT, Marshall NL. Gender and relationship between parent-role quality and psychological distress. J Fam Issues 1994;15:229.
16. Hyde JS, Plant EA. Factor structure of Barnett's Partner Rewards and Concerns scale. Unpublished manuscript, University of Wisconsin, 1996.
17. Barnett RC, Marshall NL, Raudenbush SW, Brennan RT. Gender and the relationship between job experiences and psychological distress: A study of dual-earner couples. J Pers Soc Psychol 1993;65:794.
18. MacDermid SM, Barnett R, Crosby F, et al. The measurement of work/life tension: Recommendations of a virtual think tank. Boston, MA: Alfred P. Sloan Foundation, 2000.
19. Spielberger, CD. In: Form Y, ed. Manual for the state-trait anxiety inventory. Palo Alto: Consulting Psychologist Press, 1983.
20. Brennan RT, Barnett, RC. Negative affectivity: How serious a threat to self-report studies of psychological distress? Women's health: Research on gender, behavior, and policy. 1998;4:369.
21. Burke MJ, Brief AP, George JM. The role of negative affectivity in understanding relations between self-reports of stressors and strains: A comment on the applied psychology literature. J Appl Psychol 1993; 78:402.
22. Radcliffe Public Policy Center. Life's work: Generational attitudes toward work and life integration. Cambridge, MA: Harvard University, Radcliffe Institute for Advanced Study, 2000.
23. Boston Bar Association Task Force on Professional Challenges and Family Needs. Facing the grail: Confronting the cost of work-family imbalance. Boston, MA: Boston Bar Association, 1999.
24. Galinsky, E. Dual-centric: A new concept of work-life. New York: Families and Work Institute, 2003.
25. National Resident Matching Program. Results and data 2004 Match. AAMC Publications, Washington, DC, 2004.

Address reprint requests to:
Phyllis L. Carr, M.D.
Associate Dean of Student Affairs
Boston University School of Medicine
715 Albany Street, Room L109
Boston, MA 02118

E-mail: plcarr@bu.edu

DOI: 10.1111/j.1475-6773.2005.00436.x

Developing a Quality Measure for Clinical Inertia in Diabetes Care

Dan R. Berlowitz, Arlene S. Ash, Mark Glickman, Robert H. Friedman, Leonard M. Pogach, Audrey L. Nelson, and Ashley T. Wong

Objective. To develop a valid quality measure that captures clinical inertia, the failure to initiate or intensify therapy in response to medical need, in diabetes care and to link this process measure with outcomes of glycemic control.
Data Sources. Existing databases from 13 Department of Veterans Affairs hospitals between 1997 and 1999.
Study Design. Laboratory results, medications, and diagnoses were collected on 23,291 patients with diabetes. We modeled the decision to increase antiglycemic medications at individual visits. We then aggregated all visits for individual patients and calculated a treatment intensity score by comparing the observed number of increases to that expected based on our model. The association between treatment intensity and two measures of glycemic control, change in HbA1c during the observation period, and whether the outcome glycosylated hemoglobin (HbA1c) was greater than 8 percent, was then examined.
Principal Findings. Increases in antiglycemic medications occured at only 9.8 percent of visits despite 39 percent of patients having an initial HbA1c level greater than 8 percent. A clinically credible model predicting increase in therapy was developed with the principal predictor being a recent HbA1c greater than 8 percent. There were considerable differences in the intensity of therapy received by patients. Those patients receiving more intensive therapy had greater improvements in control ($p < .001$).
Conclusions. Clinical inertia can be measured in diabetes care and this process measure is linked to patient outcomes of glycemic control. This measure may be useful in efforts to improve clinicians management of patients with diabetes.

Key Words. Diabetes mellitus, outcomes assessment, quality of health care

Central to improving clinical practice is reliable and valid measures of the quality of care. While process and outcome measures each play an important role in quality measurement, it has long been recognized that the development of process measures that are linked to outcomes is an important

health services research goal (Brook, McGlynn, and Cleary 1996). The presence of such links helps validate the process measure and quality improvement resources may be redirected to those processes shown to have the greatest impact on patient outcomes (Hammermeister et al. 1995). Yet establishing links between process and outcome measures in observational studies is not easy and paradoxical results in which more intensive care produces worse outcomes often arise because of confounding by indication (Rubin, Pronovost, and Diette 2001). We now attempt to develop a quality measure for describing clinicians' practices in the pharmacological management of diabetes mellitus and to link this process measures to important intermediate outcomes of glycemic control.

Diabetes is a common medical problem that often has a significant negative impact on a patient's health status. Studies of the quality of diabetes care have frequently demonstrated a wide gap between recommended medical practices and the care that diabetes patients actually receive (Saaddine et al. 2002). This is particularly true for the management of hyperglycemia in which many patients have inadequate glycemic control (O'Connor et al. 1996; Chin et al. 2000; Harris 2000; El-Kebbi et al. 2001). Guidelines published by the American Diabetes Association (2003) identify hemoglobin A1c (HbA1c) of less than 7 percent as desirable and have long recommended that clinicians take action when HbA1c levels are greater than 8 percent. Diabetes treatment protocols typically encourage stepped intensification of pharmacological therapy until goals for glycemic control are achieved (Mazze et al. 1994; Abraira et al. 1995; El-Kebbi et al. 1997). Yet medical practice studies find that clinicians often do not follow these

---

Address correspondence to Dan R. Berlowitz, M.D., M.P.H., CHQOER, ENRM VA Hospital, 200 Springs Road, Bedford, MA 01730. Dan R. Berlowitz, Director, Mark Glickman, Ph.D., and Ashley T. Wong, M.A., are with the Center for Health Quality, Outcomes & Economic Research, Edith Nourse Rogers VA Hospital, Bedford MA. Drs. Berlowitz and Glickman are also with the Department of Health Services, Boston University School of Public Health, Boston. Arlene S. Ash, Ph.D., is with the Health Care Research Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center, Boston. Robert H. Friedman, M.D., is with the Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center, Boston. Leonard M. Pogach, M.D., M.B.A., is with the VA New Jersey Health Care System for Health Care; Center for Healthcare Knowledge Management, University of Medicine and Dentistry of New Jersey, New Jersey Medical School, East Orange, NJ. Audrey L. Nelson, Ph.D., R.N., F.A.A.N., Associate Chief of Nursing Service, is with the James A. Haley Hospital, Tampa, FL. Dr. Nelson is Director, Patient Safety Center of Inquiry, and Director, HSR&D REAP on Patient Safety Outcomes, and also a Co-Director, AHRQ Patient Safety Research Center on Safe Patient Transitions, VA Patient Safety Center of Inquiry, Tampa, FL.

recommendations, failing to increase antiglycemic medications despite persistently elevated blood glucose and HbA1c levels (El-Kebbi et al. 1997; Wetzler and Snyder 2000).

Recently, the term clinical inertia has been proposed to define the phenomenon by which health care providers fail to initiate or intensify therapy when indicated (Phillips et al. 2001). Moreover, it is increasingly recognized as among the most important barriers to achieving control of chronic medical conditions (Javors and Bramble 2003). Efforts to improve care by identifying and minimizing clinical inertia may then have a significant impact in enhancing outcomes for patients with diabetes. Before this can be accomplished, however, one must be able to measure clinical inertia.

Two measures of clinical inertia in diabetes care have been used previously: the proportion of patients having an intensification of pharmacological therapy at visits with elevated measures of glycemic control (El-Kebbi et al. 1997; Wetzler and Snyder 2000) and the proportion of patients with elevated HbA1c levels on a greater-than-starting dose of medications (Grant et al. 2002). These measures have typically examined only a short time-frame and do not reflect the many clinical factors that influence treatment decisions. Furthermore, they have not been linked to outcomes, perhaps because such attempts are likely to suffer from confounding by indication (Rubin, Pronovost, and Diette 2001). Confounding by indication refers to the fact that poor glycemic control is simultaneously an indication for treatment intensification and a predictor of poor outcome. Thus, unless we adjust for variations in patient status during the course of treatment, we are likely to see more intense therapy associated with worse glycemic control. Recently, we proposed an alternate approach to measuring intensity of pharmacological therapy in hypertension care that compares the observed number of increases in therapy over an extended time period to the number predicted by empirically modeling factors that affect the probability of an increase in therapy at individual visits (Berlowitz et al. 1998). Improved performance on this process measure was associated with better blood pressure control, reversing the negative association (likely because of confounding by indication) that appeared when predicting outcomes from aggregated measures of treatment intensity adjusted only for baseline patient characteristics. This intensity score approach has since been placed within the general framework of causal modeling for the effect of a time-varying treatment via structural nested mean models (Brumback et al. 2003).

We now develop a similar quality measure for measuring clinical inertia in diabetes care by evaluating the pharmacologic management of diabetes in

patients with ready access to clinicians and medications through the Department of Veterans Affairs (VA). To accomplish this, we perform the following three tasks. First, we identify factors associated with the decision to increase antiglycemic medications at individual patient care visits. Second, using the resulting model, we describe the intensity of antiglycemic medication therapy received by patients over an approximately 16-month period. Finally, we examine the link between this process measure and outcomes by determining whether those patients receiving more intensive therapy had improved glycemic control.

METHODS

Data Sources

Study data were obtained from two sources. First, we used two national VA databases, the Outpatient Clinic File (OPC) and the Patient Treatment File (PTF), which contain information on all outpatient and inpatient medical encounters. This included diagnoses described as ICD-9-CM codes, dates of encounters, and types of outpatient clinic visits. Second, we used the Computerized Patient Record System, available through each VA medical center, to obtain pharmacy records and the results of laboratory studies. Pharmacy records described the name, dosage, and refill date for each dispensed medication and diabetes-related equipment. Because these treatments are available to eligible veterans either for free or a small copayment, the VA is likely to be their primary source for diabetes-related medications.

Study Subjects

We studied veterans with diabetes mellitus receiving regular medical care at 13 VA medical centers located in two Veterans Integrated Service Networks, New England and Florida–Puerto Rico. Study patients met the following criteria. First, they had to have at least two primary care visits during the 6-month period beginning October 1, 1997. We randomly selected one primary care visit from this period as the "index" visit. Second, patients had to have at least one primary care visit between 1 and $1\frac{1}{2}$ years after the index visit. If more than one such visit was available, we randomly selected one as the "outcome" visit. Third, patients needed a diagnosis of diabetes mellitus. Diabetes was considered present if a diabetes ICD-9-CM code (250.x [diabetes mellitus], 357.2 [polyneuropathy in diabetes], or 362.0x [diabetic retinopathy]) was recorded in VA databases on two occasions at least 7 days apart during the

1-year period beginning April 1, 1997. At least one diabetes diagnosis had to be from an outpatient encounter and at least one had to be no later than the index visit. Ninety-five patients with more than 40 visits per year were excluded from this sample. Since onset of diabetes as a child would preclude military service, most patients had type 2 diabetes.

Variable Selection and Statistical Analyses

Analyses were performed in three steps. For the initial analytic step, we considered each individual visit contributed by patients in our study sample. We used visits as the focus of our analysis because we consider each visit as an opportunity for the clinician to assess the patient and make an informed judgment about whether or not to alter medication therapy. We constructed a model to predict the probability that an individual patient-care visit would result in an increase in antiglycemic medications based on characteristics at the time of that visit. We considered only those visits in which diabetes management was likely to occur, specifically visits to all primary care clinics and selected subspecialties of internal medicine such as diabetes/endocrinology. These visits were identified using the clinic stop variable from the OPC. An increase in antiglycemic medications was considered present at a visit if a new antiglycemic medication was started or the dosage of an existing medication was increased. Since increases in medications identified in the pharmacy database may not occur on the same day as a visit, we assigned any increase to the prior visit if it was within 14 days after the visit; otherwise it was assigned to the next visit. We chose a 14-day window after examining dates of increases in relation to visit dates. Use of a window acknowledges that clinicians may order laboratory tests at a visit and then subsequently contact a patient to change his medications. Each visit was thus classified as antiglycemic medication "increase" or "no increase." Visits occurring after a patient was prescribed an insulin "sliding scale," as indicated by this specific designation in the *sig* statement of the prescription, were not considered as these patients were likely adjusting their own medications. This was present in 594 patients.

We used clinical judgment to identify available factors likely to influence the decision to increase antiglycemic medication therapy. These potential predictors included demographic characteristics of the patients; results of laboratory tests such as measures of glycemic control, lipid levels, renal function, and urinalysis; measures of health care utilization including recent hospitalizations and the time since the patient's last outpatient visit; pharmacy inform-

ation including steroid use, insulin, and prescription of glucose monitoring equipment; and diagnoses. Diagnoses were used to identify the presence of specific diabetic complications (neuropathy, retinopathy, nephropathy, or peripheral vascular disease), other cardiovascular risk factors (hypertension, hyperlipidemia, tobacco abuse disorder), presence of any psychiatric or substance abuse disorders, and a summary measure of comorbidity burden as described by the Deyo modification of the Charlson index (Charlson et al. 1987; Deyo, Cherkin, and Ciol 1992). Diagnoses were considered present if there was a single ICD-9-CM code indicating the appropriate condition in the year prior to the index visit, an approach that we have previously found to have good agreement with medical records (Borzecki et al. 2004). Predictors were coded in alternate ways before deciding on the exact variation used in modeling. For example, for HbA1c at the time of a visit we considered different ways of coding whether any value was available, how long since the last determination, the results of the most recent determination, and results of the second most recent determination. We settled on eight different categories of availability and level (not available, ≤ 6, > 6 but ≤ 7, > 7 but ≤ 8, > 8 but ≤ 8.5, > 8.5 but ≤ 9, > 9 but ≤ 10, and > 10), in addition to a separate variable for time since last determination. Of note, VA databases do not accurately differentiate fasting from nonfasting laboratory values.

We modeled the decision to increase antiglycemic medications at a visit using classification and regression trees (CART) (Breiman 1984), a recursive partitioning algorithm as implemented in *S-plus* software version 3.3. This procedure repeatedly splits the data to create a division into subgroups of visits that are internally similar with respect to their probability of a medication increase. At each step, the procedure splits an existing group into two based on the values of a single independent variable. The variable to be split is chosen to result in the greatest degree of separation in the outcome. The model-estimated probability of a medication increase at a visit is the empirical frequencies of the outcome in each terminal group. The choice of the number of terminal groupings was guided by a cross-validation procedure in which we divided the sample into five equal-sized subsets. Then, for each possible number of terminal groups, ranging from 2 to 150, a model of that size was fit to each combination of four of the five subsets, and applied to the remaining subset. Based on the actual outcomes and estimated probabilities, *c*-statistics were calculated (Altham 1973). The final number of terminal groupings was chosen by examining models with high cross-validated *c*-statistics that conform to clinical judgments about appropriate explanatory variables.

We were concerned that many increases in insulin therapy might not be recorded in the pharmacy database. Thus, we initially developed separate models for visits in which the patient was receiving insulin and for noninsulin visits. However, as the models were similar, and increases in insulin therapy were frequently evident in the pharmacy database, we combined all visits to develop a single model.

For the next step in our analysis, we aggregated information from each individual visit for a patient to create a patient-specific measure of the intensity of medication therapy. This measure is norm-based in that it compares the intensity any given patient receives relative to the experience of other patients. We defined treatment intensity as a ratio with the numerator equal to the actual number of medication increases from the index to outcome visit minus the expected number of increases as predicted by the model for each visit. The denominator was the total number of visits. Scores on this treatment intensity measure must lie between -1 and $+1$; positive values indicate more increases in therapy than expected.

In the final step, we examined the association between this process measure describing treatment intensity and two outcome measures of glycemic control. We used a linear regression model to explain the change in HbA1c levels between the index and outcome visit. The index visit HbA1c was the determination closest in date to the index visit during the 6 months surrounding the visit. As laboratory data after the outcome visit were not available, the outcome visit HbA1c was the determination closest to the outcome visit during the 6 preceding months. We also used a logistic regression model to explain the dichotomous outcome of "outcome visit HbA1c >8 percent" yes/no, adjusting for index visit HbA1c. For both models, we adjusted for other variables likely associated with glycemic control (Zhang et al. 2000). Patients without a HbA1c determination for both the index and outcome visit were excluded from these regression analyses.

RESULTS

The study sample consisted of 23,291 patients with 266,309 visits. Patient characteristics are described in Table 1. Nearly 29 percent of patients were on insulin at baseline, and 27 percent were on insulin for the entire study period. Among the 15,437 patients with a HbA1c determination at the index visit, the mean was 7.8 ± 1.9 percent; 39.1 percent had a value above 8 percent. Patients were frequent utilizers of VA health care. Yet following 15.8 ± 1.9

Table 1: Characteristics of the Study Patients*

	Total Sample (N= 23,291)	*With Index and Outcome HbA1c (N= 12,523)*	*Without Both Index and Outcome HbA1c (N= 10,768)*	*p-Value*†
Age (years)	65.3 ± 10.4	65.3 ± 10.1	65.3 ± 10.8	.69
Male gender (%)	97.2	97.4	97.0	.05
Diabetic complication (%)	41.6	43.1	39.7	<.0001
Additional cardiovascular risk factor (%)	78.2	79.5	76.7	<.0001
Charlson index‡	1.09 ± 1.57	1.06 ± 1.52	1.14 ± 1.63	<.0001
On insulin at baseline (%)	28.6	32.2	24.5	<.0001
Index HbA1c§	7.8 ± 1.9	7.8 ± 1.9	7.6 ± 1.9	—
Index HbA1c>8% (%)§	39.1	39.8	36.0	—
Outcome HbA1c¶	7.8 ± 1.6	7.8 ± 1.6	7.7 ± 1.7	—
Outcome HbA1c>8% (%)¶	39.4	40.3	36.6	—
No. of visits	11.4 ± 6.6	12.2 ± 6.7	10.6 ± 6.4	<.0001
Months between index and outcome visits	15.8 ± 1.9	15.9 ± 1.6	15.7 ± 2.2	<.0001

*Continuous variables described as mean ± standard deviation.

†*p*-value compares group with index and outcome HbA1c to group without.

‡Charlson index excludes diabetes.

§*N*= 15,437 for total sample and *N*= 2,914 for without both sample.

¶*N*= 16,800 for total sample and *N*= 4,277 for sample without both.

HbA1c, hemoglobin A1c.

months of care, with 8.8 ± 5.1 visits per year, the mean HbA1c level among the 16,800 patients with a determination was unchanged and 39.4 percent had a value above 8 percent.

An increase in antiglycemic medication therapy occurred at 9.8 percent of visits. An increase occurred at 13.1 percent of the visits when a patient was on insulin and 7.8 percent of visits with no insulin therapy. Insulin was started on 1,884 patients during the study period.

The strongest predictor of an increase was the most recent HbA1c being greater than 8 percent (Figure 1). Other factors associated with an increase included higher serum glucose, longer interval since the last visit (when several visits are close together, an increase is less likely at any individual visit), HbA1c being performed within 3 months, being on insulin, and having received self-monitoring of blood glucose supplies. An increase occurred 32.0 percent of the time among the 10,581 visits in which the most recent HbA1c was greater than 8 percent, that HbA1c had been performed within the past

Figure 1: Model Used to Describe Factors at Individual Visits Associated with the Decision to Increase Antiglycemic Medications

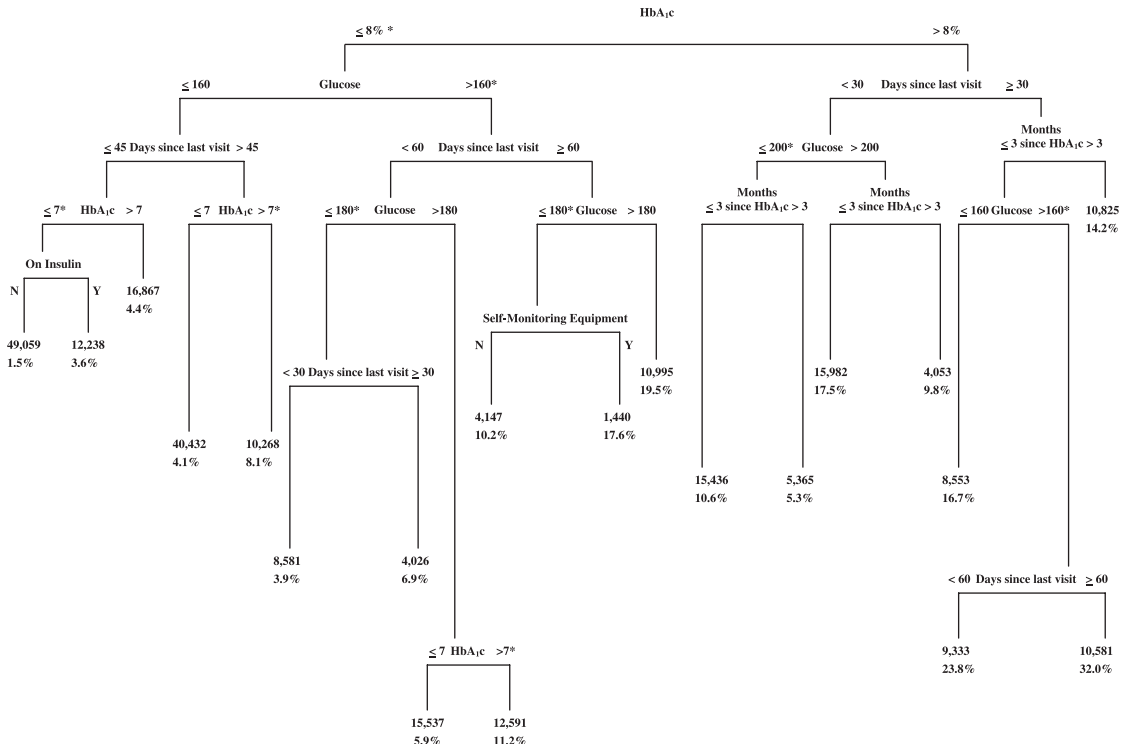


Table 2: Multivariate Regression Models* Relating Intensity of Antiglycemic Medication Therapy to Glycemic Control

	Change in HbA1c		*Outcome HbA1c > 8%*	
Variable	*Coefficient*	*p-Value*	*Odds Ratio*	*95% CI*
Intensity of therapy (per increase of 0.1)	− 0.064	<.0001	0.91	0.88, 0.98
Index visit HbA1c	—	—	1.76	1.72, 1.81
Age (per increase of 10 years)	0.013	.39	0.78	0.75, 0.81
Male gender	0.147	.13	1.12	0.86, 1.45
Any diabetes complication	− 0.131	<.0001	1.29	1.18, 1.41
Charlson index	0.026	.01	1.02	0.99, 1.05
Psychiatric diagnosis	0.156	.0002	1.03	0.92, 1.15
Alcohol/substance abuse diagnosis	− 0.169	.06	0.79	0.62, 1.00
Oral steroid use	− 0.022	.75	0.83	0.69, 1.00

*A linear model was used when modeling outcome visit HbA1c minus index visit HbA1c; a negative coefficient indicates improved glycemic control. A logistic model was used to predict odds of an outcome HbA1c > 8%; an odds ratio less than 1.0 indicates better control.

HbA1c, hemoglobin A1c; CI, confidence interval.

three months, the serum glucose was greater than 160 mg/dl, and the prior visit was more than 60 days previously. In contrast, an increase occurred at only 1.5 percent of visits when the most recent HbA1c was less than 7 percent, the serum glucose was less than or equal to 160 mg/dl, the prior visit was within the past 45 days, and the patient was not on insulin therapy.

The expected number of increases per patient varied with some patients having all their visits in the group with the lowest predicted probability of an increase (1.5 percent), and others having each visit with a 32 percent predicted probability of an increase. Patient treatment intensity scores also varied among patients with a range from − 0.32 to 0.96. Twenty-eight percent of patients had intensity scores greater than 0.05, values we viewed as more intensive therapy than the norm, and 34 percent had values less than –0.05, indicating less intensive therapy. Mean intensity scores were slightly higher for patients with an index HbA1c > 8 percent compared with those with one < 8 percent (0.02+0.15 versus 0.01+0.11, $p < .001$).

More intensive antiglycemic medication therapy was significantly associated with improved glycemic outcomes. An index and outcome HbA1c was available in 12,523 patients (54 percent of the sample). These patients differed from the 10,768 patients without these determinations (Table 1). Among patients in the highest quintile of treatment intensity scores, with scores greater than 0.09, HbA1c values declined by 0.12 percent between the index

and outcome visit. In patients receiving less intensive therapy, the change in HbA1c values worsened progressively. In those patients in the lowest quintile, with intensity scores less than −0.08, HbA1c values increased by 0.09 percent. Fifty-seven percent of patients in the lowest quintile had an outcome visit HbA1c greater than 8 percent, as compared with 36 percent among the remaining patients. These associations were significant in regression models that adjusted for other patient characteristics (Table 2). Each 0.1 increase in the intensity score, equivalent to one additional visit during which there was an increase in therapy beyond that expected over 10 visits, was associated with an additional 0.06 percent decline in the HbA1c and an odds ratio of 0.91 that the outcome HbA1c would be greater than 8 percent. Regression models were almost identical when limited to the subsample of patients on insulin therapy for the entire study period.

DISCUSSION

Clinical trials have convincingly demonstrated that intensive medication management of diabetes improves glycemic control and prevents the development of microvascular disease (The Diabetes Control and Complications Trial Research Group 1993; U.K. Prospective Diabetes Study [UKPDS] Group 1998). Despite these facts, many patients with diabetes do not achieve recommended levels of glycemic control. Clinical inertia is now recognized as an important barrier contributing to inadequate glycemic control (Phillips et al. 2001). We now propose a quality measure that can be used to measure this phenomenon. This measure works by interpreting the total treatment received in the context of the level of treatment expected as a result of a sequence of conditional decisions. Thus, the same amount of treatment is viewed as more intense for a person with better glycemic control during the treatment period than for a person who was seen repeatedly with poor control.

Our approach to measuring clinical inertia in diabetes care has numerous advantages over past efforts (El-Kebbi et al. 1997; Wetzler and Snyder 2000; Grant et al. 2002). First, by evaluating care over an extended time period, our method recognizes that at any single visit, not increasing pharmacological therapy might be an appropriate action. However, over a longer time period, several increases might be indicated if treatment goals are not met. Second, our approach recognizes that many clinical factors might be considered in the decision to increase therapy, including not only the level of hyperglycemia, but also the recency of these lab tests, the types of therapy

being received, and the time since the previous visit. Third, our measure has face validity in that it is based on widely accepted clinical practice recommendations from the time frame of this study. The strongest predictor of an increase in therapy was a HbA1c greater than 8 percent. This is consistent with the American Diabetes Association recommendations that encourage action when the HbA1c is above this level, and also reflects many clinicians' beliefs that lifestyle modifications will not be sufficient to achieve tight control at these higher levels. Interestingly, HbA1c levels further above 8 percent were not associated with a greater likelihood of an increase although higher blood glucose levels were an important predictor. Fourth, our measure appears to be sensitive to differences in practice. There were large differences in the intensity of therapy that patients received with some patients receiving many more increases in therapy than the norm.

Finally, and most importantly, our process measure was linked to outcomes in that patients who received more intensive management were more likely to achieve better outcomes in glycemic control. We evaluated two measures of glycemic control and found for each that more intensive therapy was associated with significantly better results. Establishing such links between process and outcome measures is an important goal in quality assessment. It further validates our measure of intensity of therapy and suggests that we are capturing an important aspect of care. These results also emphasize that interventions to improve care should focus on the problem of clinical inertia.

Past studies in diabetes care have generally failed to establish such links between process and outcomes (Williams et al. 1967; Romm and Hulka 1979, 1980). Considerable advances have been made since these studies in developing methods, especially those that use a "propensity to receive treatment" score, to reduce bias when using observational studies to determine treatment effects (D'Agostino 1998). Approaches that only adjust for baseline covariates, however, do not work well when examining a sequence of treatment encounters, such as in diabetes care, where there may be time-dependent confounding. A number of analytic strategies have been described for this type of problem (Robins 1997; Robins, Hernan, and Brumback 2000) that recognize the need to interpret the care given at each encounter in the context of the patient's status at that encounter. As recently discussed, our intensity score is a readily interpreted measure in the spirit of G-estimators, which represent the effect of a "generalized treatment regime," for structural nested models (Brumback et al. 2003). Our diabetes intensity of medication therapy measure was able to associate more care with better outcomes, as was a similar intensity measure that we developed in hypertension (Berlowitz et al. 1998). These

findings are consistent with numerous studies showing that normative practice in these areas is insufficiently aggressive. While we cannot prove that the measures are bias-free, they are helpful in addressing the problem of confounding by indication.

Our intensity of therapy measure may be used in a variety of ways. Quality improvement programs in diabetes care could profile clinicians' practices not only on the basis of their glycemic control but also in their pharmacological management of diabetes. Clinicians and settings with higher intensity scores and better glycemic control could be identified for benchmarking. Interventions to improve care could be developed that provide clinicians with feedback on their performance on this measure. Such interventions that focus on how clinicians decide to increase therapy may be particularly successful (Cook et al. 1999) and could be directed at that subset of clinicians whose patients have poor glycemic control in the setting of low intensity therapy. As our intensity measure relies on data available from existing databases, such efforts would be relatively easy to implement. The measure may also be used to study care provided to important subgroups of patients such as ethnic minorities and individuals with mental health illnesses. For example, we recently studied hypertension control in patients with diabetes and found not only did they have worse blood pressure control than patients without diabetes, but they also were receiving less intensive therapy (Berlowitz et al. 2003). Thus, we were able to conclude that poor blood pressure control in patients with diabetes is not solely related to more difficult to treat disease, but also to the fact that clinicians were treating these patients differently.

Our results provide additional evidence regarding the presence of clinical inertia. They are similar to past studies in that many patients had inadequate glycemic control. Over 39 percent had an initial HbA1c greater than 8 percent, and following an additional 16 months of care with more than 11 visits on average, glycemic control in the population was essentially unchanged. Despite less than optimal control, increases in antiglycemic medication therapy were relatively infrequent, occurring in only 9.8 percent of visits. Even among those visits with the greatest indication for an increase, an actual increase occurred only 32 percent of the time. Thus, many opportunities for increasing therapy to achieve better glycemic control were being missed. In deciding on whether to increase therapy, our results suggest that clinicians tend to focus on relatively few factors.

The process measure that we propose is norm-based; it compares clinicians to the usual performance of other providers. While this is often the standard approach when comparing outcomes such as risk-adjusted mortality,

it is used less frequently for process measures that are often absolute. As with many HEDIS-type measures, either the appropriate process is performed or it is not. The specific situation in which performance of the process is indicated is usually well identified, for example "when the HbA1c is elevated at a visit, therapy should be increased." This approach has a number of advantages including being easy to comprehend, indicating specific subgroups in which intensification is indicated, and being less susceptible to gaming by increasing therapy in patients with well-controlled diabetes. However when examining a complex behavior such as medication intensification, which typically takes place over several visits and in which there is significant provider discretion with many factors contributing to the decision, we believe that a norm-based approach has many advantages. Because our approach is norm-based, care is required in using the model to calculate intensity scores for non-VA settings.

Unique to our approach of measuring clinical inertia is that we combine information from multiple visits. We determine an expected probability for an increase in therapy at each individual visit and then sum the expected probabilities over all visits. It is this value we compare with the observed number of increases. In calculating these expected probabilities, we used CART modeling because of our belief that clinicians' actions are very much guided by clusters of signs and symptoms as well as discrete thresholds for action as emphasized in national guidelines. Additionally, CART models are useful for identifying interactions among predictors. However, CART models have also been criticized because subgroups may be difficult to interpret clinically and simple rules may not be uncovered (Marshall 2001). We also evaluated a logistic model and found it had similar discriminative ability as the CART model.

Many factors go into the clinical decision on increasing therapy at an individual visit; our model captures only a few such factors. We cannot determine whether individual decisions to increase or not increase therapy might be appropriate for a patient. Such a determination would require a careful weighing of the risks and benefits of changes as well as patient preferences. This level of detailed information could not be obtained from existing databases. This emphasizes the importance of applying our measure to studying groups of patients rather than the decisions made for individuals.

We did not measure adherence to medications in our study. Clinicians may not be intensifying therapy because of concerns that patients were nonadherent. In at least one small study, however, this was not found to be the case (Javors and Bramble 2003). Further work should examine this issue. It could be that even a stronger link between intensity of therapy and glycemic

control would be demonstrated after considering adherence. However, the fact that we did demonstrate such a link without considering adherence suggests the validity of our approach.

Several additional limitations should be noted. First, our study population was highly selected, consisting mostly of elderly men with type 2 diabetes and access to medical care. Second, we had no information on non-VA medical care. However, most patients are likely to be using the VA for obtaining medications, and the population had high rates of utilization. Third, we relied on pharmacy databases to identify increases in therapy. While these databases are likely to capture most episodes of initiating a new therapy, increases in dosage of a prescribed medication, particularly for insulin, may be missed. Fourth, HbA1c levels were missing for many patients and these patients differed from patients with such information. We cannot be certain how this affected our results. Finally, the magnitude of changes in HbA1c associated with increased treatment intensity were not large. However, we followed patients for relatively short periods of time with some patients having as little as 6 months between determinations. This may not have been adequate time to see large changes.

It should also be noted that we focused on the decision by clinicians to increase antiglycemic medication therapy. We were interested in studying whether clinicians, when confronted with poor glycemic control, changed pharmacologic management. We did not capture other important aspects of care such as which medication was selected or how much of an increase in dosage occurred, as well as other actions that might be appropriate including enhanced patient education, efforts to improve dietary and medication adherence, or referral to a diabetes team. Additionally, we did not capture information related to potential overmedication, side effects, and risks of hypoglycemia. This may be a particular concern in an older population.

We do not know why clinicians did not increase therapy more frequently. Phillips et al. (2001), in describing clinical inertia, ascribed this phenomenon to three problems in clinical practice: physician overestimation of care provided, use of soft reasons to rationalize decisions not to increase therapy, and lack of training and organizational focus necessary to achieve therapeutic goals. However, El-Kebbi et al. (1999) have also described how failures to intensify therapy in a diabetes clinic are related to clinicians' perceptions that glycemic control is improving or that patients are nonadherent to therapy. Whatever the cause, overcoming clinical inertia is not likely to be easy, but it is essential if we are to substantially improve health outcomes for patients with diabetes. Our measure of treatment intensity may

be a first step in improving how clinicians prescribe medications used in the management of diabetes.

ACKNOWLEDGMENTS

This research was supported by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service, DII 99-205. 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

Abraira, C., J. A. Colwell, F. Q. Nuttall, C. T. Sawin, N. J. Nagel, J. P. Comstock, N. V. Emanuele, S. R. Levin, W. Henderson, and H. S. Lee. 1995. "Veterans Affairs Cooperative Study on Glycemic Control and Complications in Type II Diabetes (VA CSDM)." *Diabetes Care* 18 (8): 1113–23.

Altham, P. M. E. 1973. "A Non-Parametic Measure of Signal Discriminability." *British Journal of Mathematics, Statistics and Psychology* 26: 1–12.

American Diabetes Association. 2003. "Standards of Medical Care for Patients with Diabetes Mellitus." *Diabetes Care* 26: S33–50.

Berlowitz, D. R., A. S. Ash, E. C. Hickey, R. H. Friedman, M. Glickman, B. Kader, and M. A. Moskowitz. 1998. "Inadequate Management of Blood Pressure in a Hypertensive Population." *New England Journal of Medicine* 339 (27): 1957–63.

Berlowitz, D. R., A. S. Ash, E. C. Hickey, M. Glickman, R. Friedman, and B. Kader. 2003. "Hypertension Management in Diabetes: The Need for More Aggressive Therapy." *Diabetes Care* 26 (2): 355–9.

Borzecki, A. M., A. T. Wong, E. C. Hickey, A. S. Ash, and D. R. Berlowitz. 2004. "Identifying Hypertension-Related Comorbidities from Administrative Data: What's the Optimal Approach?" *American Journal of Medical Quality* 19 (5): 201–6.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees.* Monterey, CA: Wadsworth and Brooks/Cole.

Brook, R. H., E. A. McGlynn, and P. D. Cleary. 1996. "Measuring Quality of Care." *New England Journal of Medicine* 335 (13): 966–70.

Brumback, B., S. Greenland, M. Redman, N. Kiviat, and P. Diehr. 2003. "The Intensity-Score Approach to Adjusting for Confounding." *Biometrics* 59 (2): 274–85.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longtudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40 (5): 373–83.

Chin, M. H., S. B. Auerbach, S. Cook, J. F. Harrison, J. Koppert, L. Jin, F. Thiel, T. G. Karrison, A. G. Harrand, C. T. Schaefer, H. T. Takashima, N. Egbert, S. C. Chin, and W. L. McNabb. 2000. "Quality of Diabetes Care in Community Health Centers." *American Journal of Public Health* 90 (3): 431–34.

Cook, C. B., D. C. Ziemer, I. M. El-Kebbi, D. L. Gallina, V. G. Dunbar, K. L. Ernst, and L. S. Phillips. 1999. "Diabetes in Urban African-Americans. XVI. Overcoming Clinical Inertia Improves Glycemic Control in Patients with Type 2 Diabetes." *Diabetes Care* 22 (9): 1494–500.

D'Agostino Jr., R. B. 1998. "Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group." *Statistics in Medicine* 17 (19): 2265–81.

Deyo, R. A., D. C. Cherkin, and M. A. Ciol. 1992. "Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Databases." *Journal of Clinical Epidemiology* 45 (6): 613–9.

El-Kebbi, I. M., D. C. Ziemer, C. B. Cook, C. D. Miller, D. L. Gallina, and L. S. Phillips. 2001. "Comorbidity and Glycemic Control in Patients with Type 2 Diabetes." *Archives of Internal Medicine* 161 (10): 1295–300.

El-Kebbi, I. M., D. C. Ziemer, D. L. Gallina, V. Dunbar, and L. S. Phillips. 1999. "Diabetes in Urban African-Americans. XV. Identification of Barriers to Provider Adherence to Management Protocols." *Diabetes Care* 22 (10): 1617–20.

El-Kebbi, I. M., D. C. Ziemer, V. C. Musey, D. L. Gallina, A. M. Bernard, and L. S. Phillips. 1997. "Diabetes in Urban African-Americans. IX. Provider Adherence to Management Protocols." *Diabetes Care* 20 (5): 698–703.

Grant, R. W., E. Cagliero, P. Murphy-Sheehy, D. E. Singer, D. M. Nathan, and J. B. Meigs. 2002. "Comparison of Hyperglycemia, Hypertension, and Hypercholesterolemia Management in Patients with Type 2 Diabetes." *American Journal of Medicine* 112 (8): 603–9.

Hammermeister, K. E., A. L. Shroyer, G. K. Sethi, and F. L. Grover. 1995. "Why Is It Important to Demonstrate Linkages between Outcomes of Care and Processes and Structures of Care?" *Medical Care* 33 (10 Suppl): OS5–16.

Harris, M. I. 2000. "Health Care and Health Status and Outcomes for Patients with Type 2 Diabetes." *Diabetes Care* 23 (6): 754–8.

Javors, J. A., and J. E. Bramble. 2003. "Uncontrolled Chronic Disease: Patient Non-Compliance or Clinical Mismanagement?" *Disease Management* 6 (3): 169–78.

Marshall, R. J. 2001. "The Use of Classification and Regression Trees in Clinical Epidemiology." *Journal of Clinical Epidemiology* 54 (6): 603–9.

Mazze, R. S., D. D. Etzwiler, E. Strock, K. Peterson, C. R. McClave, J. F. Meszaros, C. Leigh, L. W. Owens, L. C. Deeb, A. Peterson, and M. Kummer. 1994. "Staged Diabetes Management." *Diabetes Care* 17 (Suppl 1): 56–66.

O'Connor, P. J., W. A. Rush, J. Peterson, P. Morben, L. Cherney, C. Keogh, and S. Lasch. 1996. "Continuous Quality Improvement Can Improve Glycemic Control for HMO Patients with Diabetes." *Archives of Family Medicine* 5 (9): 502–6.

Phillips, L. S., W. T. Branch, C. B. Cook, J. P. Doyle, I. M. El-Kebbi, D. L. Galina, C. D. Miller, D. C. Ziemer, and C. S. Barnes. 2001. "Clinical Inertia." *Annals of Internal Medicine* 135 (9): 825–34.

Robins, J. M. 1997. "Causal Inference from Complex Longitudinal Data." In *Latent Modelling with Applications to Causality*, edited by M. Berkane, pp. 69–117. New York: Springer Verlag.

Robins, J. M., M. A. Hernam, and B. A. Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11 (5): 550–60.

Romm, F. J., and B. S. Hulka. 1979. "Care Process and Patient Outcome in Diabetes Mellitus." *Medical Care* 17 (7): 748–57.

———. 1980. "Peer Review in Diabetes and Hypertension: The Relationship between Care Process and Patient Outcome." *Southern Medical Journal* 73 (5): 564–8.

Rubin, H. R., P. Pronovost, and G. B. Diette. 2001. "The Advantage and Disadvantages of Process-Based Measures of Health Care Quality." *International Journal for Quality in Health Care* 13 (6): 469–74.

Saaddine, J. B., M. M. Engelgau, G. L. Beckles, E. W. Gregg, T. J. Thompson, and K. M. Narayan. 2002. "A Diabetes Report Card for the United States: Quality of Care in the 1990s." *Annals of Internal Medicine* 136 (8): 565–74.

The Diabetes Control and Complications Trial Research Group. 1993. "The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus." *New England Journal of Medicine* 329 (14): 977–86.

U.K. Prospective Diabetes Study (UKPDS) Group. 1998. "Intensive Blood-Glucose Control with Sulphonylureas or Insulin Compared with Conventional Treatment and Risk of Complications in Patients with Type 2 Diabetes (UKPDS 33)." *Lancet* 352 (9131): 837–53.

Wetzler, H. P., and J. W. Snyder. 2000. "Linking Pharmacy and Laboratory Data to Assess the Appropriateness of Care in Patients with Diabetes." *Diabetes Care* 23 (11): 1637–41.

Williams, T. F., D. A. Martin, M. D. Hogan, J. D. Watkins, and E. V. Ellis. 1967. "The Clinical Picture of Diabetic Control, Studied in Four Settings." *American Journal of Public Health* 57 (3): 441–51.

Zhang, Q., M. Safford, J. Ottenweller, G. Hawley, D. Repke, J. F. Burgess Jr., S. Dhar, H. Cheng, H. Naito, and L. M. Pogach. 2000. "Performance Status of Health Care Facilities Changes with Risk Adjustment of HbA1c." *Diabetes Care* 23 (7): 919–27.

ELSEVIER

Journal of Psychosomatic Research 59 (2005) 307–314

Journal of Psychosomatic Research

Support providers of sexual minority women with breast cancer Who they are and how they impact the breast cancer experience

Ulrike Boehmer[a,b,*], Karen M. Freund[c], Rhonda Linde[d]

[a]*Department of Health Services, School of Public Health, Boston University, Boston, MA, USA*
[b]*Center for Health Quality, Outcomes and Economic Research (CHQOER), Bedford, MA, USA*
[c]*Women's Health Unit, Evans Department of Medicine, Boston University School of Medicine, Boston, MA, USA*
[d]*The Fenway Institute, Fenway Community Health, Boston, MA, USA*

Abstract

Objective: The aim of this study was to identify the factors associated with adjustment to breast cancer among sexual minority women with breast cancer and their support person. **Methods:** In a cross-sectional study, sexual minority women with breast cancer and their support provider were asked to self-report social support, distress, and coping, using standardized measures. **Results:** Twenty-three (77%) women had a support provider participating in the study. Disclosure of sexual orientation, less helpless-hopeless coping, and support provider perception of high fighting spirit were related to lower patient distress. Lower support provider distress was related to more patient disclosure of sexual orientation, a larger social network, and an underestimation of fatalistic patient coping. An overestimation of patients' anxious preoccupation coping was linked to higher support provider distress. **Conclusions:** Providing opportunities to sexual minority patients and their support providers to focus on issues such as disclosure of sexual orientation and coping may lower patient and support provider distress.

Keywords: Adaptation, psychological; Breast neoplasms; Homosexuality, female; Social support; Stress, psychological

Introduction

Spouses are the most frequently reported support person for married heterosexual women with breast cancer [1], as well as the most important source of support [2,3]. Studies that focus on spouses indicate that they provide breast cancer patients with emotional support and physical assistance [4] and that the provision of emotional and instrumental support is associated with better adjustment [5] and perceived as the most helpful forms of support [3,6,7]. Some studies indicate that the distress of spouses equals or even surpasses patients' distress [8–10] and a number of studies noted that patients' spouses lack social support [4,11].

The assertion that breast cancer is a family disease, that primarily affects the couple, has led to a neglect of research of women who are not in traditional marriages. To date, we lack knowledge about sexual minority women with breast cancer, i.e., lesbians, bisexuals, and women who partner with women, who are expected to have a higher incidence of breast cancer than married heterosexual women [12–16].

Different studies focusing on sexual minority women suggest that they perceive less support from their family than do heterosexual women and instead are more likely to receive support from their friends [17–21]. Yet, little is known about the social support of sexual minority women with breast cancer. One study by Fobair et al. [22], which compared lesbian and heterosexual women with breast cancer, determined that lesbians were less frequently partnered, yet were more likely to obtain social support from their partners and friends than heterosexuals do, confirming the findings of previous research in that heterosexual patients received greater social support from their relatives, were closer to their family members, more likely to count on them for help, and were more satisfied with their family members than were lesbian breast cancer patients.

* Corresponding author. CHQOER, 200 Springs Road (152), Bedford, MA 01730, USA. Tel.: +1 781 687 3220; fax: +1 781 687 3106.
E-mail address: boehmer@bu.edu (U. Boehmer).

0022-3999/05/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jpsychores.2005.06.059

Research on sexual minority women and their sources of social support, as well as the cancer literature that considers breast cancer patients' husbands and their support functions, inform the present study. Our research objectives are first to identify the most important support person for sexual minority women with breast cancer. Because not all sexual minority women will be partnered and the support choices of this population are mostly unknown, we decided against a mandatory recruitment of sexual minority women and their partners. Second, we examine the support persons' impact on sexual minority women with breast cancer. In particular, previous studies indicated dyadic influences; that is, the patient and support provider's collective coping and social support affects patient's and support providers' adjustment [23]. It has been shown that the perception that the spouse uses maladaptive coping was related to elevated distress in the individual who rated the spouse as coping poorly [24]. The present study assesses if these influences exist among sexual minority women with breast cancer as well.

Method

Study design

The data used for this paper are derived from a retrospective cross-sectional study of sexual minority women with breast cancer and their support persons. 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 [25,26] and is well suited for a comprehensive exploration of lesbians' 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 participants approval for this study. All recruitment materials announced that participants would receive US$20 for their participation. Recruitment activities consisted of newspaper advertisements, Internet postings, and the distribution of flyers to organizations, at events, and medical centers. Extensive efforts were made to reach African-American and Latina lesbians. We used a separate flyer that explicitly stated that researchers were looking for African-American and Latina participants. This flyer was posted on the Internet, handed to known African-American or Latina lesbians, distributed to churches, organizations, and health centers with a high proportion of African-American and Latina patients, and at African-American or Latino centered events.

Participants

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 non-metastatic breast cancer, (3) having completed active treatment for breast cancer within the last 5 years, and (4) English proficiency. We defined sexual minority status as stating a lesbian or bisexual sexual 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. While we required that women had completed their treatment, we defined active treatment as surgery, radiation, and/or chemotherapy to include women in this study who were currently using hormonal treatment, such as Tamoxifen.

Procedures

After consent was obtained, we conducted one tape-recorded, in-depth, semistructured interview. In addition to this one-time qualitative interview with women with breast cancer and their support providers, we collected information on participants through a self-administered questionnaire. In the course of the qualitative interview, the women with breast cancer were asked 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. Having a significant support person or the support person's willingness to participate was not eligibility criteria for women with breast cancer. Once a support person was identified, we contacted this person and asked for participation in the study. After consent was obtained, we conducted one tape-recorded, in-depth, semistructured interview with each support provider and collected quantitative information through a self-administered questionnaire.

Ten percent of women with breast cancer (n=3) indicated not having a significant support person. Another 10% (n=3) identified a support person, but passively refused the support person's participation by not providing contact information. This included two women who identified their partner, one refused contact information because of her strained relationship with her partner, whereas the second woman reported that her partner had had breast cancer prior to her own breast cancer and that an interview will be too upsetting for her partner. Thus, 80% (n=24) identified a support person and provided contact information, including one woman who identified two support providers, because at the time of her breast cancer diagnosis, she had a strained relationship with her partner and was relying on a friend when her partner was unavailable. After we contacted all 25 identified support providers, all but 1 agreed to participate and were interviewed. The support provider who refused participation explained that she was no longer in contact with the woman with breast cancer.

Measures

Our key dependent variable is emotional distress, which we measured by a brief version of the Profile of Mood

States (POMS). The POMS assesses mood disturbance by asking respondents to indicate to which degree mood-descriptive adjectives (e.g., weary, blue) applied to their emotional state during the past week [27]. The high correlation of the brief 11-item version with the full version of the POMS [28] led us to choose the brief version to reduce respondent burden. The 11-item POMS yields one total mood disturbance score, with higher scores indicating more mood disturbance.

While we assessed mood disturbance in both women with breast cancer and their support providers, coping was assessed differently. Women with breast cancer were asked to report their coping, while the support providers were asked to serve as a proxy and report how they perceived the woman with breast cancer to cope. Coping was measured by the abbreviated 29-item version of the Mental Adjustment to Cancer scale (MAC; [29]). The mini-MAC measures coping responses to a cancer diagnosis, which are then summarized into five subscales that describe the coping styles of (1) fighting spirit, (2) helplessness-hopelessness, (3) anxious preoccupation, (4) fatalism, and (5) cognitive avoidance. Each subscale has acceptable reliability, with alpha coefficients ranging from .62 (fatalism) to .87 (helplessness-hopelessness; [29]). High scores on the scales helplessness-hopelessness, anxious preoccupation, fatalism, and cognitive avoidance can be interpreted as an indication of a maladaptive coping style in response to cancer, whereas a high score on fighting spirit is considered a positive reaction to the cancer diagnosis [29,30].

Independent variables

The questionnaire asked women with breast cancer and their support providers to self-report the following demographics: age at the time of the interview; race categorized into White, African American, Hispanic, Asian, American Indian, mixed race, or other; years of education; health insurance; and current employment and, if answered in the affirmative, hours worked per week. Participants were asked to report their income with respect to the following ranges less than US$20,000, US$20,000 to 29,999, US$30,000 to 39,999, US$40,000 to 49,999, US$50,000 to 59,999, US$60,000 to 69,999, US$70,000 to 79,999, US$80,000 to 89,999, US$90,000 to 99,999, and over US$100,000. In the absence of an agreed upon standardized measure of sexual orientation [31], we followed the recommendations of the Institute of Medicine Report on Lesbian Health, in that we defined our population using two of three (identity, behavior, and desire) dimensions of sexual orientation that we deemed essential for this research project's focus on breast cancer. Women with breast cancer were asked to report the sexual identity or the relationship behavior that best describes them, allowing for the responses lesbian, bisexual, partner with women, or other. In the questionnaire for support providers, the category "heterosexual" was added. Disclosure of sexual orientation to others was assessed by asking participants to report whether they disclosed their sexual orientation (=1), disclosed and openly talked about (=2), or did not disclose (=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. Thus, higher scores indicate a greater level of disclosure.

The illness factors, stage of the disease, time since diagnosis, and time since completion of treatment were collected from women with breast cancer using the questionnaire. Information on women's cancer treatments, such as type of surgery, radiation, chemotherapy, and hormonal treatment, were derived from the qualitative interview data.

In the questionnaires for both participant groups, an abbreviated social support scale based on five items from the Medical Outcome Study (MOS; [32]) was included to assess social support. This measure assesses the frequency with which respondents perceive that different dimensions of social support are available to them: none of the time, a little of the time, some of the time, most of the time, or all of the time. The dimensions of social support are measured by questions such as "How often is someone available to you if you need it, to confide in or talk to about yourself or your problems?" The social support dimensions include emotional and informational support, positive social interaction, tangible support, and affectionate support [32]. We assessed the size of participants' social network, using one item from the MOS that asked, "how many close friends and close relatives do you have?" Current relationship status was categorized into single, being in a committed relationship, or other.

Analysis

Using SAS, we calculated summary statistics to describe participants' illness and demographic characteristics. We used *t* tests, chi square, and Fisher's Exact Test to determine significant differences between women with and those without support providers. Mean differences between women with breast cancer and their support providers in mood disturbance, social support characteristics, and level of disclosure were calculated and their significant differences were assessed using univariate analyses. Women with breast cancer's coping and the proxy rating of their support providers was assessed using the intraclass correlation, a reliability measure of ratings performed by two different raters. We used regression analyses to determine which factors predicted the most variance in emotional distress of support providers and women with breast cancer.

Results

The illness-related information on the women with breast cancer are presented in Table 1. At the time of assessment, more than 1 year had passed since women with breast

Table 1
Medical/Illness characteristics of women with breast cancer

Characteristic	Women without SP (n=7)	Women with SP (n=23)
Stage of breast cancer, % (n)		
In situ	14.3 (1)	26.1 (6)
I	28.6 (2)	34.8 (8)
II	28.6 (2)	30.4 (7)
III	14.3 (1)	0 (0)
Unknown, but report "early stage"	14.3 (1)	8.7 (2)
Mean number of months since diagnosis±S.D. (range)	31.1±5.7 (17–58)	21.4±3.3 (5–52)
Mean number of months since completion of treatment±S.D. (range)	23.3±17.6 (7–56)	14.7±16.0 (1–47)
Treatment, % (n)		
Lumpectomy	42.9 (3)	52.2 (12)
Mastectomy	28.6 (2)	30.4 (7)
Bilateral mastectomy	28.6 (2)	17.4 (4)
Breast reconstruction	14.3 (1)	30.4 (7)
Radiation	71.4 (5)	60.9 (14)
Chemotherapy	57.1 (4)	56.5 (13)
Hormonal treatment (e.g., Tamoxifen)	57.1 (4)	73.9 (17)

cancer had completed their treatment, which consisted for about half of breast conserving surgery followed by adjuvant treatments.

Using bivariate analyses, we examined if there were differences between the 23 women with a participating support provider and the 7 women without such a person. The two groups did not differ on any illness-related characteristics.

In Table 2, we added the demographic information on the 24 participating support providers to the demographic information on women with and women without support providers. There were no significant differences on any of the demographic characteristics between women with and without support providers. However, we did find differences with respect to social support in that women with a support provider had a significantly higher level of perceived support and those with a support provider were more likely to be partnered (Fisher's P=.0596) compared with women without a support provider. When we compared the two groups of women with breast cancer on coping styles and level of distress, there were no significant differences (results not shown).

All support providers were female and 79% (19) were relationship partners, 3 were friends, 1 was a sister, and 1 a mother to the 23 women with breast cancer. Women with breast cancer had the respective relationships with their support provider, on average, for almost 17 years, with a range from 3 to 45 years. The 19 women who chose their partner as their support provider had this partner relationship for an average of more than 14 years, with a range from 3 to 33 years.

For the remainder of this paper, we focus only on the women with breast cancer who had a support provider and their respective support providers, without further distinguishing if the support person is a partner or has another type of relationship to the woman with breast cancer. Due to the fact that 1 of the 23 women identified two support providers, we duplicated the data on this 1 woman, thereby increasing the sample size of women with breast cancer from 23 to 24, to obtain 24 matched pairs of women with breast cancer and support provider.

We calculated the mean difference in the level of disclosure, size of social network, social support, and mood disturbance between women with breast cancer and their support providers. Support providers had a significantly lower level of perceived support, and there was a trend toward a smaller social network compared with women with breast cancer.

Table 2
Demographic and social support characteristics of women with breast cancer and their support providers

Characteristic	Women without support provider (n=7)	Women with support provider (n=23)	Support provider (n=24)
Age, mean±S.D. (range)	53.3±2.9 (45–66)	50.8±1.6 (37–73)	49.5±8.6 (37–70)
Race/Ethnicity, % (n)			
White	85.7 (6)	91.3 (21)	91.7 (22)
Latina	0	4.4 (1)	0
African-American	14.3 (1)	4.4 (1)	4.2 (1)
Asian	0	0	4.2 (1)
Education, % (n)			
High school	14.3 (1)	0 (0)	4.2 (1)
College	57.1 (4)	47.8 (11)	29.2 (7)
Graduate school	28.6 (2)	52.2 (12)	66.7 (16)
Full or part time work, % (n)	100 (7)	73.9 (17)	83.3 (20)
Income range: <US$10,000–>100,000			
Supports one person, mean±S.D. (n)	4.4±1.7 (5)	4.1±3.0 (16)	4.5±3.1 (14)
Supports more, mean±S.D. (n)	5±2.8 (2)	5±3.2 (6)	7.5±2.8 (10)
Missing income		(1)	
Having health insurance, % (n)	100 (7)	96 (22)	95.8 (23)
Sex orientation, % (n)			
Lesbian	85.7 (6)	87.0 (20)	79.2 (19)
Bisexual	0 (0)	4.4 (1)	8.3 (2)
Partner with women	14.3 (1)	8.7 (2)	0
Heterosexual	0	0	12.5 (3)
Disclosure, mean±S.D. (range)	17.3±4.2 (12–24)	18.5±5.5 (6–24)	17.1±5.9 (2.4–24)
Partnered relationship, % (n)	42.9 (3)*	82.6 (19)	100 (24)
Social network, mean±S.D.	11.3±17.3 (2–50)	14.0±12.0 (3–50)	8.5±6.3 (2–30)
Social support, mean±S.D.	58.6±26.4 (20–90)**	85.9±15.6 (45–100)	67.3±22.5 (15–95)

* P<.10.
** P<.05.

The mean difference between patients' self-rating and the support providers' proxy ratings of coping indicated whether support providers over- or underestimated the coping of women with breast cancer. A negative score indicates that the support provider underestimated how poorly a woman is coping with breast cancer, unless the coping strategy is fighting spirit, which is reversed. Support providers' estimation of women's fatalistic coping was significantly lower than the women's self-rating.

In the last column of Table 3, we reported the intraclass correlation of women's self-ratings and support providers' other-ratings of coping. Women with breast cancer and their support providers agreed moderately on the coping strategies of cognitive avoidance, anxiety preoccupation, and fatalism. They had, however, only fair agreement on fighting spirit and even less agreement on helpless-hopeless coping.

The impact of demographic characteristics, such as age and education, on the coping with and adjustment to breast cancer among sexual minority women has been assessed elsewhere [33]. The focus here is on examining women with breast cancer and their support providers' level of disclosure, social support, and self or other coping on adjustment. To understand the predictors of level of distress experienced by women with breast cancer, we regressed women's social support characteristics and their coping ratings on their level of mood disturbance. To understand the impact of women's social support providers, we also regressed the support providers' social support characteristics, the support providers' ratings of women's coping, the difference in social support characteristics between women with breast cancer and support providers, as well as the over- and underestimations of coping on mood disturbance in women with breast cancer. To assess the predictors of the variance in support provider distress, the same factors were regressed on support provider distress (Table 4).

The separate regression analyses revealed that women's helpless-hopeless coping increased their emotional distress. Women with breast cancer were also more distressed if there was discordance in the level of disclosure of sexual orientation between the support provider and the woman with breast cancer, in that the support provider was more open than was the woman with breast cancer. Lower emotional distress in women with breast cancer was related to being open about their sexual orientation and support providers' perception of high fighting spirit, and there was a trend towards lower level of distress in women with breast cancer whose support provider perceived them as having more fatalistic coping. The strongest predictor of support providers' level of distress was disclosure of sexual orientation among women with breast cancer, indicating that women who are more open about their sexual orientation lower support providers' distress level. Support providers' level of distress was lower if they had a larger network of friends and family and if there was a trend that the network of women with breast cancer surpasses that of support providers. Support providers' overestimation of anxiety and preoccupation in the women with breast cancer increased support provider distress, whereas support providers' underestimation of fatalistic coping in women with breast cancer lowered support providers' mood disturbance.

Table 3
Comparison of women with breast cancer and their support providers (n=24)

	WBC	SP	Mean difference	P value	Intraclass correlation
Disclosure	18.3±5.5	17.1±5.9	−0.9±7.2	ns	N/A
Social network	13.8±11.8	8.5±6.3	−5.2±14.6	.0938	N/A
Social support	85.2±15.6	67.3±22.5	−17.9±23.8	.0012	N/A
POMS	8.7±6.8	9.5±8.3	0.8±8.7	ns	N/A
Fatalism	15.7±2.3	14.1±2.4	−1.5±1.7	.0002	.58
Cognitive avoidance	8.8±2.3	8.7±2.2	−0.1±2.1	ns	.54
Anxiety/ preoccupation	18.2±4.6	18.5±4.9	0.3±5.1	ns	.43
Helplessness/ hopelessness	10.1±2.7	11.3±3.8	1.2±4.2	ns	.19
Fighting spirit	12.9±2.4	12.7±2.0	−0.2±2.8	ns	.24

Table 4
Predictors of mood disturbance in women with breast cancer and in support providers

	WBC POMS		SP POMS	
	b	R^2	b	R^2
Women with breast cancer (WBC)				
WBC friends	−0.12	.04	0.20	.08
WBC soc sup	−0.11	.07	−0.15	.08
WBC outness	**−0.66***	**.27**	**−0.87****	**.32**
WBC fatalism	−0.40	.02	1.08	.08
WBC cognitive avoidance	−0.06	.01	−0.01	.00
WBC anxiety/preoccupation	−0.13	.01	−0.23	.02
WBC helplessness/ hopelessness	**1.59****	**.39**	0.36	.01
WBC fighting spirit	−0.52	.03	0.59	.03
Support provider (SP)				
SP friends	−0.22	.04	**−0.16***	**.18**
SP soc sup	−0.05	.03	−0.45	.12
SP outness	0.04	.01	−0.43	.09
SP fatalism	**−1.06*****	**.13**	−0.18	.01
SP cognitive avoidance	−0.36	.01	−1.03	.08
SP anxiety/preoccupation	0.22	.02	**0.68*****	**.16**
SP helplessness/hopelessness	0.14	.01	0.26	.01
SP fighting spirit	**−1.95****	**.32**	−0.21	.01
Difference support provider–women				
SP–WBC friends	0.03	.01	**−0.22*****	**.15**
SP–WBC soc sup	0.01	.01	−0.08	.05
SP–WBC outness	**0.41***	**.18**	0.21	.03
SP–WBC fatalism	−1.38	.12	**−2.06***	**.18**
SP–WBC cognitive avoidance	−0.33	.01	−1.15	.08
SP–WBC anxiety/ preoccupation	0.29	.05	**0.77***	**.23**
SP–WBC helplessness/ hopelessness	−0.53	.11	0.07	.01
SP–WBC fighting spirit	−0.60	.06	−0.54	.03

* P<.10.
** P<.05.
*** P<.01.

Discussion

Sexual minority women, i.e., is lesbian or bisexual-identified women and women who partner with women, have been recognized as an underserved population in the United States [34], for which it has been suggested that certain negative health outcomes, including the cancer burden, are more prevalent compared with heterosexual women [31,35]. Because of their sexual orientation, these women are exposed to society's negative attitudes, stigma, prejudice, and discrimination, including while receiving health care, which constitutes a cultural barrier that interferes with their access to care [31]. How these women fair when diagnosed with breast cancer has hardly been researched, despite evidence that breast cancer incidence is higher in sexual minority women than in heterosexual women [34,36,37]. Our study sought to make a contribution to this area by focusing on the social support that is available to these women. In particular, about three quarters (77%) of the sexual minority women with breast cancer in our sample had a support provider available to them, and almost a quarter that did not have such a person refused to identify the person or whose support provider refused participation.

Comparisons of women with support providers with those without such a person suggest that relationship status may be of importance for the availability of such a person, while no other demographic or illness-related characteristics appeared to explain the availability of a support provider. More importantly, patients with a support provider reported a significantly higher level of perceived support, while the size of the network did not differ between these two groups of women. We interpreted this as an indication that perceived support is directly linked to having a support provider, thereby underlining the importance of such a person for sexual minority women with breast cancer. On the other hand, the availability of a support provider was not linked to improved adjustment or particular coping styles.

Our findings on sexual minority women appear to mirror the conclusions of studies of heterosexual or presumed heterosexual women with breast cancer to the extent that the relationship partner was the most frequently identified type of support provider [1,38,39]. In addition, as in studies of the spouses of heterosexual breast cancer patients, we found a lower level of perceived support among support providers than among women with breast cancer and a trend towards a smaller network compared with women with breast cancer [40–43]. Finally, we also found that patients' and their support providers' distress levels are equal [43–46].

The particular value of our study is, however, its contribution to the understanding of distress in sexual minority women and their support providers. Sexual minority women's level of disclosure is the only aspect that has an about equal impact on the distress of women with breast cancer and their support providers, highlighting the importance of this aspect in this special population. Moreover, discordance in the level of disclosure, in that the support provider is more open about sexual orientation than is the woman with breast cancer, is linked to more emotional distress. This is an important finding that deserves further examination in a future study. Support providers' level of disclosure or the social support resources of women with breast cancer or their support providers are not linked to the distress level of women with breast cancer. Support provider distress is linked to their network size, and in addition, support providers appear to benefit if women with breast cancer have a larger network than their support provider. This leads one to hypothesize that the emotional well-being of women with breast cancer is dependent on the support provider, whereas support providers' emotional well-being depends on their own connections to others. In addition, one can hypothesize that support providers' distress is eased if the women with breast cancer who they are supporting are less isolated.

Support providers' perception of how women cope with breast cancer is foremost associated with the adjustment of women dealing with the disease, whereas support providers' perception of women's coping has a limited relationship to their own distress. On the other hand, support providers' distress appears to be eased if they are supporting women whose coping is less anxious or more fatalistic than judged by the support provider.

The findings of this exploratory study provides some insights that may be useful for the later design of cancer support interventions that target sexual minority women and their support providers, including sexual minority women who are without such a support person. However, before such interventions can be developed, it will be necessary to conduct additional research to overcome some of the limitations of this study. Consistent with the conclusions of previous researchers who targeted lesbian, gay, and bisexual populations, we think that the targeted sampling strategy was most appropriate for reaching a special population inflicted by a rare event, such as breast cancer [47]. However, an inherent limitation of this approach is that only participants who are motivated enough to participate in a study such as ours are recruited. We attempted to counteract this motivational bias by offering a stipend, in an attempt to provide an equal monetary motivation. Furthermore, the small sample size prevents us from making unequivocal conclusions about differences between women with and without support providers. In addition, our support provider sample consisted of only 5 nonpartners and 19 support providers who were partners to the woman with breast cancer. A larger sample size would have allowed for an assessment of differences between women partnered with their support providers and women who rely on support providers who are not partners. We hope that future studies will build on our findings and further expand our exploratory study of social support among sexual minority women. In particular, future studies should collect information on the types of social support provided by different

sources of support as well as collect data on support providers' own coping.

These limitations and suggestions for future research notwithstanding, valuable and new information has been gained about a particular group of breast cancer patients who have previously been neglected. While we concede that our findings need confirmation by additional studies, they make a contribution to improving the current care of sexual minority women with breast cancer by providing new insights into this group's social support conditions and other relevant aspects of their breast cancer experience.

Acknowledgments

We are grateful to the participants who took the time to talk to us and share their thoughts and feelings. 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, PI: U. Boehmer. Institutional support and work space for the research was provided by the Department of Veterans Affairs, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA. 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.

References

[1] Kilpatrick MG, Kristjanson LJ, Tataryn DJ, Fraser VH. Information needs of husbands of women with breast cancer. Oncol Nurs Forum 1998;25(9):1595–601.
[2] Rose JH. Social support and cancer: adult patients' desire for support from family, friends, and health professionals. Am J Community Psychol 1990;18(3):439–64.
[3] Neuling SJ, Winefield HR. Social support and recovery after surgery for breast cancer: frequency and correlates of supportive behaviours by family, friends and surgeon. Soc Sci Med 1988;27(4):385–92.
[4] Northouse LL. Social support in patients' and husbands' adjustment to breast cancer. Nurs Res 1988;37(2):91–5.
[5] Courtens AM, Stevens FC, Crebolder HF, Philipsen H. Longitudinal study on quality of life and social support in cancer patients. Cancer Nurs 1996;19(3):162–9.
[6] Smith E, Redman P, Burns T, Sagert K. Perceptions of social support among patients with recently diagnosed breast, endometrial, and ovarian cancer. J Psychosoc Oncol 1985;3:65–81.
[7] Dunkel-Schetter C. Social support and cancer: findings based on patient interview and their implications. J Soc Issues 1984;40(4): 77–98.
[8] Given B, Given CW. Patient and family caregiver reaction to new and recurrent breast cancer. J Am Med Womens Assoc 1992;47(5): 201–6, 212.
[9] Keitel MA, Zevon MA, Rounds JB, Petrelli NJ, Karakousis C. Spouse adjustment to cancer surgery: distress and coping responses. J Surg Oncol 1990;43(3):148–53.
[10] Matthews BA, Baker F, Spillers RL. Family caregivers and indicators of cancer-related distress. Psychol Health Med 2003;8(1):45–56.
[11] Glasdam S, Jensen AB, Madsen EL, Rose C. Anxiety and depression in cancer patients' spouses. Psychooncology 1996;5(1):23–9.
[12] Dibble SL, Roberts SA, Nussey B. Comparing breast cancer risk between lesbians and their heterosexual sisters. Womens Health Issues 2004;14(2):60–8.
[13] Cochran S, Mays V, Bowen D, Gage S, Bybee D, Roberts S, et al. Cancer-related risk indicators and preventive screening behaviors among lesbians and bisexual women. Am J Public Health 2001;91(4): 591–7.
[14] Kavanaugh-Lynch MHE, White E, Daling JR, Bowen DJ. Correlates of lesbian sexual orientation and the risk of breast cancer. J Gay Lesbian Med Assoc 2002;6(3–4):91–5.
[15] Fish J, Wilkinson S. Understanding lesbians' healthcare behaviour: the case of breast self-examination. Soc Sci Med 2003; 56(2):235–45.
[16] Rankow EJ, Tessaro I. Mammography and risk factors for breast cancer in lesbian and bisexual women. Am J Health Behav 1998;22(6):403–10.
[17] Kurdek LA. Perceived social support in gays and lesbians in cohabitating relationships. J Pers Soc Psychol 1988;54(3):504–9.
[18] Kurdek LA, Schmitt JP. Relationship quality of partners in heterosexual married, heterosexual cohabiting, and gay and lesbian relationships. J Pers Soc Psychol 1986;51(4):711–20.
[19] Kurdek LA, Schmitt JP. Perceived emotional support from family and friends in members of homosexual, married, and heterosexual cohabiting couples. J Homosex 1987;14(3–4):57–68.
[20] Mays VM, Beckman LJ, Oranchak E, Harper B. Perceived social support for help-seeking behaviors of Black heterosexual and homosexually active women alcoholics. Psychol Addict Behav 1994;8(4): 235–42.
[21] Bradford J, Ryan C, Rothblum ED. National lesbian health care survey: implications for mental health care. J Consult Clin Psychol 1994;62(2):228–42.
[22] Fobair P, O'Hanlan K, Koopman C, Classen C, Dimiceli S, Drooker N, et al. Comparison of lesbian and heterosexual women's response to newly diagnosed breast cancer. Psychooncology 2001; 10(1):40–51.
[23] Ben-Zur H, Gilbar O, Lev S. Coping with breast cancer: patient, spouse, and dyad models. Psychosom Med 2001;63(1):32–9.
[24] Ben-Zur H. Your coping strategy and my distress: inter-spouse perceptions of coping and adjustment among breast cancer patients and their spouses. Fam Syst Health 2001;19(1):83–94.
[25] Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. Soc Probl 1989;36(4):416–30.
[26] Biernacki P, Waldorf D. Snowball sampling: problems and techniques of chain referral sampling. Sociol Methods Res 1981;10(2):141–63.
[27] McNair DM, Lorr M, Droppleman LF. POMS manual. San Diego: Educational and Industrial Testing Service, 1971.
[28] Cella DF, Jacobsen PB, Orav EJ, Holland JC, Silberfarb PM, Rafla S. A brief POMS measure of distress for cancer patients. J Chronic Dis 1987;40(10):939–42.
[29] Watson M, Law M, dos-Santos M, Greer S, Baruch J, Bliss JM. The Mini-MAC: further development of the Mental Adjustment to Cancer scale. J Psychosoc Oncol 1994;12(3):33–46.
[30] Watson M, Greer S, Young J, Inayat Q, Burgess C, Robertson B. Development of a questionnaire measure of adjustment to cancer: the MAC scale. Psychol Med 1988;18(1):203–9.
[31] Institute of Medicine Committee on Lesbian Health Research Priorities. Lesbian health. Current assessment and directions for the future. Washington, DC: Institute of Medicine, Committee on Lesbian Health Research Priorities, Neuroscience and Behavioral Health Program. Health Sciences Policy Program, Health Sciences Section, 1999.
[32] Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32(6):705–14.
[33] Boehmer U, Linde R, Freund K. Sexual minority women's coping and psychological adjustment after a diagnosis of breast cancer. J Women's Health 2005;14(3):216–26.

[34] Gay and Lesbian Medical Association, LGBT health experts. Healthy People 2010 companion document for lesbian, gay, bisexual and transgender (LGBT) health. http://wwwglmaorg/policy/hp2010/ 2001(April 12, 2001).
[35] American Cancer Society. Cancer facts and figures 2004. webversion ed. Atlanta: American Cancer Society, 2004.
[36] Valanis BG, Bowen DJ, Bassford T, Whitlock E, Charney P, Carter RA. Sexual orientation and health: comparisons in the women's health initiative sample. Arch Fam Med 2000;9:843–53.
[37] Dibble SL, Vanoni JM, Miaskowski C. Women's attitudes toward breast cancer screening procedures: differences by ethnicity. Womens Health Issues 1997;7(1):47–54.
[38] Mellon S, Northouse LL. Family survivorship and quality of life following a cancer diagnosis. Res Nurs Health 2001;24(6):446–59.
[39] Northouse LL, Mood D, Kershaw T, Schafenacker A, Mellon S, Walker J, et al. Quality of life of women with recurrent breast cancer and their family members. J Clin Oncol 2002;20(19):4050–64.
[40] Northouse LL, Mood D, Templin T, Mellon S, George T. Couples' patterns of adjustment to colon cancer. Soc Sci Med 2000;50(2):271–84.
[41] Kornblith AB, Herr HW, Ofman US, Scher HL, Holland JC. Quality of life of patients with prostate cancer and their spouses. Cancer 1994;73:2791–802.
[42] Northouse LL, Dorris G, Charron-Moore C. Factors affecting couples' adjustment to recurrent breast cancer. Soc Sci Med 1995;41(1):69–76.
[43] Northouse LL, Templin T, Mood D, Oberst M. Couples' adjustment to breast cancer and benign breast disease: a longitudinal analysis. Psychooncology 1998;7(1):37–48.
[44] Baider L, Walach N, Perry S, Kaplan De-Nour A. Cancer in married couples: higher or lower distress? J Psychosom Res 1998;45(3): 239–48.
[45] Compas BE, Worsham NL, Epping-Jordan JE, Grant KE, Mireault G, Howell DC, et al. When mom or dad has cancer: markers of psychological distress in cancer patients, spouses, and children. Health Psychol 1994;13(6):507–15.
[46] Walker BL. Adjustment of husbands and wives to breast cancer. Cancer Pract 1997;5(2):92–8.
[47] Meyer IH. Why lesbian, gay, hisexual, and transgender public health? Am J Public Health 2001;91(6):856–9.

JOURNAL OF WOMEN'S HEALTH
Volume 14, Number 3, 2005

Sexual Minority Women's Coping and Psychological Adjustment after a Diagnosis of Breast Cancer

ULRIKE BOEHMER, Ph.D.,[1,2] RHONDA LINDE, Ph.D.,[3] and KAREN M. FREUND, M.D., M.P.H.[4]

ABSTRACT

Objective: **To determine factors that influence sexual minority women's coping responses and adjustment to breast cancer.**

Methods: **We conducted a cross-sectional study with 64 sexual minority women with breast cancer who were recruited through targeted community-based sampling. In this study, sexual minority women consisted of three sexual orientation groups: those who self-reported partnering with women and those with a lesbian or bisexual identity. We determined the number of years of sexual minority status and disclosure of sexual orientation and used standardized measures to assess these women's coping and adjustment to breast cancer. Data were analyzed using statistical methods as appropriate for the level of data.**

Results: **We determined that sexual minority factors, such as sexual orientation group, influenced coping and adjustment even after illness and social support factors were controlled. In multivariate analyses, women who identified as lesbians or bisexuals used less maladaptive coping compared with women who reported partnering with women. The association between reporting a lesbian identity and lower distress approached significance in multivariate regression equations.**

Conclusions: **Of the sexual minority factors that were considered, sexual orientation group, number of years of sexual minority status, and disclosure of sexual minority status, only sexual orientation group was related to coping and lower distress. Contrary to expectations, disclosure of sexual orientation did not relate to coping and lower distress. The findings support the need for future studies to include different aspects of sexual minority status, in particular, clearly defined sexual orientation groups.**

INTRODUCTION

MEDIA AND RESEARCH ATTENTION has focused on the likelihood of lesbians' increased risk factors for breast cancer.[1–5] However, there is little research addressing how lesbians respond and adjust when diagnosed with breast cancer. One potential method of inquiry is to compare the coping of lesbians with that of heterosexual women to look for health disparities based on sex-

[1]Department of Health Services, Boston University School of Public Health, Boston Massachusetts.
[2]Center for Health Quality, Outcomes and Economic Research (CHQOER), Bedford, Massachusetts.
[3]The Fenway Institute, Fenway Community Health, Boston, Massachusetts.
[4]Women's Health Unit, Evans Department of Medicine, Boston University School of Medicine, Boston Massachusetts.

ual minority status. A study by Fobair et al.[6] that used this methodology reported no differences in mood disturbance; however, in coping styles, lesbians had significantly less fatalism and less cognitive avoidance than heterosexuals.[6] Yet Fobair et al. also found that lesbians experienced less fighting spirit compared with heterosexual women.[6] In order to understand the influences on sexual minority women's coping and adjustment to breast cancer, we chose a methodology of comparing coping and adjustment within a group of sexual minority women. In this study, we focused on examining the variability in predictors of coping and adjustment within sexual minority women with breast cancer.

The existing literature, generated from studying presumed heterosexual populations, states that coping and adjustment to breast cancer are influenced by three broad categories of factors: illness, person, and social support factors. Coping strategies are related to the illness phase and the responsiveness of others in a person's environment,[7] thereby giving social support a central role in improving coping with cancer.[8] The importance of social support has also been confirmed for adjustment, in that some studies indicate single women are more vulnerable to poor adjustment[9,10] and propose to recognize singles as a vulnerable subpopulation among breast cancer patients.[11,12] Coping styles differ by patients' age, in that older patients use more frequently optimistic and palliative strategies and rate them as more effective.[13]

Active coping strategies, such as fighting spirit, problem solving, seeking social support, and focusing on the positive, are linked to improved psychosocial adjustment.[14–18] This knowledge about associations of coping and adjustment with certain illness, person, and social support factors provides the foundation for the development of psychosocial programs and interventions to impact women's coping and adjustment to breast cancer.

Few studies have examined whether the same coping styles and other social factors in lesbians are associated with improved adjustment.[6,19] Prior research studies of noncancer populations of gay men and lesbians linked disclosure of sexual minority status to better mental health and adjustment.[20–22] Morris et al.'s survey[23] of a national sample of lesbian and bisexual women suggested an empirically tested model of well-being that stated lesbians and those who had a sexual minority identity for more years were more open about their sexual minority status, and this openness predicted better psychological adjustment. The study by McGregor et al.[24] contradicted one component of this model in that disclosure did not relate to lower distress in lesbians with breast cancer. Based on these contradictory findings, we wished to examine the suggested model in our sample of lesbians with breast cancer and were particularly interested in examining the relationship of disclosure to improved adjustment.

MATERIALS AND METHODS

Study design

The data for this paper were derived from two cross-sectional studies of lesbians with breast cancer. Both studies relied on targeted community sampling, which was further enhanced through the use of snowball sampling where participants were asked to refer others who might be willing to participate in the study. This sampling strategy has been widely used to overcome the challenges of recruiting members of vulnerable or hidden populations into research studies[25,26] and has been noted as essential for targeted research on rare populations.[27]

We obtained institutional human subjects approval for both studies, and all recruitment materials announced that participants would receive $20 for their participation. Recruitment activities consisted of newspaper advertisements, Internet postings, and the distribution of fliers to organizations and at events and medical centers. Extensive efforts were made to reach African American and Latina lesbians, and we used a separate flier that explicitly stated researchers were looking for African American and Latina participants. This flier was posted on the Internet, handed to known African American or Latina lesbians, and distributed to churches, organizations, and health centers with a high proportion of African American and Latino patients and at African American or Latino-centered events.

Participants

Women were eligible for this study if they met three eligibility criteria: (1) sexual minority status, (2) a diagnosis of breast cancer, and (3) fluency in English. We defined sexual minority sta-

tus as stating a lesbian or bisexual identity or partnering with women, as previous research on sexual orientation had shown that older, African American, or Hispanic women were more likely to report same-sex partner choice than a lesbian or bisexual identity.[28]

Procedures

After consent was obtained, we conducted tape-recorded in-depth, semistructured interviews. At the end of each one-time qualitative interview, we asked participants to complete a self-administered questionnaire from which we derived almost all data for this report.

Measures

Our key dependent variable is emotional distress, which we measured by a brief version of the Profile of Mood States (POMS). The POMS assesses mood disturbance by asking respondents to indicate to which degree mood-descriptive adjectives (e.g., weary, blue) applied to their emotional state during the past week.[29] The high correlation of the brief 11-item version with the full version of the POMS[30] led us to choose the brief version to reduce respondent burden. The 11-item POMS yields one total mood disturbance score, with higher scores indicating more mood disturbance.

Coping was measured by the abbreviated 29-item version of the Mental Adjustment to Cancer (MAC) scale.[31] The Mini-MAC measures coping responses to a cancer diagnosis, which are then summarized into five subscales. The fighting spirit subscale includes such items as "I see my illness as a challenge." The helplessness-hopelessness subscale includes such items as "I feel like giving up." The anxious preoccupation subscale includes such items as "I am a little frightened." The fatalism subscale includes such items as "I count my blessings," and the fifth and last subscale, cognitive avoidance, includes such items as "Not thinking about it helps me cope." Each subscale has acceptable reliability with alpha coefficients ranging from 0.62 to 0.87.[31] High scores on the helplessness-hopelessness, anxious preoccupation, fatalism, and cognitive avoidance scales can be interpreted as an indication of a maladaptive coping style in response to cancer, whereas a high score on fighting spirit is considered a positive reaction to the cancer diagnosis.[31,32] As previous researchers suggested, we examined the correlation between the fighting spirit and helplessness-hopelessness scales to determine if they ought to be combined into a single scale.[33,34] Because the correlation between the two scales approached statistical significance in our sample ($r = -0.22$; $p = 0.08$), we chose to combine both scales into a modified fighting scale. A high score on this modified fighting scale indicates positive coping.

Independent variables

The questionnaire asked women to self-report four sexual minority factors. In the absence of an agreed on standardized measure of sexual orientation,[35] we followed the recommendations of the Institute of Medicine (IOM) Report on Lesbian Health, in that we defined our population using two of three dimensions (identity and behavior but not desire) of sexual orientation, which we deemed essential for this research project's focus on breast cancer. Women were asked to report the sexual identity or the relationship behavior that best described them, allowing for the responses: lesbian, bisexual, partner with women, or other. Number of years women had their particular sexual orientation was calculated by subtracting women's responses to the question "How old were you when you first identified this way?" from their current age. 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 groups of people: children, grandchildren, nieces and nephews, parents, grandparents, siblings, other blood relatives, heterosexual friends, colleagues at work, one's boss, neighbors, or heterosexuals in general. Higher scores indicate a greater level of disclosure. Women's disclosure of their sexual orientation to any of their treating providers (yes vs. no) was assessed using the information provided in the qualitative interview.[36]

The illness factors, stage of the disease, and time since diagnosis were collected via questionnaire, and the extent of women's cancer treatment was derived from the qualitative interview data.

Person factors were assessed with the questionnaire. Respondents were asked about their age at the time of the interview; race categorized into white, African American, Hispanic, Asian, American Indian, mixed race, or other; years of education; health insurance; current employment

and if answered in the affirmative, hours worked per week. Income was determined by providing respondents with the following ranges: <$20,000, in $10,000 increments to $99,999, and >$100,000.

Social support was assessed using an abbreviated social support scale from the Medical Outcome Study (MOS).[37] This short version of the social support scale consists of five items and assesses the frequency with which emotional, tangible, affection, and positive interaction support are available: none of the time, a little of the time, some of the time, most of the time, or all of the time. The individual items are (1) someone available to confide in or talk to about yourself or your problems (emotional/informational support), (2) someone to get together with for relaxation (positive social interaction), (3) someone to help with daily chores if you are sick (tangible support), (4) someone to turn to for suggestions about how to deal with a personal problem (emotional/informational support), and (5) someone to love and make you feel wanted (affectionate support).[37] The responses to the five items were summarized into one scale that has a 0–100 range, with higher scores indicating more support. We assessed the size of women's social network, using one item from the MOS that asked "How many close friends and close relatives do you have?" Current relationship status was categorized into single, being in a committed relationship with a woman, or other. Current or former membership in a cancer-related support group was derived from the qualitative interview data.

Analysis

We performed descriptive statistics, such as means and frequencies, to show the distribution of each of the study variables. We used *t* tests, Fisher's exact test, Pearson correlations, and analysis of variance (ANOVA) with Duncan multiple comparison tests to examine the bivariate associations between study variables: sexual minority, person, illness, and social support factors, coping, and mood disturbance. Based on the bivariate analysis results, we selected variables that were significantly ($p < 0.05$) associated with coping and mood disturbance to be included in a multiple regression model to determine the independent association of these variables on each of the coping styles and on mood disturbance separately. For some analyses, we recoded sexual orientation into dichotomous variables. One recoded variable compared lesbian-identified women with the combined group of bisexual-identified women and women who partner with women. The second variable compared lesbian-identified or bisexual-identified women with women who partner with women. Previous studies suggested that anxious preoccupation coping is conceptually similar to our outcome of emotional distress measured by the POMS,[33,34] which we confirmed in that anxious preoccupation coping and the POMS were significantly correlated ($r = 0.36$; $p < 0.01$). Based on this, we related this coping style to the independent measures yet did not consider it for inclusion into a model of mood disturbance that included coping style as a predictor. The statistical package SAS was used for all analyses.

RESULTS

Descriptive statistics on sexual minority status and person factors in this study sample are shown in Table 1. On average, women were 50 years old, and the majority were white, had a high education level, were currently employed, and had health insurance. One quarter of the sample reported an income in the $60,000–69,999 range and supported more than one person. Sexual orientation was foremost reported as a sexual identity in that 86% had a lesbian identity and 9% a bisexual identity. Five percent reported their sexual orientation as behavior in that they partnered with women. On average, women had had this sexual orientation for 27 years. Seventy-eight percent of the participants disclosed their sexual orientation to breast cancer care providers, and women's average level of disclosure to others in their social environment was 18.1 on a scale that ranged from 0 to 24.

The descriptive statistics on women's social support and illness characteristics are shown in Table 2. Seventy-two percent of women were in a committed relationship. On average, women reported having about 15 close friends or relatives, and the mean level of perceived support was 81.8 on a scale that ranged from 0 to 100. About half of the women reported membership in a cancer-related support group. The majority of women reported early breast cancer, stage 0, I, or II, and women received their respective diagnoses an average of 4 years ago. More than half of the sample received breast-conserving surgery. Sixty-

TABLE 1. DEMOGRAPHICS OF STUDY PARTICIPANTS WITH BREAST CANCER (n = 64)

Characteristic	*Data*
Mean age, ± SD (range)	50.5 ± 8.1 (26–73)
Race/ethnicity, % (n)	
White	92 (59)
Latina	5 (3)
African American	3 (2)
Education, % (n)	
High school	3 (2)
College	48 (31)
Graduate school	48 (31)
Employment, % (n)	
No	16 (10)
Part-tiime	31 (20)
Full-time	53 (34)
Income range: <$20K–>$100K	
Supporting 1 person, % (n) mean ± SED	75 (47) 4.0 ± 2.5
Supporting more than 1 person, % (n) mean ± SD	25 (16) 6.1 ± 3.0
Missing income	(1)
Having health insurance, % (n)	97 (62)
Sexual orientation, % (n)	
Lesbian	86 (55)
Bisexual	9 (6)
Partner with women	5 (3)
Mean number of years sexual orientation, ± SD (range)	26.6 ± 9.9 (8–56)
Disclosure of sexual orientation to providers, % (n)	
Yes	78 (50)
No	22 (14)
Mean disclosure to family, work, others, ± SD (range)	18.1 ± 4.9 (6–24)

seven percent of women received radiation therapy, 58% chemotherapy, and 53% hormone therapy, such as tamoxifen. Of the women treated surgically with a single or bilateral mastectomy, 38% chose breast reconstruction.

Table 3 reports women's coping responses to breast cancer and their psychological adjustment. In the fourth column of Table 3, we show the scale developers' suggested cutoff points. Generally, scores above this point indicate maladaptive coping, whereas for fighting spirit, maladaptive coping is below the cutoff point. The proportion of our sample in this negative range is reported in the last column. Seventeen percent of fatalism, 3% of cognitive avoidance, 9% of anxious preoccupation, and 9% of fighting spirit coping was in the negative range, indicating coping below levels established in other cancer populations. Five percent of the study population experienced distress higher than 2 SD from the mean.

As suggested by Morris et al.,[23] we related number of years with sexual minority status and sexual orientation groups to both of our disclosure measures. Number of years with sexual minority status was significantly related neither to disclosure to providers nor to disclosure to others. When comparing lesbian-identified women with all others, we found that lesbian-identified women were significantly more likely to disclose to their breast cancer providers than bisexual-identified and women who partner with women combined (odds ratio = 6.39, CI = 1.43-28.5, $p < 0.05$), yet disclosure to others was not significantly related to sexual orientation group.

Additional bivariate analyses focused on relating sexual orientation to other independent measures. We related the three groups, lesbian-identified, bisexual-identified, and partnering with women, to age, race, educational level, employment status, having insurance, and income. There were no significant differences in these demographics among the three sexual orientation groups. Because other researchers had focused on lesbian-identity in particular, we also compared lesbian-identified women with the other two groups combined and confirmed our earlier findings of no significant association. When we compared the three sexual orientation groups on treatments (surgery, radiation, chemotherapy, hormone treatment, and breast reconstruction), we found no significant differences.

We also related sexual orientation, disclosure,

TABLE 2. SOCIAL SUPPORT AND MEDICAL/ILLNESS CHARACTERISTICS OF STUDY PARTICIPANTS WITH BREAST CANCER (n = 64)

Characteristic	*Data*
Relationship status, % (n)	
Single	28 (18)
Partnered	72 (46)
Mean MOS social support, ± SD (range)	81.8 ± 18.4 (20–100)
Mean number friends, ± SD (range)	15.3 ± 14.4 (2–65)
Attended support group, % (n)	
Yes	52 (33)
No	48 (31)
Stage of breast cancer, % (n)	
In situ	23 (15)
I	34 (22)
II	25 (16)
III	5 (3)
IV	6 (4)
Unknown	6 (4)
Mean time since diagnosis (months), ± SD (range)	49.4 ± 47.4 (5–215)
Surgical treatment, % (n)[a]	97 (62)
Lumpectomy	53 (34)
Mastectomy (including bilateral)	45 (29)
Adjuvant treatment, % (n)	
Radiation	67 (43)
Chemotherapy	58 (37)
Hormone treatment (e.g., tamoxifen)	53 (34)
Breast reconstruction, % (n)	
Yes	38 (11)
No	62 (18)

[a]One participant did not receive any surgical treatment.

and years of sexual minority status to social support. The sexual orientation groups did not differ on relationship status, social support, number of friends, and support group membership. Number of years of sexual minority status was unrelated to all four social support measures. Women who were more open about their sexual minority status ($r = 0.37, p < 0.01$) and those who disclosed to breast cancer providers perceived greater levels of social support (mean = 84.7 vs. mean = 71.4, $p < 0.05$). Disclosure to others and to providers was unrelated to number of friends. Disclosing to providers related to being in a committed relationship (Fisher's test, $p = 0.0151$), whereas disclosing to others was unrelated to relationship status.

The next set of bivariate analyses related all independent measures, that is, sexual minority, person, social support, and illness factors, to the two sets of outcome variables, coping and distress. The results are reported in Table 4.

There were no significant associations between any of the demographic factors and any of the different coping styles or distress. Only one of the sexual minority status measures, sexual orientation, was related to coping and distress. Women who reported a lesbian or bisexual identity used significantly less cognitive avoidance coping.

TABLE 3. COPING STRATEGIES AND PSYCHOLOGICAL ADJUSTMENT

	Mean ± SD	*Range*	*Cutoff*	*Negative range (%)*
Coping (Mini-MAC)				
Fatalism	14.8 ± 2.7	9–20	>17	17.2
Cognitive avoidance	8.4 ± 2.3	4–15	>12	3.1
Anxious preoccupation	18.7 ± 4.6	9–32	>24	9.4
Fighting spirit	30.6 ± 4.3	18–37	<25	9.4
POMS	8.9 ± 7.7	0–28	Mean ± 2 SD	4.7

TABLE 4. RELATIONSHIP OF SEXUAL MINORITY, PERSON, SOCIAL SUPPORT, AND ILLNESS FACTORS TO COPING AND ADJUSTMENT

	Fighting spirit	*Fatalism*	*Anxious preoccupation*	*Cognitive avoidance*	*POMS*
Sexual minority					
Lesbian				M^a = 8.2A[b]	M = 8.2A
Bisexual				M = 7.5A	M = 10.3AB
Partnered with women				M = 12.7B	M = 18.7B
Years of sexual orientation					
Disclosure providers					
Disclosure others					
Person					
Age					
Race					
Education					
Employment					
Income					
Insurance					
Support					
Partnered					
Social support	r = 0.32***				
Friends	r = 0.25**				
Support group					
Yes				M = 9.1***	M = 10.8**
No				M = 7.6	M = 6.8
Illness					
BC stage					
0					M = 11.0AB
1					M = 5.5A
2					M = 9.5AB
3					M = 12.3AB
4					M = 16.0B
Unknown					M = 7.8AB
Months since diagnosis					
Lumpectomy		M = 14.1**			M = 6.9**
Mastectomy		M = 15.5			M = 11.6
Radiation					
Chemotherapy					
Yes		M = 15.4**			
No		M = 14.0			
Hormone therapy					
Reconstruction					

[a]*M*, mean, *r*, correlation.
[b]Groups with different letters (A,B) indicate they are significantly ($p < 0.05$) different from each other.
**$p < 0.05$; ***$p < 0.01$.

Lesbian-identified women's level of distress was significantly less than distress levels in the two other sexual orientation groups. Number of years with sexual minority status and disclosure was unrelated to coping and distress. Being in a committed relationship was unrelated to coping or distress. However, women's perceived level of social support and number of friends were both positively related to more fighting spirit. Women in support groups reported significantly more cognitive avoidance coping and more distress. The illness factors, time since diagnosis, radiation therapy, hormone therapy, and breast reconstruction were unrelated to coping and level of distress. Women who received breast-conserving surgery used significantly less fatalism coping and had lower distress. Chemotherapy was related to more fatalism coping. Women's distress was related to the stage of their breast cancer; women with stage I were significantly less distressed than women with metastatic breast cancer. None of the measures we considered (sexual minority, person, illness, and social support factors) were linked to anxious preoccupation coping.

TABLE 5. UNSTANDARDIZED REGRESSION COEFFICIENTS OF COPING STYLES AND MOOD DISTURBANCE

	Dependent variables				
				POMS[b]	
Independent variables	*Fighting spirit*	*Fatalism*	*Cognitive avoidance*[a]	*Model 1*	*Model 2*
Sexual orientation			−3.90***	−4.75*	−3.68
Social support	0.08***				
Friends	0.05				
Support group			1.21**	3.98**	3.41*
Cancer stage				0.10	−0.09
Surgery		1.17*		3.67*	4.31**
Lumpectomy = 0					
Mastectomy = 1					
Chemotherapy		1.08*			
Fighting spirit					−0.46**
R^2	0.16	0.11	0.23	0.20	0.26

[a]Sexual orientation was recoded to compare lesbian and bisexual identity with women who partner with women.
[b]Sexual orientation was recoded to compare lesbian-identified women with the combination of women who partner with women and bisexual-identified women.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

We built on the bivariate analyses by introducing all significant associations simultaneously into multiple regression equations for each coping style and for mood disturbance (Table 5). Sixteen percent of the variance in fighting spirit was explained by perceived social support and the number of friends in the social network, but number of friends was no longer significant. It is of note that the correlation between social support and number of friends was $r = 0.23$, $p = 0.07$, indicating that these two measures of social support captured different dimensions of social support. In the model of fatalism coping, 11% of the variance was explained by type of surgery and chemotherapy together, but introducing both variables together diminished both variables' significance level. Twenty-three percent of the variance in cognitive avoidance coping was explained by having attended a support group and sexual orientation, indicating that women with a lesbian or bisexual identity used less cognitive avoidance coping. In the multivariate equation of mood disturbance (Table 5, Model 1), 20% of the variance was explained by sexual orientation, support group, surgery, and cancer stage, yet cancer stage was no longer significant, and the effect of sexual orientation was reduced to a trend, indicating lesbian identity was related to less distress.

Our remaining analyses focused on mood disturbance only. We correlated fighting spirit, fatalism, and cognitive avoidance with mood disturbance to determine the relationship of coping styles and adjustment. Fighting spirit was significantly correlated ($r = 0.27$, $p < 0.05$), whereas fatalism and cognitive avoidance were not significantly correlated with mood disturbance. Adding fighting spirit into the multivariate model of mood disturbance increased the amount of explained variance by 6%. In this model, surgery remained a significant independent measure, and fighting spirit was significantly associated with less distress, whereas lesbian identity became nonsignificant.

DISCUSSION

The purpose of this study was to assess coping and adjustment in a sample of sexual minority women. Some of our findings were consistent with the literature on presumed heterosexual samples of breast cancer patients. Social support's positive relationship to fighting spirit, was a direct confirmation of prior studies,[38] and the reverse association between fighting spirit and distress, even after other factors were controlled, confirmed prior research.[18]

Because we deemed identity and partner behavior important concepts in a study of adaptation and coping in response to breast cancer, we included women who identified as lesbians or bisexuals and women who reported partnering with women. Although we had only few women

who reported a sexual orientation other than lesbian identity, differences in sexual orientation were linked to cognitive avoidance coping and approached significance with distress. These findings highlight the importance of using different measures of sexual orientation in future studies to continue the evaluation of salient differences among these different groups.

Our data provided weak support for Morris et al.'s model[23] of predictors and outcomes of disclosure among lesbian and bisexual women, in that in our sample, lesbian-identified women were more likely to disclose to providers. However, disclosure did not have the anticipated positive outcomes suggested by Morris et al. Instead, we confirmed the earlier finding by McGregor et al.[24] of disclosure's lack of relationship to lower distress in lesbians with breast cancer. Also of note is our finding that disclosure was unrelated to coping.

Other important contributions of our study are our findings of relationships between disclosure and social support. Our study provided evidence that women in relationships more frequently disclosed to providers, most likely indicating the need to disclose to providers if women with breast cancer wished to involve their partner in their care. Finally, the association of greater perception of social support among women who disclosed to providers was a hopeful sign, which might mean that women who disclosed to their physicians received their support. This interpretation is consistent with the literature that linked disclosure to providers with satisfaction with care and greater use of routine care.[39–41]

Our study had several noteworthy negative results. None of the demographic factors, including younger age, that studies often linked to poorer adjustment[42–44] were related to coping or adjustment in our study. In previous studies, there is some recognition that unmarried women are fairing poorer or are at least at risk for poor adjustment.[45] In our study, however, being in a committed relationship, the equivalent of being married except for the absence of legal same-sex marriage at the time of the study, had no positive influence on coping or adjustment. Fatalism and cognitive avoidance coping, linked by prior studies to distress, were unrelated to the distress level in our sample.[18] Overall, the sexual minority women in this sample experienced low levels of emotional distress, lower than the mean (10.18) mood disturbance score of the scale construction sample.[30]

Previous research found little or inconsistent impact of illness factors, such as time since diagnosis or type of medical treatment, on coping and adjustment.[46–49] Our data revealed an association of chemotherapy and mastectomy with fatalism and an association of mastectomy with distress. In our data, time since diagnosis was consistently unrelated to any of our coping and mood measures. The lack of influence of time raises questions if coping and distress responses in this population are stable and linked to the diagnosis.

Findings that membership in a cancer-related support group related to higher distress and more cognitive avoidance coping can be understood to indicate emotionally distressed women and those who use more cognitive avoidance coping seek out support groups more frequently. We hope studies will include support group membership as a variable in the future and evaluate its influence more thoroughly.

A number of limitations have to be considered when interpreting the results of this study. Our cross-sectional design allowed us to determine relationships without, however, understanding the causality of these relationships or change over time. We neglected to collect information about long-term side effects, which might also affect coping and adjustment to the disease. Our findings were further limited by the use of a convenience sample, thereby preventing us from generalizing to the population of lesbians with breast cancer or being confident about the replication of our findings in a larger study. Although this is a valid criticism of our methodology, in other areas of lesbian, gay, bisexual, and transgender health research, the findings of smaller studies that used this sampling strategy were later replicated by methodologically sound population-based research studies,[27] thereby underlining the importance of targeted studies, such as ours.

CONCLUSIONS

This study indicated that in sexual minority women with breast cancer, factors directly related to this sexual minority status influenced coping and adjustment, even after illness and social support factors were controlled. Lesbian or bisexual identity related to less cognitive avoidance coping, and lesbian identity related to lower distress. Contrary to expectations, disclosure of sexual orientation did not relate to coping and lower distress.

ACKNOWLEDGMENTS

We are grateful to the participants who took the time to talk to us and share their thoughts and feelings. Support for this research was provided by the Susan G. Komen Breast Cancer Foundation grant # POP0100158; the Massachusetts Department of Public Health Cancer Research Program, FY02; and the American Cancer Society, ROG-03-105-01, PI: U. Boehmer. Institutional support and work space for the research was provided by the Department of Veterans Affairs, Veterans Health Administration, and Health Services Research and Development Service. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Susan G. Komen Breast Cancer Foundation, the Massachusetts Department of Public Health, the American Cancer Society, or the Department of Veterans Affairs.

REFERENCES

1. Dibble SL, Roberts SA, Nussey B. Comparing breast cancer risk between lesbians and their heterosexual sisters. Wom Health Issues 2004;14:60.
2. Cochran S, Mays V, Bowen D, et al. Cancer-related risk indicators and preventive screening behaviors among lesbians and bisexual women. Am J Public Health 2001;91:591.
3. Kavanaugh-Lynch MHE, White E, Daling JR, Bowen DJ. Correlates of lesbian sexual orientation and the risk of breast cancer. J Gay Lesb Med Assoc 2002;6:91.
4. Fish J, Wilkinson S. Understanding lesbians' healthcare behaviour: The case of breast self-examination. Soc Sci Med 2003;56:235.
5. Rankow EJ, Tessaro I. Mammography and risk factors for breast cancer in lesbian and bisexual women. Am J Health Behav 1998;22:403.
6. Fobair P, O'Hanlan K, Koopman C, et al. Comparison of lesbian and heterosexual women's response to newly diagnosed breast cancer. Psychooncology 2001;10:40.
7. Heim E, Augustiny KF, Schaffner L, Valach L. Coping with breast cancer over time and situation. J Psychosom Res 1993;37:523.
8. Holland JC, Gooen-Piels J. Principles of psycho-oncology. In: Bast RC, Jr, Kufe DW, Pollock RE, et al., eds. Cancer medicine, 5th ed. Hamilton, Ontario: B.C. Decker Inc, 2000:943.
9. Holmberg SK, Scott LL, Alexy W, Fife BL. Relationship issues of women with breast cancer. Cancer Nurs 2001;24:53.
10. Broeckel JA, Jacobsen PB, Balducci L, Horton J, Lyman GH. Quality of life after adjuvant chemotherapy for breast cancer. Breast Cancer Res Treat 2000;62:141.
11. Gluhoski VL, Siegel K, Gorey E. Unique stressors experienced by unmarried women with breast cancer. J Psychosoc Oncol 1997;15:173.
12. Weisman AD, Worden JW. The existential plight in cancer: Significance of the first 100 days. Int J Psychiatry Med 1976;7:1.
13. Halstead MT, Fernsler JI. Coping strategies of long-term cancer survivors. Cancer Nurs 1994;17:94.
14. Livneh H. Psychosocial adaptation to cancer: The role of coping strategies. J Rehabil 2000;66:40.
15. Shapiro DE, Boggs SR, Rodrigue JR, et al. Stage II breast cancer: Differences between four coping patterns in side effects during adjuvant chemotherapy. J Psychosom Res 1997;43:143.
16. Stanton AL, Danoff-Burg S, Cameron CL, et al. Emotionally expressive coping predicts psychological and physical adjustment to breast cancer. J Consult Clin Psychol 2000;68:875.
17. Watson M, Greer S, Rowden L, et al. Relationships between emotional control, adjustment to cancer and depression and anxiety in breast cancer patients. Psychol Med 1991;21:51.
18. Akech T, Okuyama T, Imoto S, Yamawaki S, Uchitomi Y. Biomedical and psychosocial determinants of psychiatric morbidity among postoperative ambulatory breast cancer patients. Breast Cancer Res Treat 2001;65:195.
19. Fobair P, Koopman C, Dimiceli S, et al. Psychosocial intervention for lesbians with primary breast cancer. Psychooncology 2002;11:427.
20. Jordan KM, Deluty RH. Coming out for lesbian women: Its relation to anxiety, positive affectivity, self-esteem, and social support. J Homosex 1998;35:41.
21. Griffith KH, Hebl MR. The disclosure dilemma for gay men and lesbians: "Coming out" at work. J Appl Psychol 2002;87:1191.
22. D'Augelli AR, Grossman AH, Hershberger SL, O'Connell TS. Aspects of mental health among older lesbian, gay, and bisexual adults. Aging Ment Health 2001;5:149.
23. Morris JF, Waldo CR, Rothblum ED. A model of predictors and outcomes of outness among lesbian and bisexual women. Am J Orthopsychiatry 2001;71:61.
24. McGregor BA, Carver CS, Antoni MH, et al. Distress and internalized homophobia among lesbian women treated for early stage breast cancer. Psychol Women Q 2001;25:1.
25. Watters JK, Biernacki P. Targeted sampling: Options for the study of hidden populations. Soc Probl 1989; 36:416.
26. Biernacki P, Waldorf D. Snowball sampling: Problems and techniques of chain referral sampling. Sociol Methods Res 1981;10:141.
27. Meyer IH. Why lesbian, gay, bisexual, and transgender public health? Am J Public Health 2001;91:856.
28. Laumann EO, Gagnon JH, Michael RT, Michaels S. The social organization of sexuality. Sexual practices in the United States. Chicago: University of Chicago Press, 1994.
29. McNair DM, Lorr M, Droppleman LF. POMS manual. San Diego: Educational and Industrial Testing Service, 1971.
30. Cella DF, Jacobsen PB, Orav EJ, et al. A brief POMS measure of distress for cancer patients. J Chronic Dis 1987;40:939.

31. Watson M, Law M, dos-Santos M, et al. The Mini-MAC: Further development of the Mental Adjustment to Cancer scale. J Psychosoc Oncol 1994;12:33.
32. Watson M, Greer S, Young J, et al. Development of a questionnaire measure of adjustment to cancer: The MAC scale. Psychol Med 1988;18:203.
33. Classen C, Koopman C, Angell K, Spiegel D. Coping styles associated with psychological adjustment to advanced breast cancer. Health Psychol 1996;15:434.
34. Cordova MJ, Giese-Davis J, Golant M, et al. Mood disturbance in community cancer support groups. The role of emotional suppression and fighting spirit. J Psychosom Res 2003;55:461.
35. Institute of Medicine Committee on Lesbian Health Research Priorities, Lesbian Health. Current assessment and directions for the future. Neuroscience and Behavioral Health Program. Health Sciences Policy Program, Health Sciences Section, 1999.
36. Boehmer U, Case P. Physicians don't ask, some patients tell: Disclosure of sexual orientation among women with breast cancer. Cancer 2004;101:1882.
37. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32:705.
38. Spiegel D. Psychosocial aspects of breast cancer treatment. Semin Oncol 1997;24:S1.
39. Bradford J, Ryan C, Rothblum ED. National Lesbian Health Care Survey: Implications for mental health care. J Consult Clin Psychol 1994;62:228.
40. Barbara AM, Quandt SA, Anderson RT. Experiences of lesbians in the health care environment. Women Health 2001;34:45.
41. White JC, Dull VT. Health risk factors and health-seeking behavior in lesbians. J Wom Health 1997;6: 103.
42. Cimprich B, Ronis DL, Martinez-Ramos G. Age at diagnosis and quality of life in breast cancer survivors. Cancer Pract 2002;10:85.
43. Bloom JR, Kessler L. Risk and timing of counseling and support interventions for younger women with breast cancer. J Natl Cancer Inst Monogr 1994;16:199.
44. Ganz PA, Hirji K, Sim MS, et al. Predicting psychosocial risk in patients with breast cancer. Med Care 1993;31:419.
45. Ganz PA, Greendale GA, Petersen L, Kahn B, Bower JE. Breast cancer in younger women: Reproductive and late health effects of treatment. J Clin Oncol 2003;21:4184.
46. Remie M, Garssen B. Non-expression of emotions in cancer patients. In: Vingerhoets A, van Bussel F, Boelhouwer J, eds. The (non)expression of emotions in health and disease. Tilburg: Tilburg University Press, 1997:237.
47. Hoskins CN, Budin WC, Maislin G. Medical factors and patterns of adjustment to breast cancer. Psychooncology 1996;5:31.
48. Parker PA, Baile WF, de Moor C, Cohen L. Psychosocial and demographic predictors of quality of life in a large sample of cancer patients. Psychooncology 2003;12:183.
49. Goodwin PJ, Ennis M, Bordeleau LJ, et al. Health-related quality of life and psychosocial status in breast cancer prognosis: Analysis of multiple variables. J Clin Oncol 2004;22:4184.

Address reprint requests to:
Ulrike Boehmer, Ph.D.
CHQOER
200 Springs Road (152)
Bedford, MA 01730

E-mail: boehmer@bu.edu

This article has been cited by:

1. Ulrike Boehmer, Jennifer Potter, Deborah J. Bowen. 2009. Sexual Functioning After Cancer in Sexual Minority Women. *The Cancer Journal* **15**:1, 65-69. [CrossRef]
2. Jessica P. Brown, J. Kathleen Tracy. 2009. Lesbians and cancer: an overlooked health disparity. *Cancer Causes & Control* **19**:10, 1009-1020. [CrossRef]
3. Ulrike Boehmer, Rhonda Linde, Karen M. Freund. 2007. Breast Reconstruction following Mastectomy for Breast Cancer: The Decisions of Sexual Minority Women. *Plastic and Reconstructive Surgery* **119**:2, 464-472. [CrossRef]

ORIGINAL ARTICLE

"Shotgun" Versus Sequential Testing

Cost-Effectiveness of Diagnostic Strategies for Vaginitis

Phyllis L. Carr, MD,[1] Michael B. Rothberg, MD, MPH,[2,3] Robert H. Friedman, MD,[1] Donna Felsenstein, MD,[4] Joseph S. Pliskin, PhD[5]

[1]Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [2]Division of General Medicine and Geriatrics, Department of Medicine, Baystate Medical Center, Boston, MA, USA; [3]Tufts University School of Medicine, Boston, MA, USA; [4]Division of Infectious Disease, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; [5]Departments of Industrial Engineering and Management, and Health Systems Management, Ben Gurion University of the Negev, Beer Sheva, Israel.

BACKGROUND: Although vaginitis is a common outpatient problem, only 60% of patients can be diagnosed at the initial office visit of a primary care provider using the office procedures of pH testing, whiff tests, normal saline, and potassium hydroxide preps.

OBJECTIVE: To determine the most cost-effective diagnostic and treatment approach for the medical management of vaginitis.

DESIGN: Decision and cost-effectiveness analyses.

PARTICIPANTS: Healthy women with symptoms of vaginitis undiagnosed after an initial pelvic exam, wet mount preparations, pH, and the four criteria to diagnose bacterial vaginosis.

SETTING: General office practice.

METHODS: We evaluated 28 diagnostic strategies comprised of combinations of pH testing, vaginal cultures for yeast and *Trichomonas vaginalis*, Gram's stain for bacterial vaginosis, and DNA probes for *Neisseria gonorrhoeae* and Chlamydia. Data sources for the study were confined to English language literature.

MEASUREMENT: The outcome measures were symptom-days and costs.

RESULTS: The least expensive strategy was to perform yeast culture, gonorrhoeae and Chlamydia probes at the initial visit, and Gram's stain and Trichomonas culture only when the vaginal pH exceeded 4.9 ($330, 7.30 symptom days). Other strategies cost $8 to $76 more and increased duration of symptoms by up to 1.3 days. In probabilistic sensitivity analysis, this strategy was always the most effective strategy and was also least expensive 58% of the time.

CONCLUSIONS: For patients with vaginitis symptoms undiagnosed by pelvic examination, wet mount preparations and related office tests, a comprehensive, pH-guided testing strategy at the initial office visit is less expensive and more effective than ordering tests sequentially.

KEY WORDS: vaginitis; diagnosis; treatment; cost-effectiveness; decision analysis.
DOI: 10.1111/j.1525-1497.2005.0188.x
J GEN INTERN MED 2005; 20:793–799.

Vaginitis is one of the 25 most common medical reasons for consulting a physician in the United States,[1] resulting in 5 to 10 million office visits per year.[2] Although they encounter vaginitis frequently, primary care practitioners have difficulty making an etiologic diagnosis in the office setting.[3–9] Specialized centers report diagnosing 80% to 90% of women from the initial visit,[10] but primary care providers report only a 50% to 60% chance of reaching a diagnosis despite extensive laboratory testing,[4] because the typical practitioner lacks the office resources and expertise found in specialized centers.

The cost of diagnosis in patients presenting with vaginitis symptoms can vary dramatically depending on the amount of testing conducted at the initial visit. Most experts agree that diagnosis should begin with a complete pelvic examination, determining the source of the discharge (cervical or vaginal), gross evaluation of the discharge for consistency, and adherence to the vaginal walls or cervix; followed by wet mount preparations with saline looking for clue cells, white blood cells, and *Trichomonas vaginalis*; potassium hydroxide (KOH) testing for yeast and the whiff test; and pH testing.[11] If these are nondiagnostic, however, should the practitioner: 1) perform extensive testing at the initial visit, or 2) limit initial testing to simple, inexpensive bacteriologic tests to reduce costs and burden to the patient, realizing that more patients will require follow-up visits and testing to reach a definitive diagnosis? We performed decision and cost-effectiveness analyses to compare these strategies for patients who present with symptoms of vaginitis, but who remain undiagnosed after initial office-based evaluation.

METHODS

We constructed a decision tree using a standard computer program (Decision Maker 7.06, Pratt Medical Group, Boston) and data derived from relevant peer-reviewed articles in the English language. We then analyzed the outcomes of 28 different office-based diagnostic strategies for the medical management of vaginitis. We considered a population of healthy women who present with vaginal discharge, pruritus, irritation or odor who could not be diagnosed by initial office evaluation consisting of pH; wet mount (KOH and normal saline [NS]) preparations for Candida species, *Trichomonas vaginalis*, and mucopurulent discharge; and the four criteria to diagnose bacterial vaginosis (thin, homogeneous discharge; pH>4.5; clue cells and a positive whiff test). We assumed that practitioners would not perform Gram's stain in the office. Pregnant patients and those who used over-the-counter treatment for vaginitis were excluded.

The authors do not have any conflicts of interest, financial or otherwise, to declare.

Dr. Carr and Dr. Rothberg contributed equally to this article.

Address correspondence and requests for reprint to Dr. Carr: Office of Student Affairs, Boston University School of Medicine, 715 Albany Street, L-109, Boston, Massachusetts 02118 (e-mail: plcarr@bu.edu).

*See editorial by Willet and Centor, p. 871

Received for publication February 11, 2005
and in revised form April 26, 2005
Accepted for publication May 3, 2005

Table 1. Probabilities That Specific Etiological Agents Cause Vaginitis

Etiologic Agents	Base* Case	Low†	High†	References
Candida species	.25	.20	.33	2,10,11
Bacterial vaginosis	.35	.28	.50	7,10,11
Trichomonas vaginalis	.15	.10	.20	2,10
Chlamydia trachomatis	.05	.02	.07	5,7,12,13
Neisseria gonorrhoeae	.02	0	.02	5,12,13
Herpes	.02	.01	.02	7,12
Other	.16			

*Base case is the estimate used in the model, derived from the literature, cited in column 5.
†"Low" and "High" refer to low and high ranges of the probabilities, derived from the literature, cited in column 5.

Diagnosis and Treatment

The possible etiologies of vaginitis are yeast, *Trichomonas vaginalis*, and bacterial vaginosis. Cervicitis caused by *Chlamydia trachomatis* and *Neisseria gonorrhoeae* can also mimic vaginitis. Herpes infection can present similarly to vaginitis. Other etiologies are less common (Table 1). For simplicity, we assumed that each patient would have only one causative organism.

We considered initial diagnostic strategies that incorporated the following tests, either alone or in combination: vaginal pH, vaginal cultures for Candida and Trichomonas, Gram's stain for bacterial vaginosis (BV), and DNA probes for *N. gonorrhoeae* and Chlamydia (GC/Chlamydia probes). We assumed that pH test results would be available during the examination and could be used to guide further testing at the visit with a normal pH excluding BV, trichomonas, and atrophic vaginitis. We estimated that all other test results besides the initial office evaluations would take two days. Specific treatments were based on Centers for disease Control and Prevention (CDC) guidelines (Table 2). We also considered 2 empirical treatment strategies: 1) treatment guided by vaginal pH (treatment with single dose fluconazole for Candida when the pH is less than 4.9, or treatment with 2 g of metronidazole to cover Trichomonas and/or BV when the pH is greater than 4.9) or 2) treatment with both fluconazole and metronidazole.

Further Evaluation

Patients who responded to initial treatment were considered cured. We assumed that patients who failed the initial therapy would receive the prescription for the second therapy by telephone, but those patients who failed empiric therapy or who were undiagnosed after the initial round of tests would return for an office visit and undergo all previously unordered tests. We estimated that symptoms from causes of vaginitis other than those modeled would resolve 20% of the time without treatment before a second visit.

Referral to Specialists

The model presumed that patients not responding to two courses of therapy and those who were undiagnosed despite a complete battery of tests would be referred to an infectious disease or gynecology specialist, who would repeat all tests and treat all diagnosable patients appropriately. For patients with symptoms due to causes other than those modeled, we estimated that specialists could successfully treat half.

Adverse Outcomes

The probability of treatment side effects appears in Table 2. For simplicity, we assumed that side effects would last 2 days and be equal in severity to the vaginitis symptoms.

Outcome Measures

The model summed the costs of all diagnostic tests, office visits, and referrals. Effectiveness was expressed as change in symptom days. We assumed that all vaginitis symptoms would be of equal severity regardless of etiology, would persist until properly treated, and would disappear on the third day of successful treatment.[12]

Sensitivity Considerations

All quantitative assumptions were subject to one-way sensitivity analysis to discern their relative impact on the cost-effectiveness of different strategies. We also conducted a probabilistic analysis in which we varied all inputs simultaneously to determine confidence intervals for the results. We

Table 2. Treatment Regimens for Vaginitis

Etiological Agents and Treatments	Course	Side Effects (%)	Average Wholesale Price ($)	Cure Rate (%)
Candida species				
Fluconazole	150 mg PO × 1	10 (5 to 13)	11.89	85 (72 to 93)* 14–16
Terconazole,† 0.8% cream	HS × 3 nights	10 (5 to 18)	30.96	84 (80 to 94)* 16,17
Bacterial vaginosis				
Metronidazole	500 mg PO BID × 7 days	10 (10 to 15)	3.36	80 to 92* 18
Metronidazole	2 g PO × 1	7 (5 to 10)	0.48	70 to 87* 18
Trichomonas vaginalis				
Metronidazole	2 g PO × 1	7 (5 to 10)	0.48	90 (82 to 93)* 10,18
Metronidazole	500 mg PO BID × 7 days	10 (10 to 15)	3.36	93 (90 to 95)* 10,18
Cervicitis				
Ceftriaxone	250 mg IM	10 (8 to 22)	8.82	98 (95 to 98)* 18
Doxycycline	100 mg PO BID × 7 days	10 (8 to 15)	1.68	95 (90 to 98)* 18
Azithromycin	2 g PO × 1	7 (5 to 10)	20.98	98 (96 to 98)* 18

*References course of treatment and its cure rate.
†Terconazole was chosen for second treatment of vaginitis as it covers Candida glabrata *and* Candida tropicalis *in addition to* Candida albicans.
PO, per os; IM, intramascular; BID, twice daily; HS, nightly.

Table 3. Test Characteristics of Diagnostics Tests Used in Vaginitis

	Base* Case	Low†	High†	References
Sensitivity of wet mount in				
Candida species	.60	.40	.80	10,11,20,21
Bacterial vaginosis	.90	.80	.95	3,8,10,11,22
Trichomonas vaginalis	.67	.40	.80	2,8,10,11,20,23
Chlamydia trachomatis	.30	.18	.42	25
Neisseria gonorrhoeae	.30	.18	.42	25
Other	0	0	0	Expert assumption
Sensitivity of				
Candida culture	.95	.95	.95	11
Gram's stain	.95	.93	.95	8,10,11,22,26
Trichomonas vaginalis culture	.95	.89	.95	8,10,11
DNA probe (GC)	.90	.90	1	27–30
DNA probe (Chlamydia)	.90	.90	1	12,19,31–33
Probability of pH > 4.9 in				
Candida species	.29	.25	.34	7,11
Bacterial vaginosis	1	1	1	7,11
Trichomonas vaginalis	1	1	1	7,11
Chlamydia trachomatis	.67	.60	.74	7
Neisseria gonorrhoeae	.67	.60	.74	Expert assumption
Other	.67	.60	.74	7

*Base case is the estimate used in the model, derived from the literature, cited in column 5.
†"Low" and "High" refer to low and high ranges of the probabilities, derived from the literature, cited in column 5.

performed 1,000 Monte Carlo simulations, each time choosing random values from within each variable's 95% confidence interval using logit distributions.

Data and Estimates

Etiologic Agents. Yeast causes 20% to 33% of vaginitis symptoms[2,10,11] (Table 1), BV 28% to 50% [7,10,11], Trichomonas 10% to 20% [2,10,11] and cervicitis 2% to 7%.[5,7,13,19] Initial office evaluation correctly diagnoses Candida species 60% of the time [10,11] (Table 3), Trichomonas 70% of the time,[2,8,10,11,23,24] GC or Chlamydia 30% of the time,[25] and BV 90% of the time.[3,8,10,11] Using these data, we calculated the conditional probability of each of these etiologies given a negative initial office evaluation, as well as the probability of each diagnosis depending on vaginal pH (Table 4). For simplicity, we assumed all subsequent tests were conditionally independent and 100% specific.

Treatment Efficacy and Side Effects. We based our estimates of the efficacy and side effects of treatments on data from randomized clinical trials (Table 2). We assumed vaginal creams would cause contact dermatitis in 10% of patients, and that fluconazole would cause gastrointestinal symptoms in the same proportion.[14–16] Metronidazole causes secondary yeast infection and gastrointestinal symptoms in 7% to 10% of patients, depending on duration of treatment.[18]

Costs. All costs were in US dollars for the year 2003 and assumed the societal perspective (Table 5). We included all direct medical costs combining the costs of diagnostic testing, physician visits and prescription medications, as well as indirect costs from lost productivity during physician visits. Costs for diagnostic tests and physician visits were based on the 2003 Medicare Fee Schedule. Drug costs reflect average wholesale prices. Labor costs were based on US average employee compensation for 2003.[34]

RESULTS

Prevalence of Disease

Table 1 shows the prevalence of the common causes of vaginitis in patients presenting to practitioners. Using the sensitivity of office evaluation described above for each of the causes, the model calculated the prevalence of each cause in the subset of patients with non-diagnostic office wet mount preparations, both overall and based on vaginal pH (Table 4). Regardless of pH, about 45% of these patients have a diagnosis of "other" which cannot be determined by common office tests.

Table 4. Prevalence of Etiological Agents After Negative Wet Prep and pH Testing

Etiological Agents	Prevalence in General Practice*	Calculated Prevalence After Negative Wet Prep	Calculated Prevalence if pH Is >4.9	Calculated Prevalence if pH Is <4.9
Candida species	.25	.21	.09	.44
Bacterial vaginosis	.35	.09	.13	.00
Trichomonas vaginalis	.15	.13	.19	.00
Chlamydia trachomatis	.05	.09	.09	.09
Neisseria gonorrhoeae	.02	.04	.04	.03
Other	.18	.45	.46	.44

*See references in Table 2, column 5.

Table 5. Costs

Test	Cost ($)*
Wet mount preparation	8.06
Gram Stain	8.06
Vaginal culture (Candida species)	15.86
Trichomonas vaginalis culture	17.86
Neisseria gonorrhoeae DNA probe	37.86
Chlamydia DNA probe	37.86
Herpes DNA amplification probe†	66.27
Human papillomavirus testing†	66.27
Physician office visit	50.32
One hour of patient time for physician visit	24.48
Specialist consultation	164.34

**Laboratory and physician costs are from the Medicare Fee Schedule. Indirect costs from lost productivity come from the US Bureau of Labor Statistics.*
†As part of specialist's evaluation.

Diagnostic Strategies

Table 6 lists all 28 diagnostic strategies with the associated average costs and mean symptom days, as determined by the model. The least expensive diagnostic strategy was the most comprehensive: begin with pH testing, yeast cultures and DNA probes for gonorrhoeae and Chlamydia for all patients, but perform Gram's stain and Trichomonas cultures only when vaginal pH exceeded 4.9 ($330, 7.30 symptom days). Other strategies increased average costs by $5 to $81 per patient, and increased duration of symptoms by up to 1.3 days. In general, diagnostic strategies which entailed fewer tests during the initial visit, especially those not testing for yeast, resulted in higher costs because of the greater number of follow-up office visits and the high cost of referral.

Empiric Treatment

Compared to testing strategies, empiric treatment strategies resulted in fewer referrals (40% vs 41% to 46%), but more adverse effects (11% to 19% vs 6%). Diagnostic testing followed by pH-guided empirical therapy while awaiting test results was superior to both empirical treatment and testing alone. Depending on the testing strategy, the savings associated with adding empirical treatment while awaiting test results ranged from $8 to $63 (mean savings $39), and decreased symptom duration by 0.6 and 1.3 days, even after accounting for additional side effects and costs related to empirical treatment. Empirical treatment while awaiting test results was beneficial even if the side effects of the medications were three times as severe as the vaginitis itself.

Sensitivity Analyses

Strategies including empiric treatment and those without were analyzed separately. Results were similar. Despite the small

Table 6. Average Cost and Utility of 28 Strategies for Initial Evaluation of Vaginitis Symptoms

Strategy*	Average Cost ($)	Mean Symptom Days	Incremental Cost† ($)	Incremental Symptom Days‡	Incremental Cost-Effectiveness $/Symptom Day Avoided§
YSBTp	330	7.30			
YBTp	335	7.59	5	0.29	Dominated
YSBT	337	7.30	8	0.01	Dominated
YBT	338	7.59	9	0.29	Dominated
YTp	351	7.87	22	0.57	Dominated
YT	353	7.87	24	0.57	Dominated
YBp	354	7.92	24	0.62	Dominated
YB	354	7.92	25	0.62	Dominated
YSBp	358	7.68	29	0.38	Dominated
YSp	370	7.96	41	0.66	Dominated
Y	372	8.20	42	0.90	Dominated
YSB	372	7.70	42	0.40	Dominated
Yp	373	8.20	43	0.90	Dominated
BT	374	8.19	44	0.89	Dominated
BTp	374	8.19	45	0.89	Dominated
YS	382	7.98	53	0.68	Dominated
SBTp	385	7.97	56	0.67	Dominated
SBT	386	7.97	57	0.67	Dominated
T	392	8.47	62	1.17	Dominated
Tp	393	8.47	63	1.17	Dominated
B	395	8.53	66	1.23	Dominated
Bp	396	8.53	67	1.23	Dominated
SB	397	8.31	67	1.01	Dominated
STp	397	8.25	67	0.95	Dominated
ST	397	8.25	67	0.95	Dominated
SBp	397	8.31	67	1.01	Dominated
S	410	8.58	80	1.28	Dominated
Sp	411	8.58	81	1.28	Dominated

**Y, vaginal culture for Candida species; S, sexually transmitted disease testing including probes for* Neisseria gonorrhoeae *and Chlamydia; B, Gram stain for bacterial vaginosis T, vaginal culture for* Trichomonas vaginalis*; p, vaginal pH testing.*
†Incremental costs represent the difference between the strategy and the next best nondominated strategy.
‡Incremental symptom days represents the difference between the strategy and the next best nondominated strategy.
§The difference in cost divided by the difference in quality-adjusted life expectancy for each strategy compared with the next best nondominated strategy. A dominated strategy costs more and is less effective than another available strategy.

differences between the strategies in cost and effectiveness, the model was robust with respect to the preferred strategies throughout the ranges tested. Under most circumstances, comprehensive testing guided by pH at the initial visit was least expensive. Under certain circumstances, however, ordering GC/Chlamydia probes at the initial visit was expensive (Fig. 1).

Initial GC/Chlamydia testing was least cost-effective when the sensitivities of the initial office evaluations to detect Yeast, BV or Trichomonas were low. As a result, these entities would be relatively more prevalent and GC and Chlamydia less prevalent. One benefit of GC/Chlamydia testing at the initial visit is avoidance of future visits. However, if the cost of an office visit is low, or undiagnosable patients get better without treatment, initial GC/Chlamydia probes are less cost-effective.

Under some circumstances checking pH resulted in a worse outcome. The cost-effectiveness of pH testing relies on the ability of low pH to rule out BV and Trichomonas. BV by definition has a high pH and studies of Trichomonas show that it uniformly causes a high pH. However, if the sensitivity of pH testing for Trichomonas is less than 85%, pH testing should not be done, because false negative tests will lead to unacceptable delay in the diagnosis of Trichomonas.

Results of Probabilistic Sensitivity Analysis

In 1000 Monte Carlo simulations, only two diagnostic strategies were ever cost-effective. Testing for yeast, BV and Trichomonas without GC/Chlamydia probes was the least expensive strategy 42% of the time. Comprehensive testing including GC/Chlamydia probes was always the most effective strategy, and was also the least expensive strategy in 58% of the iterations.

DISCUSSION

Understanding the monetary costs and clinical outcomes of managing common medical problems is increasingly necessary for health care providers and to society.[35–37] Unfortunately, the costs and benefits of alternative diagnostic approaches are not always apparent. In medicine, the traditional approach to diagnosis includes a history and physical examination, followed by the development of a differential diagnosis and the performance of serial testing to arrive at the most likely diagnosis. This approach is particularly appealing when tests are potentially harmful, have a high false positive rate, or are expensive.[38] The "shotgun" approach, where the clinician orders every possible test during the initial contact, is considered wasteful at best, and potentially harmful, in that it can increase the likelihood of test complications and diagnostic confusion.[39] However, in the current healthcare environment, the cost of tests may be small in comparison to the costs of additional office visits, referrals, or emergency department visits resulting from delayed diagnosis.

Symptoms of vaginitis are common in medical practice. The history and physical findings are seldom helpful in diagnosing the etiology of vaginitis symptoms, including the character of the discharge or the presence of an odor.[4] While a majority of patients can be diagnosed with simple wet mount preparations and related initial office tests, a substantial number remain undiagnosed. The cost of obtaining a diagnosis for these patients is high, at least $330 (Table 6), more than 6 times the average cost of the office visit ($50.32) itself.

The task of choosing a diagnostic strategy is daunting. The clinician must choose from among five tests, each with a different sensitivity, specificity, and cost, then consider empirical treatment while awaiting test results. The clinician must consider the three common causes of vaginitis, each

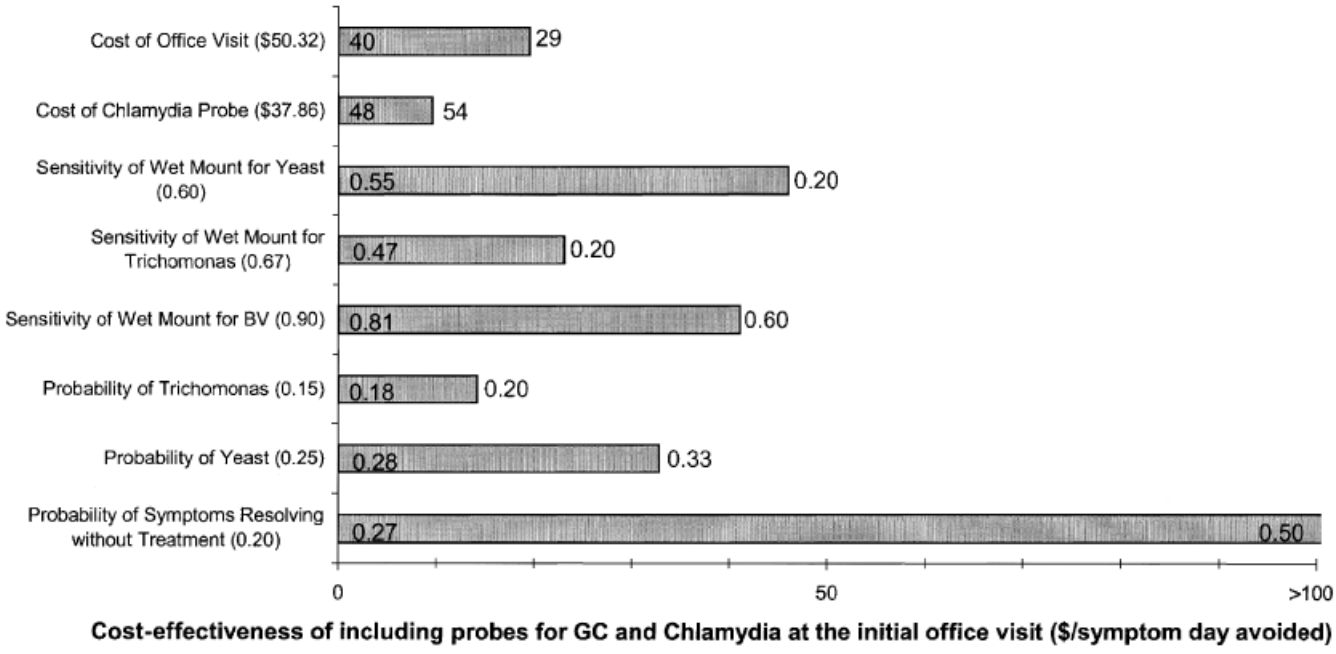


FIGURE 1. One-way sensitivity analysis. Bars show the range of incremental cost-effectiveness of adding *Neisseria gonorrhoeae* and Chlamydia probes to yeast culture and pH guided testing for bacterial vaginosis and Trichomonas. Baseline values for each variable are shown in parenthesis. Left-hand numbers represent the threshold at which GC/Chlamydia probes become cost-saving. Right-hand numbers represent the upper bound in sensitivity analysis.

with a different prevalence, as well as a number of less common causes of these symptoms, such as gonorrhoeae, and Chlamydia. Ordering every potentially useful test at the initial visit may result in the quickest diagnosis, but is it a reasonable use of resources? Our analysis suggests that immediate testing for all diagnostic possibilities—except BV and trichomonas when vaginal pH is normal—is not only reasonable, but the least expensive clinical strategy, considering the cost of follow-up visits and referrals.

Empirical treatment with fluconazole, metronidazole or both in place of testing was not cost-effective, because the majority of patients would not be cured and many would incur unnecessary treatment with attendant side effects, delayed diagnosis and associated costs. By contrast, adding pH-guided empiric therapy to any testing strategy while waiting for culture results both shortened symptom duration and decreased cost. Patients treated empirically by pH had immediate relief, without having to wait for culture results, and if cured, did not have to return for further office visits.

Patients who are at high risk of developing complications from vaginal infections, such as pregnant women or women who are scheduled for an abdominal or vaginal procedure require accurate, diagnostically guided treatment. Empiric treatment would be inappropriate for these women since partial treatment may interfere with interpretation of subsequent tests.

There are a number of limitations to our study. Some results may not be generalizable to all medical practices. For example, the prevalence of disease among patients with negative initial office tests is unknown. For our model, we calculated this probability using the prevalence of each of these etiologies of vaginitis symptoms and the known sensitivity of the initial office tests. In individual practices, however, these results will vary depending upon the prevalence of these etiologies in the specific community and the clinician's skill in interpreting the initial office tests. For example, if a clinician does not easily identify clue cells, undiagnosed patients will have a higher prevalence of BV, and a lower prevalence of GC and Chlamydia, making GC/Chlamydia probes less cost-effective for that clinician. Alternatively, if the patient is at high risk for sexually transmitted diseases, or if the practitioner has difficulty distinguishing between the discharge of cervicitis and that of vaginitis, GC and Chlamydia probes will invariably be cost-saving.

Our study was also limited by available data. Although 45% of the patients in our model could not be diagnosed by simple tests available to the average clinician in office practice, it is unclear how much medical intervention helps these women. We tested these variables in the sensitivity analysis and found that if symptoms resolved in more than one-third of these women without treatment, then testing should be deferred to the follow-up visit, as many women would improve before a follow-up appointment. Because these women constitute such a large percentage of vaginitis patients, more studies are needed to characterize the etiology and prognosis of this condition. For simplicity, we also assumed that each case of vaginitis was caused by a single organism. In reality, concurrent diagnoses in genital tract infections are frequent. This fact only strengthens the argument for broad testing on the initial visit rather than testing for, and then treating, one entity at a time.

There are a number of strengths to our study. A decision analysis allowed us to study this complex problem when doing so in a randomized clinical trial would have been impractical. A second strength is the inclusive nature of our study, which makes the results broadly applicable.

While the cost differences among the diagnostic strategies are modest (no more than $81 per case, on average), the high prevalence of vaginitis makes these differences expensive in aggregate. Choosing one diagnostic strategy over another could result in savings of tens of millions of dollars annually on a national basis. Similarly, 1.3 symptom-days may seem trivial, but when multiplied by one million patients it represents 3,600 patient-years of vaginitis.

CONCLUSION

Vaginitis is common, yet often difficult for primary care practitioners to diagnose effectively in the office setting. Our study suggests that considerable savings and decreased symptoms can be achieved by using vaginal pH to guide testing and treatment at the initial office visit for those patients who are undiagnosed after a complete pelvic examination, evaluation of the discharge, whiff test, pH and wet mount preparations. Under most circumstances, testing for GC and Chlamydia will also improve outcomes and save money.

REFERENCES

1. **Paavonen J, Stamm WE.** Lower genital tract infections in women. Infect Dis Clin North Am. 1987;1:179.
2. **Sparks JM.** Vaginitis. J Repro Med. 1991;36:745–52.
3. **Nugent RP, Krohn MA, Hillier SL.** Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. J Clin Microbiol. 1991;29:297–301.
4. **Schaaf VM, Perez-Stable EJ, Borchardt K.** The limited value of symptoms and signs in the diagnosis of vaginal infections. Arch Intern Med. 1990;150:1929–33.
5. **Berg AO, Heidrich FE, Fihn SD, et al.** Establishing the cause of genitourinary symptoms in women in a family practice. JAMA. 1984;251: 620–5.
6. **Carr PL, Thabault P, Levenson S, Friedman RH.** Vaginitis in a community based practice. Clin Res. 1992;40:554A.
7. **Bleker OP, Folkertsma K, Dirks-Go SIS.** Diagnostic procedures in vaginitis. Eur J Obstet Gynec Reprod Biol. 1989;31:179–83.
8. **Eschenbach DA, Hillier SL.** Advances in diagnostic testing for vaginitis and cervicitis. J Reprod Med. 1989;34:555–65.
9. **Anderson MR, Klink K, Cohrssen A.** Evaluation of vaginal complaints. JAMA. 2004;291:1368–79.
10. **Sobel JD.** Vaginitis in adult women. Obs Gyn Clin N Am. 1990;17: 851–79.
11. **Sobel JD.** Vaginitis. N Engl J Med. 1997;337:1896–1903.
12. **Holmes KK.** Lower genital tract infections in women: cystitis/urethritis vulvovaginitis and cervicitis. In Holmes KK, Sparling PF, Mardh P, et al., eds. Sexually Transmitted Diseases. New York: McGraw-Hill; 1999:527–45.
13. **Gratton CA, Lim-fong R, Prasad E, Kibsey PC.** Comparison of a DNA probe with culture for detecting *Chlamydia trachomatis* directly from genital specimens. Mol Cell Probes. 1990;4:25–31.
14. **Sobel JD, Brooker D, Stein GE, et al.** Single oral dose fluconazole compared with conventional clotrimazole topical therapy of Candida vaginitis. Am J Obstet Gynecol. 1995;172:1263–8.
15. Oral fluconazole for vaginal candidiasis. Med Lett. 1994;36:81–2.
16. **Reef SE, Levine WC, McNeil MM, et al.** Treatment options for vulvovaginal candidiasis, 1993. Clin Infec. Dis. 1995;20(suppl 1):S80–90.
17. **Clark C, Cooper CL, Gordon SF, et al.** A multicenter comparison of one-dose tioconazole ointment with three–dose terconazole cream in vulvovaginal candidiasis. J Women' Health. 1993;2:189–96.
18. **Centers for Disease Control and Prevention**. Guidelines for treatment of sexually transmitted disease. MMWR. 1993;42:171.
19. **Iwen PC, Blair TM, Woods GL.** Comparison of the Gen-Probe PACE 2 system, direct fluorescent–antibody, and cell culture for detecting *Chlamydia trachomatis* in cervical specimens. Am M Clin Pathol. 1991;95: 578–82.
20. Sexually transmitted disease guidelines, CDC. MMWR. 1993;42:171.

21. **Bertholf MZ, Stafford MJ.** Colonization of *Candida albicans* in vagina, rectum and mouth. J Fam Pract. 1983;95:919–24.
22. **Amsel R, Totten PA, Spiegel CA, et al.** Nonspecific vaginitis: diagnostic criteria and microbial and epidemiological associations. Am J Med. 1983;74:14–22.
23. **Krieger JN, Tam MR, Stevens CE, et al.** Diagnosis of trichomoniasis: comparison of conventional wet mount preparation with cytological studies, cultures and monoclonal antibody staining of direct specimens. JAMA. 1988;259:1223.
24. **Cotch MF, Pasterek JG, Nugent RP, Yerg DE, Martin DH, Eschenbach DA.** Demographics and behavioral predictors of Trichomonas vaginitis infection among pregnant women. Obstet Gynecol. 1991;78: 1087–92.
25. **Landers DV, Weisenfeld HC, Heine RP, Krohn MA, Hillier SL.** Predictive value of the clinical diagnosis of lower genital tract infection in women. Am J Obstet Gynecol. 2004;190:1004–10.
26 **Spiegel SA, Amsel R, Holmes KK.** Diagnosis of bacterial vaginosis by direct gram stain of vaginal fluid. J Clin Microbiol. 1983;18:170–77.
27. **Lewis JS, Kranig-Brown D, Trainor DA.** DNA probe confirmatory test for *Neisseria gonorrhosae*. J Clin Microbiol. 1990;28:2349–50.
28. **Schwebke JR, Zajackowski ME.** Comparison of DNA probe (Gen-Probc) with culture for the detection of Neisseria gonorrhoeae in an urban STD programme. Genitourin Med. 1996;72:108–10.
29. **Granato PA, Franz MR.** Evaluation of a prototype DNA probe test for the noncultural diagnosis of gonorrhea. J Clin Microbiol. 1989;27: 632–5.
30. **Szell A, Tisza T, Horvath A.** A comparative study for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* with DNA probe. Acta Microbiol Immunol Hung. 1994;41:291–3.
31. **Loeffelholz MJ, Lewinski CA, Silver SR, et al.** Detection of *Chlamydia trachomatis* in endocervical specimens by polymerase chain reaction. J Clin Microbiol. 1992;30:2847–51.
32. **Lees MI, Newman DM, Garland SM.** Comparison of a DNA probe assay with culture for the detection of *Chlamydia trachomatis*. J Med Microbiol. 1991;35:159–61.
33. **Mercer LJ, Robinson DC, Sahm DF, Lawrie MJ, Hajj SN.** Comparison of chemiluminescent NA probe to cell culture for the screening of *Chlamydia trachomatis* in a gynecology clinic population. Obstet Gynecol. 1990;76:114–7.
34. **U.S. Bureau of Labor Statistics.** Employer Cost for Employee Compensation. Accessed December 2, 2003 (http://www.bls.gov/ncs/ect/home.htm#data).
35. **Bloom BS, Wierz DJ, Pauly MV.** Cost and price of comparable branded and generic pharmaceuticals. JAMA. 1986;256:2523–30.
36. 1993 Red Book. Montvale, NJ: Medical Economics Data Inc; 1993.
37. **Edelson JT, Weinstein MC, Tosteson ANA, et al.** Long-term cost-effectiveness of various initial monotherapies for mild to moderate hypertension. JAMA. 1990;263:407–13.
38. **Pauker SG, Kassirer JP.** The threshold approach to clinical decision making. N Engl J Med. 1986;302:1109–17.
39. **Thibault GE.** Diagnostic strategy—the shotgun versus the arrow. N Engl J Med. 1995;332:321–5.

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Contraception

Contraception 71 (2005) 157–160

Original research article

Emergency contraception knowledge among women in a Boston community

Cynthia H. Chuang[a,*], Karen M. Freund[b]

Massachusetts Emergency Contraception Network

[a]Division of General Internal Medicine, Penn State Milton S. Hershey Medical Center, P.O. Box 850, Hershey, PA 17033-0850, USA

[b]Women's Health Unit, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA 02215, USA

Received 2 August 2004; accepted 23 August 2004

Abstract

This study assesses the baseline knowledge of emergency contraception (EC) in a Boston neighborhood. A written survey was distributed to women aged 18–44 years in the Boston neighborhood of Jamaica Plain. Of the 188 participants, 82% have heard of EC. Knowledge disparities by race/ethnicity groups were seen, with only 51% of Latina women and 75% of Black women having heard of EC compared with 99% of White women ($p<.0001$ and $p=.002$, respectively). Of the entire cohort, 39% knew that EC works by preventing pregnancy, 48% knew that it should be taken within 72–120 h of unprotected intercourse and 44% knew that it is only available by prescription in Massachusetts. Only 25% of women have ever discussed EC with a health care provider, and only 12% have ever received an advance prescription. A community education campaign aimed at reproductive-age women, health care providers and pharmacists has been tailored to address these knowledge deficits.

Keywords: Contraception; Emergency contraception; Family planning

1. Introduction

Emergency contraception (EC) has great potential for reducing the number of unintended pregnancies and abortions in the United States [1]. Although only 1% of women have ever used EC [2], as many as 51,000 abortions were likely averted by its use in 2000 [3]. Barriers that have been associated with nonuse of EC are limited knowledge, shame associated with asking for EC and problems in gaining access to EC [4]. Educating reproductive-age women, health care providers and pharmacists about EC is the first step in reducing these barriers to ensure prompt availability of this method when needed.

The Massachusetts EC Network is a coalition of community organizations, medical providers and government agencies who share a common goal of reducing unintended pregnancies by improving knowledge of and access to EC in Massachusetts. The network has launched an educational campaign in the community of Jamaica Plain in Boston, MA, which began in the fall of 2003. The goal of the campaign is to increase awareness about EC through outreach to the public, health care providers and pharmacists. This survey assessed baseline knowledge about EC among reproductive-age women living in the community before the campaign for the purpose of guiding the design and content of the educational intervention.

2. Methods

This survey project was conducted in May 2003 before the start of the Massachusetts EC Network's educational campaign in the community of Jamaica Plain in Boston, MA. This community was chosen because of its socioeconomic and ethnic heterogeneity. Additionally, local community groups and health centers had indicated willingness to participate in the community-based intervention. The Institutional Review Board of the Boston Medical Center approved this study.

* Corresponding author. Tel.: +1 717 531 8161; fax: +1 717 531 7726.
E-mail address: cchuang@psu.edu (C.H. Chuang).

0010-7824/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.contraception.2004.08.007

Table 1
Age and race/ethnicity characteristics ($N=188$)

Characteristic	*n*
Age (years)	
18–24	58 (31%)
25–29	42 (22%)
30–34	41 (22%)
35–39	26 (14%)
40–44	21 (11%)
Race/Ethnicity	
Hispanic/Latina	55 (29%)
White, not Hispanic/Latina	106 (57%)
Black, not Hispanic/Latina	20 (11%)
Other	6 (3%)

2.1. Study participants

Participants were recruited at public street locations frequented by local residents (near a grocery, retail stores, a neighborhood street fair). We intentionally did not recruit near any health center to avoid oversampling women routinely accessing health care services. Research assistants approached women and asked them if they were Jamaica Plain residents. If so, they were asked to participate in the study. Subjects were informed that the anonymous survey was for the purpose of medical research and that their participation was voluntary. Women who agreed to participate were given the survey. After completing the survey, women were offered a small bag containing chocolates, condoms and a pamphlet with basic facts about EC.

2.2. Survey instrument

The survey was a single-sided, 11-item written questionnaire designed for this study. Demographics on age, race/ethnicity and community residence were collected. Women who were not aged 18–44 years nor were community residents of Jamaica Plain were excluded from the study. Questions about EC are shown in Appendix A. The survey was designed at a 5th-grade literacy level and was available in both English and Spanish. The Spanish questionnaire was produced by translating the English questionnaire to Spanish and was then translated back to check for accuracy. The survey was pilot tested, in both English and Spanish, to a representative sample of women.

The main outcome was determined by the question, "Have you ever heard of emergency contraception (also called the 'morning-after pill')?" If a participant answered "no," she was finished with the survey. If she answered "yes," she continued answering the remaining questions about EC. We assumed that if a woman has never heard of EC or the morning-after pill, that she has no further detailed knowledge about it.

2.3. Statistical analysis

Data analysis was performed using SAS for Windows Version 8.2 (SAS Institute, Cary, NC). Frequencies of demographic characteristics were determined. To assess whether our sampling method recruited a cohort representative of the community, demographic characteristics of the study participants were compared with the 2000 U.S. Census data for demographic characteristics of women, aged 18–44 years, residing in Jamaica Plain, MA, using χ^2 tests for homogeneity. Responses to EC questions were compared by race/ethnicity using pairwise χ^2 analysis. All analyses were two-tailed, using a p value of .05 as criterion for statistical significance.

3. Results

Of the 238 surveys collected, 46 women were ineligible because they were not between the ages of 18 and 44 years, and 4 women were ineligible because they were not residents of the Jamaica Plain community. The remaining 188 surveys were analyzed. The age and race/ethnicity characteristics are shown in Table 1. The race/ethnicity distribution of the women in the sample was representative of women living in Jamaica Plain according to the 2000 U.S. Census data [5]; however, the age distribution did differ significantly ($p=.003$), with our study population overrepresenting the youngest age group.

Table 2
Responses (%) to EC questions by race/ethnicity ($N=188$)

Question	White ($n=106$)	Latina ($n=55$)	Black ($n=20$)	Other ($n=6$)	All women ($n=188$)
Has heard of EC	99	51*	75**	100	82
Has heard of EC and knows that it works by preventing pregnancy	44	31	35	33	39
Has heard of EC and knows of the 72 to 120h time window	64	24*	35**	50	48
Has heard of EC and knows that it is available only by prescription in Massachusetts	57	29**	25**	33	44
Has heard of EC and has discussed it with a health care provider	27	22	25	17	25
Has heard of EC and has received an advance prescription	13	11	10	0	12

* $p<.0001$ compared with White women.
** $p<.02$ compared with White women.

Responses to the EC questions are shown in Table 2. Overall, 82% of women have heard of EC or the morning-after pill. Only 39% of women knew that EC works by preventing pregnancy. About half of the women (48%) knew that EC should be taken within 72–120 h of unprotected intercourse (18% never heard of EC, 23% believed it had to be taken sooner, 1% believed it could be taken up to a week and 9% didn't know). When asked how to get EC in Massachusetts, 44% knew that it was available only by prescription in Massachusetts (18% never heard of EC, 2% thought it was not available at all, 4% thought it was sold over the counter and 31% didn't know). Still fewer women have ever discussed EC with a health care provider (25%) or received an advance prescription for it (12%).

Responses to the EC questions differed by race/ethnicity, as seen in Table 2. Only 51% of Latina women and 75% of Black women had heard of EC compared with 99% of White women ($p<.0001$ and $p=.002$, respectively). Fewer Latina women than White women knew that EC should be taken within 72–120 h of unprotected intercourse (24% vs. 64%; $p<.0001$) and that it is available only by prescription in Massachusetts (29% vs. 57%; $p=.001$). Similarly, Black women were also significantly less likely than White women to know that EC should be taken within 72–120 h of unprotected intercourse (35% vs. 64%; $p=.02$) and that EC is only available by prescription in Massachusetts (25% vs. 57%; $p=.01$).

4. Discussion

Our finding that 82% of women in our sample have heard of EC is higher than that reported in earlier national surveys. In 1994, only 55% of women in a nationally representative sample have heard of EC [2], and, in 1996, only 28% of American teenagers have heard of it [6]. This increase in awareness likely represents time trends as well as differences in the populations sampled. Although recognition of EC in this study was higher than in previous reports, specific knowledge about how and when EC works and how to get it is still poor. Latina and Black women in this sample were significantly less likely to have heard of EC and to have specific knowledge about it as compared with White women.

Few women have discussed EC with a health care provider or received an advance prescription for it. Several medical organizations have incorporated guidelines on routine counseling and advance prescriptions for EC [7–9]. Although this effective form of postcoital contraception was first described more than 30 years ago, physicians still have little experience with prescribing it [10–16]. Our findings inform the specialties of internal medicine, family medicine, obstetrics and gynecology and adolescent health of the need to increase EC education and provision into routine care.

Limitations of this study include the sampling framework that may not include a representative group of women from the community. Sampling street pedestrians in shopping areas may underrepresent low-income women or those with physical limitations. Despite these potential limitations, the study sample did approximate the ethnic demographic distribution of the reproductive-age women in the community according to most recent census information.

The purpose of this survey project was to inform the Massachusetts EC Network of the baseline knowledge of EC to aid in the design of a community education campaign. The survey provided the planning process with the knowledge that although most women have already heard of EC, detailed knowledge that women need about how to get and how to take EC needs to be improved. Therefore, educational programs that target local health care providers and pharmacists are being focused on methods of delivering effective and accurate information about EC to patients and clients. Furthermore, the survey results pointed to the need for specific interventions targeting minority women to reduce the disparity of knowledge in these groups. The Massachusetts EC Network has been in collaboration with Latina community leaders and community groups serving the Latina community in an effort to reach this group of women. A second survey project is planned after this community education campaign is completed to evaluate the effectiveness of this undertaking.

Acknowledgment

The authors thank the members of the Massachusetts Emergency Contraception Network who assisted with the survey development and subject recruitment for this project (Julia Gaggin Humphreys, Melissa Kogut, Adrienne Lazaroff, Stephanie Lowitt, Brooke Berens, Walker Armfield).

Appendix 1. EC survey questions

Have you ever heard of EMERGENCY CONTRACEPTION (also called the "morning-after pill")?
☐ Yes
☐ No (If you answer NO, you have finished the survey)

True or False: EMERGENCY CONTRACEPTION *prevents* a pregnancy from starting.
☐ True
☐ False
☐ I don't know

When should a woman take EMERGENCY CONTRACEPTION for it to work?
☐ 1 day before unprotected sex
☐ Up to 1 day after unprotected sex
☐ Up to 72 hours (3 days) after unprotected sex
☐ Up to 120 hours (5 days) after unprotected sex
☐ Up to 1 week (7 days) after unprotected sex
☐ I don't know

How can a woman get EMERGENCY CONTRACEPTION in Massachusetts?
☐ It is not available at all
☐ It is available only with a prescription

☐ It is available at the drugstore without a prescription (over-the-counter)
☐ I don't know

Has a doctor, nurse or pharmacist ever talked with you about EMERGENCY CONTRACEPTION?
☐ Yes
☐ No
☐ I'm not sure

Has a doctor or nurse ever given you a prescription for EMERGENCY CONTRACEPTION to have in case you needed it in the future?
☐ Yes
☐ No
☐ I'm not sure

References

[1] Trussell J, Stewart F, Guest F, Hatcher RA. Emergency contraceptive pills: a simple proposal to reduce unintended pregnancies. Fam Plann Perspect 1992;24:269–73.
[2] Delbanco SF, Mauldon J, Smith MD. Little knowledge and limited practice: emergency contraceptive pills, the public, and the obstetrician–gynecologist. Obstet Gynecol 1997;89:1006–11.
[3] Jones RK, Darroch JE, Henshaw SK. Contraceptive use among U.S. women having abortions in 2000–2001. Perspect Sex Reprod Health 2002;34:294–303.
[4] Free C, Lee RM, Ogden J. Young women's accounts of factors influencing their use and non-use of emergency contraception: in-depth interview study. BMJ 2002;325:1393.
[5] US Census Bureau. www.census.gov [accessed March 2003].
[6] Delbanco SF, Parker ML, McIntosh M, Kannel S, Hoff T, Stewart FH. Missed opportunities: teenagers and emergency contraception. Arch Pediatr Adolesc Med 1998;152:727–33.
[7] American College of Obstetricians and Gynecologists. ACOG practice bulletin: emergency oral contraception. Washington (DC): American College of Obstetricians and Gynecologists; 2001.
[8] American Medical Association. H-75.985 access to emergency contraception. Chicago: American Medical Association; 2002.
[9] U.S. Department of Health and Human Services. Healthy People 2010. With understanding and improving health and objectives for improving health. Washington (DC): U.S. Government Printing Office; 2000.
[10] Chuang CH, Waldman LJ, Freund KM, Ash AS. Emergency contraception: prescribing practices of general internists compared with other primary care physicians. Contraception 2004;69:43–5.
[11] Grossman RA, Grossman BD. How frequently is emergency contraception prescribed? Fam Plann Perspect 1994;26:270–1.
[12] Gold MA, Schein A, Coupey SM. Emergency contraception: a national survey of adolescent health experts. Fam Plann Perspect 1997;29:15–9, 24.
[13] Delbanco SF, Stewart FH, Koenig JD, Parker ML, Hoff T, McIntosh M. Are we making progress with emergency contraception? Recent findings on American adults and health professionals. J Am Med Womens Assoc 1998;53:242–6.
[14] Sills MR, Chamberlain JM, Teach SJ. The associations among pediatricians' knowledge, attitudes, and practices regarding emergency contraception. Pediatrics 2000;105:954–6.
[15] Golden NH, Seigel WM, Fisher M, et al. Emergency contraception: pediatricians' knowledge, attitudes, and opinions. Pediatrics 2001; 107:287–92.
[16] Beckman LJ, Harvey SM, Sherman CA, Petitti DB. Changes in providers' views and practices about emergency contraception with education. Obstet Gynecol 2001;97:942–6.

ELSEVIER

Contraception 72 (2005) 182–186

Contraception

Original research article

Emergency contraception: an intervention on primary care providers

Cynthia H. Chuang[a,*], Karen M. Freund[b]

[a]Division of General Internal Medicine, Pennsylvania State College of Medicine, Hershey, PA 17033-0850, USA
[b]Women's Health Unit, Section of General Internal Medicine, Evans Department of Medicine, Boston University School of Medicine, USA

Received 4 March 2005; revised 19 April 2005; accepted 21 April 2005

Abstract

Objective: We studied whether a single educational intervention can change provider knowledge, attitudes and practice patterns with respect to emergency contraception (EC).

Materials and Methods: Primary care providers completed a preintervention survey prior to attending a lecture on EC, and again 6 months later. There were 50 physicians, 4 advanced practice nurses and 2 physician assistants in the final sample (internal medicine 48%, family medicine 34%, obstetrics-gynecology 9%, and pediatrics/adolescent medicine 9%).

Results: Following the intervention, providers were more likely to agree that advance prescriptions should be given, disagree that the number of times EC is dispensed should be restricted and disagree that repeated EC use poses health risks. The proportion of providers who had ever given an advance prescription increased from 18% to 41% ($p=.007$), and there was a trend toward a greater proportion of providers initiating counseling about EC from 36% to 54% ($p=.057$).

Conclusions: A simple educational intervention was associated with a change in primary care provider attitudes and practice patterns regarding EC.

Keywords: Emergency contraception; Contraception; Primary care providers; Educational intervention

1. Introduction

Emergency contraceptive (EC) pills are combined estrogen-progestin pills or progestin-only pills that reduce the risk of pregnancy when taken after unprotected intercourse [1,2]. In May 2004, the Food and Drug Administration (FDA) rejected a levonorgestrel EC product for over-the-counter status. As a prescription-only medication, access to EC will continue to be limited until health care providers are knowledgeable and comfortable with prescribing this method of contraception.

Since EC is more effective when taken as soon as possible after unprotected intercourse [2,3], barriers that impede the timely administration of this method need to be minimized. The health care provider should be able to determine whether the patient is eligible for EC and prescribe it without delay. Since the primary care provider is often a patient's first access to the health care system, all primary care providers should be well informed about EC. Unfortunately, surveys of primary care providers in internal medicine, family medicine and adolescent medicine show that they are not prescribing this form of contraception often [4–10].

In addition to prescribing at the time of need, health care providers can give advance prescriptions for EC. This is the practice of providing prescriptions for EC to women of reproductive age in advance of need. By having a prescription on hand, a woman has more timely access to EC. Several studies have shown that advance provision of EC is effective in increasing the likelihood of its use in the event of unprotected intercourse without compromising regular contraceptive use [11–16]. In an effort to increase the practice of providing advance prescriptions, the American College of Obstetricians and Gynecologists (ACOG) has encouraged its members to routinely provide advance prescriptions for EC during health maintenance visits [17]. However, it is not known how often advance prescribing is currently being practiced. A qualitative study of primary care providers suggests that physician misconceptions about EC lead to reluctance to provide advance prescriptions [18].

This study tests the effect of a very simple intervention that could be easily implemented and reach many health

* Corresponding author. Tel.: +1 717 531 8161; fax: +1 717 531 7726.
E-mail address: cchuang@psu.edu (C.H. Chuang).

0010-7824/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.contraception.2005.04.012

care providers. Health care providers frequently attend lectures for continuing medical education. Although passive group education has had mixed results in changing physician behavior [19,20], we hypothesized that EC may be amenable to such an intervention since providing EC is simple and requires minimal training. This study was designed to evaluate the effect of a single educational intervention about EC on primary care providers.

2. Methods

The Boston Medical Center's Women's Health Unit sent letters to directors of continuing education programs at hospitals and health centers throughout Massachusetts offering a lecture program on EC designed for an audience of primary care providers. A brief description of the study was included in the letter. The lecture program was then scheduled at responding institutions between August and December 2002. The institutional review board at the Boston Medical Center approved this study.

Attendees to the lectures were invited to participate in this study. Primary care providers with prescribing authority (physicians, advance practice nurses and physician assistants) were eligible to participate. A cover letter distributed with the survey described the purpose of the study and stated that informed consent for participating in the study was implied with return of the survey. Surveys were filled out and returned prior to the start of the lecture. Each participant was asked to provide his/her name and mailing address for the follow-up survey, which he/she would receive in 6 months. Respondents who did not describe themselves as primary care providers or who do not have women patients were excluded from the study.

2.1. Preintervention survey

The survey was a 38-item self-administered written questionnaire designed for this study to assess knowledge, attitudes and practice patterns regarding EC. Some questions were adapted from previous studies [5,9]. Personal demographic items assessed age, sex, race/ethnicity and religion. Professional demographics items described type of medical degree (physician, advance practice nurse or physician assistant), year of graduation, specialty and practice location (urban, suburban, rural). Knowledge of EC was assessed with seven multiple-choice items regarding awareness of FDA approval, availability of dedicated products, teratogenicity, mechanism of action, timing of administration and efficacy. Attitudes about EC were measured with five items using a 5-point Likert scale ranging from 1 (*strongly agree*) to 5 (*strongly disagree*). Emergency contraception practice patterns were assessed with 13 items describing frequency of requests, frequency of prescribing and advance prescribing, and frequency of counseling about EC during routine health maintenance visits. Frequency of counseling was described as *always*, *often*, *sometimes*, *seldom* or *never*.

2.2. Intervention

The 45-min presentation on EC was designed for this study. The lecture content included description of EC, recommendations from professional medical organizations, history of EC, review of EC regimens and products, safety and efficacy data, mechanism of action, patient screening and management information, side effects, and studies on advance provision of EC. A pocket-sized card with EC prescribing information, a patient information handout on EC and "Ask me about EC" buttons were distributed at the time of the lecture.

2.3. Postintervention survey

The postintervention survey was mailed to the participants 6 months following the lecture they attended. The postintervention survey was identical to the preintervention survey, excluding the demographic items. Nonrespondents

Table 1
Comparison of providers completing both pre- and postintervention surveys with providers completing preintervention surveys only

Provider characteristic	Preintervention only ($N=80$)	Pre- and postintervention ($N=56$)	p value*
Provider type — no. (%)			
Physician	69 (87)	50 (89)	.83
Advance practice nurse	8 (10)	4 (7)	
Physician assistant	2 (3)	2 (4)	
Specialty — no. (%)			
Internal medicine, primary care	38 (48)	27 (48)	.83
Family medicine	24 (30)	19 (34)	
OB-GYN, primary care	11 (14)	5 (9)	
Pediatrics/adolescent medicine	6 (8)	5 (9)	
Male — no. (%)	27 (34)	22 (39)	.54
Age — mean (S.D.)	41 (12)	39 (11)	.42
Race/ethnicity — no. (%)			
White	57 (72)	42 (75)	
Black/African American	2 (3)	1 (2)	.75
Latino/Hispanic	5 (6)	3 (5)	
Asian	12 (15)	10 (18)	
Other	3 (4)	-	
Religion — no. (%)			
Catholic	22 (29)	15 (27)	.99
Protestant/other Christian	17 (22)	13 (24)	
Jewish	16 (21)	11 (20)	
Muslim	1 (1)	-	
Hindu	5 (6)	3 (5)	
None/atheist	9 (12)	8 (15)	
Other	7 (9)	5 (9)	
Practice location — no. (%)			
Urban	58 (73)	41 (73)	.85
Suburban	19 (24)	12 (21)	
Rural	3 (4)	3 (5)	

* Pre- and postintervention values compared using χ^2 or Fisher's exact test for categorical variables and t test for continuous variables.

Table 2
Provider knowledge of EC pre- and postintervention ($n=56$)

Knowledge question	Preintervention, *N* (%)	Postintervention, *N* (%)	p value
1. Knew that EC is FDA approved	51 (94)	52 (95)	.98
2. Knew that there are dedicated products packaged specifically for EC	48 (89)	50 (91)	.73
3. Knew that if EC fails and pregnancy ensues, there is NO increased risk of teratogenicity	43 (80)	48 (87)	.28
4. Knew that the mechanism of action of EC is disruption of the implanted embryo	37 (69)	40 (73)	.63
5. Knew that the first of the two doses of EC should be taken within 72 h of unprotected intercourse	40 (74)	44 (80)	.46
6. Knew that the second dose of EC should be taken 12 h after the first dose	44 (81)	45 (83)	.96
7. Knew that if properly used, EC reduces the risk of pregnancy by >67%	49 (91)	48 (87)	.75
Mean correct score (out of 7)	5.8+1.2	6.0+1.5	.34*

* Paired *t* test.

received up to three reminder mailings at 2-week intervals after the initial mailing.

2.4. Statistical analysis

The respondents' characteristics are reported using descriptive statistics. χ^2 Tests and *t* test were used to compare the characteristics of providers who only completed the preintervention survey to those who completed both the pre- and postintervention survey. Subsequent analyses were restricted to those providers who completed both surveys, in order to assess the effect of the intervention.

The proportion of providers who correctly answered each knowledge item was compared pre- and postintervention using χ^2 analysis. The overall number of correct responses (out of seven items) was compared pre- and postintervention using the paired *t* test. Attitudes are presented as the proportion of respondents who agree (*strongly agree* and *agree*) with positive statements or disagree (*strongly disagree* or *disagree*) with negative statements about EC. The mean response on the 5-point Likert scale is also presented. Significance testing for changes in attitude scores was performed using the Wilcoxon signed-rank test. The proportion of providers who have ever prescribed EC, who initiate counseling about EC during routine health maintenance visits at least sometimes and who have ever written an advance prescription for EC was compared using χ^2 tests. The number of prescriptions and advance prescriptions given in the past 6 months was analyzed pre- and postintervention using paired *t* tests. All statistical tests were two-tailed using an alpha of .05 as criterion for statistical significance. Statistical analysis was performed using SAS, version 8.0 (SAS Institute, Cary, NC).

3. Results

Lectures were conducted for this study between August and December 2002 at seven hospital grand rounds. Of the 86 preintervention surveys collected, 5 respondents were not primary care providers (1 hospitalist and 4 subspecialists) and 1 respondent did not have women patients in his/her practice, leaving 80 surveys eligible for the preintervention analysis. Four providers did not give contact information, 3 had incorrect addresses and 1 person stated on the survey they were unable to stay for the lecture, leaving 72 providers eligible for the 6-month follow-up mail survey. Of these 72 eligible participants, 56 follow-up surveys were received for a response rate of 78%.

3.1. Characteristics of respondents

There were 50 physicians, 4 advanced practice nurses and 2 physician assistants in the postintervention analysis (Table 1). They described their specialties as primary care internal medicine (48%), family medicine (34%), obstetrics-gynecology (9%), and pediatrics/adolescent medicine (9%). The mean age of respondents was 39 years (S.D. 11), 39% were men, 75% White, 18% Asian, 5% Latino and 2% Black. Providers' religious affiliations were divided between Catholic (27%), Protestant (24%) and Jewish (20%). The majority of providers worked in urban locations (73%). Nonrespondents did not significantly differ from the respondents in these demographic characteristics.

3.2. Emergency contraception knowledge

The provider knowledge data are shown in Table 2. Preintervention knowledge about EC was high. Out of the seven knowledge items, the mean number of correct responses was 5.8 ± 1.2. Following the intervention, the mean number of correct responses remained high at 6.0 ± 1.5.

3.3. Changes in attitudes

The effect of the intervention on attitudes toward EC is shown in Table 3. The proportion of providers agreeing that advance prescriptions should be routinely given increased from 36% before the intervention to 66% after the intervention ($p=.0006$). Following the intervention, more providers disagreed that the number of times EC should be dispensed to a patient should be restricted (64%

Table 3
(a) Provider agreement with positive attitudes about EC pre- and postintervention ($n=56$)

Attitude	Agree (%)		Mean response		p value*
	Preintervention	Postintervention	Preintervention	Postintervention	
Prescriptions for EC should be given in advance of need to patients to have on hand	36	66	2.87	2.25	.0006
EC should be available over-the-counter (without a prescription)	34	39	3.02	2.95	.51

(b) Provider disagreement with negative attitudes about EC pre- and postintervention ($n=56$)

Attitude	Disagree (%)		Mean response		p value*
	Preintervention	Postintervention	Preintervention	Postintervention	
Providing EC encourages inconsistent contraceptive use	73	77	3.84	3.95	.41
The number of times EC is dispensed to an individual patient should be restricted	64	84	3.71	4.00	.04
Repeated EC use over time poses health risks	55	86	3.52	4.09	<.0001

* Wilcoxon rank sum test.

preintervention vs. 84% postintervention, p=.04) and disagreed that repeated EC use over time poses health risks (55% preintervention vs. 86% postintervention, p<.0001). Most of the providers, however, did not agree that EC should be available as an over-the-counter medication. The majority disagreed that EC encourages inconsistent contraceptive use, and this was unchanged following the lecture.

3.4. Changes in practice patterns

The effect of the intervention on provider practice patterns of EC is shown in Table 4. There was no significant change in the proportion of providers who had ever prescribed EC or in the average number of times EC was prescribed for unprotected intercourse in the past 6 months. However, the number of providers who had ever given an advance prescription for EC increased from 18% to 41% (p=.007). The average number of advance prescriptions was unchanged (1.4±4.4 preintervention vs. 1.8±5.1 postintervention). There was also a trend toward an increasing proportion of providers who reported initiating counseling about EC at least sometimes during routine health maintenance visits (36% preintervention vs. 54% postintervention, p=.057).

Table 4
Emergency contraception practice patterns pre- and postintervention ($n=56$)

Practice pattern	Preintervention	Postintervention	p value
Ever prescribed EC	82	86	.61
Number of EC prescriptions in past 6 months	2.9+5.6	3.3+7.3	.24*
Ever provided advance prescription for EC	18	41	.007
Number of advance prescriptions in past 6 months	1.4+4.4	1.8+5.1	.53*
Counsels about EC at least sometimes	36	54	.057

* Paired t test.

4. Discussion

We found that a single educational lecture resulted in a greater proportion of providers who give advance prescriptions for EC, and a trend toward an increasing proportion who routinely counsel about EC. Attitudes about EC were also improved following the program. Baseline knowledge scores were high and remained high following the intervention. Adequate knowledge may therefore not be the main barrier to providing EC. Other barriers, such as attitudes and time constraints, may play a larger role.

While EC remains available by prescription only, we must rely on medical providers to be accessible, knowledgeable and willing to provide EC without delay. There have been a few interventions aimed at improving health care provider use of EC in the United States. From 1996 to1998, a demonstration project was conducted at San Diego County Kaiser Permanente with the aim of changing EC knowledge, attitudes and practices among health care providers. The intervention involved providing repackaged oral contraceptives for use as EC, development of provider and patient information materials, and training of health care providers and staff. The result of this 2-year project was an increase in the proportion of providers who had prescribed EC in the preceding year from 30% to 49% and improvement in providers' knowledge scores of EC [9]. Other efforts toward improving access have been through programs that allow pharmacist provision of EC in Washington, California, Alaska, New Mexico and Hawaii. In these states, a pharmacist may dispense EC to a client without a prescription. Several other states are introducing legislation to have similar pharmacy agreements, but such projects may take years to implement. These types of undertakings require large coordinated efforts and resources that are not currently available in all states.

Our intervention was inexpensive and easy to replicate. Since the American public is largely unaware that there is a contraceptive that can be used after intercourse [6,21,22], efforts by providers to educate their patients are important in an attempt to increase access to this method. Advance provision of prescriptions for EC has been recommended by ACOG [17], and studies suggest that this practice could significantly increase use of EC [11–14], but whether this is being practiced has not been previously described. In our sample of primary care providers, only a minority had ever given an advance prescription for EC, but this increased in the 6 months after the program.

This study has several limitations. There was no control group, so it is not possible to ascertain whether influences outside of the study intervention may have contributed to the findings. However, our study occurred prior to recent press publicizing over-the-counter consideration of EC by the FDA. Our sample size is small and limited to providers mostly from urban areas, which limits our ability to estimate the potential effect of our intervention to a more diverse group of primary care providers. Also, providers choosing to attend a lecture on EC are likely to have more positive attitudes about EC than other providers, and perhaps more willing to change their EC prescribing practices. However, by including lectures into existing grand round schedules, and not developing them as freestanding programs, we maximized the potential to reach a broad audience.

Medical centers, primary care practices, school/college health centers and other primary care groups should invite similar programs on EC. Such efforts to improve access are the first steps toward prevention of unintended pregnancies with this effective form of contraception. Other interventions aimed at health care providers should also be developed and studied in order to continue to improve awareness of EC among primary care providers.

References

[1] Trussell J, Rodriguez G, Ellertson C. Updated estimates of the effectiveness of the Yuzpe regimen of emergency contraception. Contraception 1999;59:147–51.

[2] WHO Task Force on Postovulatory Methods of Fertility Regulation. Randomised controlled trial of levonorgestrel versus the Yuzpe regimen of combined oral contraceptives for emergency contraception. Lancet 1998;352:428–33.

[3] von Hertzen H, Piaggio G, Ding J, et al. Low dose mifepristone and two regimens of levonorgestrel for emergency contraception: a WHO multicentre randomised trial. Lancet 2002;360:1803–10.

[4] Grossman RA, Grossman BD. How frequently is emergency contraception prescribed? Fam Plann Perspect 1994;26:270–1.

[5] Gold MA, Schein A, Coupey SM. Emergency contraception: a national survey of adolescent health experts. Fam Plann Perspect 1997; 29:15–9, 24.

[6] Delbanco SF, Stewart FH, Koenig JD, Parker ML, Hoff T, McIntosh M. Are we making progress with emergency contraception? Recent findings on American adults and health professionals. J Am Med Women's Assoc 1998;53:242–6.

[7] Sills MR, Chamberlain JM, Teach SJ. The associations among pediatricians' knowledge, attitudes, and practices regarding emergency contraception. Pediatrics 2000;105:954–6.

[8] Golden NH, Seigel WM, Fisher M, et al. Emergency contraception: pediatricians' knowledge, attitudes, and opinions. Pediatrics 2001; 107:287–92.

[9] Beckman LJ, Harvey SM, Sherman CA, Petitti DB. Changes in providers' views and practices about emergency contraception with education. Obstet Gynecol 2001;97:942–6.

[10] Chuang CH, Waldman LJ, Freund KM, Ash AS. Emergency contraception: prescribing practices of general internists compared with other primary care physicians. Contraception 2004;69:43–5.

[11] Glasier A, Baird D. The effects of self-administering emergency contraception. N Engl J Med 1998;339:1–4.

[12] Raine T, Harper C, Leon K, Darney P. Emergency contraception: advance provision in a young, high-risk clinic population. Obstet Gynecol 2000;96:1–7.

[13] Ellertson C, Ambardekar S, Hedley A, Coyaji K, Trussell J, Blanchard K. Emergency contraception: randomized comparison of advance provision and information only. Obstet Gynecol 2001;98:570–5.

[14] Jackson RA, Bimla Schwarz E, Freedman L, Darney P. Advance supply of emergency contraception: effect on use and usual contraception — a randomized trial. Obstet Gynecol 2003;102:8–16.

[15] Raine TR, Harper CC, Rocca CH, et al. Direct access to emergency contraception through pharmacies and effect on unintended pregnancy and STIs: a randomized controlled trial. JAMA 2005;293:54–62.

[16] Gold MA, Wolford JE, Smith KA, Parker AM. The effects of advance provision of emergency contraception on adolescent women's sexual and contraceptive behaviors. J Pediatr Adolesc Gynecol 2004;17:87–96.

[17] American College of Obstetricians and Gynecologists. ACOG practice bulletin: emergency oral contraception. Washington (DC): American College of Obstetricians and Gynecologists; 2001.

[18] Karasz A, Kirchen NT, Gold M. The visit before the morning after: barriers to preprescribing emergency contraception. Ann Fam Med 2004;2:345–50.

[19] Davis DA, Thomson MA, Oxman AD, Haynes RB. Evidence for the effectiveness of CME. A review of 50 randomized controlled trials. JAMA 1992;268:1111–7.

[20] Wensing M, Grol R. Single and combined strategies for implementing changes in primary care: a literature review. Int J Qual Health Care 1994;6:115–32.

[21] Delbanco SF, Mauldon J, Smith MD. Little knowledge and limited practice: emergency contraceptive pills, the public, and the obstetrician–gynecologist. Obstet Gynecol 1997;89:1006–11.

[22] Delbanco SF, Parker ML, McIntosh M, Kannel S, Hoff T, Stewart FH. Missed opportunities: teenagers and emergency contraception. Arch Pediatr Adolesc Med 1998;152:727–33.

Mammography Use

Michele M David, MD, MBA, MPH; Linda Ko, MS, MPH; Nicole Prudent, MD, MPH; Eric H. Green MD, MSc; Michael A. Posner, MS; and Karen M Freund, MD, MPH
Boston, Massachusetts

Objectives: The goal of this study was to compare mammography use in Haitian women versus that of other racial/ethnic groups in the same neighborhoods and to identify factors associated with mammography use in subpopulations that are seldom studied.

Methods: A community-based, cross-sectional survey sampled a multiethnic group of inner-city women from eastern Massachusetts. Bivariate analyses and logistic regression models were used to predict lifetime and recent (within two years) mammography screening.

Results: Self-reported lifetime mammography use was similar for Haitian (82%), African-American (78%), Caribbean (81%) and Latina women (86%) but higher for white women (94%, p=0.008). Mammography use in the past two years was also similar in all groups (66–82%, p=0.41). In multivariate models, African-American (adjusted odds ratio [AOR]; 0.3; 95% CI 0.1–0.9) and Haitian women (AOR 0.3; 95% CI 0.1–0.9) had lower odds of lifetime mammography compared to white women. Factors independently related to lifetime and recent mammography included having a regular healthcare provider, greater knowledge of breast cancer screening; higher education, and private health insurance.

Conclusions: Haitian women with a regular provider and knowledge of breast cancer screening reported recent mammography use similar to women from other racial/ethnic groups. The racial/ethnic patterns of mammography use in our study do not explain racial/ethnic differences in breast cancer stage or mortality.

Key words: Haitian ■ breast cancer ■ mammogram ■ ethnicity

 From the Haitian Health Institute at Boston Medical Center and Women's Health Research Unit, Section of General Internal Medicine and Adolescent Medicine, Evans Department of Medicine, Boston University School of Medicine. Send correspondence and reprint requests for *J Natl Med Assoc.* 2005;97:253–261 to: Michele David, MD, MPH, MBA, Address: Boston University Medical Center, Women's Health Unit, 720 Harrison Ave., DOB 1108, Boston, MA 02118; phone: (617) 638 8036; fax: (617) 638 8026; e-mail: midavid@bmc.org

INTRODUCTION

Although Massachusetts leads other states in overall rates of screening mammography for black and white women,[1] no associated reduction in the gap of breast cancer deaths between black and white women has been observed.[2-5] One explanation for this unexpected concurrence is that the group of women defined as black may contain subpopulations that differ in screening behavior, which could account for continued differences in mortality. Subpopulations with a lower education, low income and recent immigration history, such as Haitian women, may have lower screening rates.[6-8]

Haitians in the United States represent a sizable and growing population, consisting of more than 500,000 documented[9] and 1.2 million undocumented immigrants, according to recent estimates.[10] Reliable data on mammography screening of Haitians living in the United States, however, is not available from large population-based surveys. The dearth of this and other data on Haitians, including reliable and valid health information, exists for a number of reasons. Often, Haitians are classified in studies as "black," or "African-American." Moreover, when surveys targeting Haitians are attempted, difficulties owing to factors associated with language, a reluctance to participate due to prior misrepresentation (stigmatization during the early years of the HIV pandemic), immigration history and recent immigration policies are often reported.[10]

There are other reasons to believe that mammography use is low in Haitian populations. In addition to the barriers cited above, some women who originate from rural areas of Haiti may view the role of screening mammography with the health attitudes, beliefs and practices thereof.[8,11-13] For these women, breast cancer screening is likely to be an unfamiliar concept, since only Haitian women of the very highest social stratum have access to screening mammography often obtained in the United States.[14] Preliminary literature suggests that Haitian women do not avail themselves of preventive health testing as compared to women

from other racial/ethnic groups.[8,15,16]

The main goal of our study was to determine whether or not Haitian women utilize mammography at rates similar to women in their communities. We also sought to determine if the predictors of mammography use in this community sample are similar to those already reported in the literature.

METHODS

Subjects

We conducted a community-based, cross-sectional interviewer-administered survey of women 40 years of age and older who spoke English or Haitian Creole in eastern Massachusetts neighborhoods having a high proportion of Haitian residents. We used an area probability sample to select subjects. After constructing a sampling frame by using the City of Boston and Cambridge Assessing Department's Property Parcel Data for fiscal year 1997, we interviewed Haitian community leaders to identify housing blocks with high concentrations of Haitian families. Blocks estimated by two or more informants to contain 10 Haitian households or at least 20% Haitian households were included in the survey universe. In each block, all housing units were enumerated, and 12 units were randomly selected for interview. Interviewers who visited housing units having Haitian residents also asked them to suggest blocks where other Haitian families were likely to reside.

Survey Instrument and Data Collection

We developed a survey instrument that included questions on ethnicity, age and education, as well as selected knowledge, attitudes and belief variables in relation to the utilization of cancer screening. Researchers adapted questions from the National Health Interview Survey (NHIS), the NHIS cancer control supplements and the Cancer Control Needs in Multi-Ethnic Communities (CCNMEC) survey conducted in New York City in 1992.[17-19] We field-tested the questionnaire by administering it to 20 community members. We incorporated their critique in the final questionnaire, hired interviewers from the population under study. The questionnaire was translated into Haitian Creole, back-translated into English by a different translator, compared translations and resolved discrepancies. The Institutional Review Board of the Boston Medical Center

Table 1. Demographic Characteristics by Ethnicity of Community-Based Sample in Eastern Massachusetts (n=326)

	Haitian N=143	White N=80	African-American N=55	Caribbean N=26	Latina N=22	P Value
			N (%)			
Age						
40–49	60 (43)	24 (31)	28 (52)	6 (24)	13 (59)	<0.01
50–59	42 (32)	16 (20)	16 (30)	8 (32)	5 (23)	
60–69	28 (20)	19 (24)	4 (7)	5 (20)	4 (18)	
70+	10 (7)	20 (25)	6 (11)	6 (24)	0	
Education						
<High school	87 (62)	7 (9)	10 (19)	13 (54)	8 (36)	<0.01
Marital Status						
Married	87 (63)	37 (46)	29 (55)	9 (35)	10 (45)	0.03
Employment Status						
Employed	84 (59)	41 (52)	32 (59)	14 (56)	14 (64)	0.81
Insurance Status						
Private	71 (51)	47 (59)	30 (56)	14 (54)	11 (52)	0.13
Public	40 (29)	29 (36)	18 (33)	8 (31)	9 (43)	
None	28 (20)	4 (5)	6 (11)	4 (15)	1 (5)	
Have regular provider	127 (91)	74 (96)	49 (91)	25 (96)	22 (100)	0.31
Age mean (SD)*	53.1 (10.8)	58.8 (13.7)	52.0 (11.0)	59.0 (11.2)	49.9 (8.0)	<0.01

* SD Standard Deviation

approved all aspects of the study.

Four trained bilingual and bicultural female interviewers conducted 25-minute interviews in the residents' homes between June and August 1997, in English- or Haitian-Creole-based on the preference of the subject. Interviewers returned at least four times to attempt to make a contact with an eligible subject. The interviewers read a prepared informed consent form and obtained verbal consent before the interview. At the end of the survey, all participants received educational materials and information on community resources for free mammograms.

STATISTICAL ANALYSIS

Major Analytic Variables

We had two primary dependent variables based on self-reported mammography history. The first, "lifetime," was defined as positive if the respondent reported any mammogram during their lifetime. The second, recent, was defined as positive only if the respondent had a mammogram within the past two years. A definition of mammography was provided to each respondent immediately before asking the questions on utilization.

Independent variables included race/ethnicity, knowledge of breast cancer screening, beliefs about breast cancer (modesty, fatalism, and efficacy), and selected demographics data. In the data analyses, women were classified into five different race/ethnic categories: white (not Hispanic), African-American, Haitian, English-speaking Caribbean (from Barbados, Jamaica, Montserrat), or Latina (from Puerto Rico, Cuba, the Dominican Republic, and Central and South America) based on self-report. Of the 31 (9.4% of sample) who answered multiple ethnicities, race/ethnicity was inferred for 28 subjects from their answers to the questions on race, country of birth and primary language. For example, a woman who described her ethnicity as both African-American and Haitian was born in Haiti and listed Haitian Creole as her primary language, and was classified as Haitian. A series of 13 items with Likert-format responses, along with one question (with a yes or no answer) taken from the CCNMEC and NHIS cancer supplement, was used to address subjects' knowledge about breast cancer screening and beliefs about cancer (Appendix 1).[17,19] Demographic variables included age categorized into deciles (40–49, 50–59, 60–69, 70+) and education dichotomized as less than a high school and as high school or more. Marital status was categorized as "unmarried" (including women who had never married, had separated or had divorced) and "married". Employment status categorized the subjects into "employed" (for women working full or part time) and "unemployed." A participant's insurance status was categorized as "private" (private health insurance), "public," (Medicaid, Medicare and/or coverage under the state of Massachusetts uncompensated care pool) or "none."

Data Analysis

We used t-test and Chi-squared analyses to compare prevalence of independent variables among the racial/ethnic groups and to determine their unadjusted association with lifetime and recent mammography use. We performed multivariate stepwise logistic regressions to predict lifetime and recent mammography use. As race was the primary predictor of interest, it was forced into each of the final models. Odds ratios and 95% confidence intervals were calculated.

RESULTS

Of the 1,103 households randomly selected, we identified 392 eligible women. Among the eligible women, 329 completed the questionnaires (response rate=84%). Forty-three percent (n=144) of the interviewed respondents described their ethnicity as Haitian, 24% (n=80) white, 17% (n=56) African-American, 8% (n=26) English-speaking Caribbean, 7% (n=22) as Latina, and the remainder (n=3, <2%) as other ethnic identity. For the purpose of analysis, we excluded women in this last category, leaving 326 for analysis.

Table 1 shows the demographic characteristics of the 326 subjects. Caribbean and white women were significantly older than other groups of women (two sample t-tests, $p<0.001$). Haitian respondents (62%) were less likely to complete high school than all others interviewed (two sample t-tests, $p=0.001$). More Haitian women (63%) but fewer Caribbean women (35%) were married ($p=0.03$). Fifty-seven percent of the respondents were employed on either a full- or part-time basis, and employment status did not vary by ethnicity ($p=0.81$). Sixty-one percent had private health insurance. Health insurance status did not differ by ethnicity ($p=0.13$). Most women (93%) reported that they had a regular healthcare provider ($p=0.31$). Haitian women immigrants spent less time in the United States compared to other immigrant women (median length of stay 14 years compared to 17 years for English-speaking Caribbean and 22 years for Latina). Two Haitian women, none of the Latina and none of the English-speaking Caribbean women were born in the United States.

The prevalence of ever having had a mammogram among respondents was high (84%). Haitian women (82%) reported "ever" mammography use similar to that of African-American (78%), Caribbean (81%) and Latina women (86%) (Table 2). There was a non-significant trend of white women having a higher rate

of ever having had a mammogram (94%, p=0.10). The overall prevalence of having had a mammogram in the past two years was also high (69%) and similar across all five ethnic groups (p=0.41).

Women with a regular healthcare provider reported a high prevalence of both ever (87% vs. 64%; p=0.004) and recent mammography (73% vs. 37%; p=0.001) compared to women without a regular

Table 2. Unadjusted Prevalence of Ever Had a Mammogram and Mammogram in the Past Two Years by Ethnicity and Selected Demographic Variables

Variables	Ever Had a Mammogram N (%)	[P Value]	Had a Mammogram in the Past Two Years N (%)	[P Value]
Ethnicity				
Haitian (N=143)	118 (82%)	[<0.10]	96 (67%)	[0.41]
White (N=80)	75 (94%)		53 (66 %)	
African-American (N=55)	43 (78%)		39 (71 %)	
Caribbean (N=26)	21 (81%)		21 (81 %)	
Latina (N=22)	19 (86%)		18 (82 %)	
Age Category				
40-49	114 (80%)	[0.44]	97 (68%)	[0.43]
50-59	80 (90%)		69 (76%)	
60-69	41 (83%)		31 (67%)	
70+	35 (90%)		26 (69%)	
Education				
<High school	108 (86%)	[0.48]	89 (71%)	[0.62]
≥High school	162 (84%)		133 (69%)	
Marital Status				
Married	149 (87%)	[0.47]	121 (70%)	[0.81]
Not married	124 (84%)		104 (70%)	
Employment Status				
Employed	161 (87%)	[0.26]	137 (74%)	[0.08]
Unemployed	113 (83%)		89 (65%)	
Insurance Status				
Private	150 (87%)	[0.61]	130 (75%)	[0.12]
Public	87 (84%)		68 (65%)	
None	35 (81%)		27 (63%)	
Have Regular Provider				
Yes	257 (87%)	[<0.01]	217 (73%)	[<0.01]
No	14 (64%)		8 (37%)	
Knowledge about Breast Cancer				
More	226 (88%)	[<0.01]	186 (72%)	[0.04]
Less	50 (72%)		41 (59%)	
Breast Modesty				
More	19 (79%)	[0.44]	13 (54%)	[0.09]
Less	250 (85%)		208 (71%)	
Fatalism on Cancer				
More	39 (76%)	[0.08]	32 (63%)	[0.24]
Less	237 (86%)		195 (71%)	
Efficacy of Preventive Medicine				
More	230 (83%)	[0.12]	190 (69%)	[0.47]
Less	48 (92%)		37 (74%)	

provider. Women with a greater knowledge about breast cancer were more likely to have ever had a mammogram (88% vs. 72%; p=0.002) and to have had one in the past two years (72% vs. 59%; p=0.04), compared to those with less knowledge. Insurance, age, education, marital and employment status were not related to ever or recent mammography in bivariate analyses. Scales measuring modesty, fatalism and efficacy were also not statistically associated with mammography use.

Significant differences by ethnicity were found in women's beliefs concerning cancer prevention and treatment. On a 0–5 scale, with highest scores measuring greater fatalistic attitudes, Caribbean (1.52) and Haitian (1.51) women showed more fatalistic attitudes towards cancer prevention and treatment than white (0.82) and African-American women (0.87) (overall Chi-square p=0.0005). No significant ethnic differences were found on breast modesty (p=0.81) or efficacy of mammogram screening (p=0.90). However, none of these belief measures were associated with ever or recent mammography uses.

The unadjusted odds ratios are presented in Table 3. In multivariate logistic regression, where ethnicity was forced into the model, Haitian and African-American women had a 70% less odds of ever having had a mammogram (OR=0.3; 95% CI, 0.1–0.9) as compared to white women. However, ethnicity was not a predictor of having had a mammography in the past two years (Table 4). Having a regular provider, knowledge about breast cancer, education and private insurance were associated with mammography use (Table 4). Women with a regular provider had a 3.4 times increased odds of ever having had a mammogram, compared to those with no regular provider (95% CI, 1.1–10.1), and 3.4 times increased odds of having had a mammogram in the past two years (95% CI, 1.3–9.1). Women with knowledge about breast cancer had increased odds of ever having a mammogram (OR 2.8; 95% CI 1.3–6.1), and having had one in the past two years (OR 1.9; 95% CI 1.0–3.4). Education remained a significant predictor of ever having had a mammogram: subjects with high-school education or greater had an odds ratio of 3.4 for having ever had a mammogram, as compared to women with a less-than-high-school education (95% CI, 1.5–7.9). Education was not a significant predictor of having had a mammogram in the past two years. Women with private insurance were 2.4 times more likely to ever have had a mammogram (95% CI, 1.2–5.0), and 1.7 more likely to have had a mammogram in the past two years (95% CI, 1.1–3.0) than those with public insurance.

DISCUSSION

This is the first study, to our knowledge, which examines mammogram use within a community-based sample of Haitian women and compares it to that of women of other ethnic groups in the same neighborhoods. We found that overall rates of ever having had a mammogram and having had a mammogram in the past two years were high in these communities, at 84% and 69%, respectively. We found that white women were more likely to have unadjusted prevalence of lifetime mammography compared to other racial/ethnic groups but did not find a difference in crude recent mammography prevalence across race/ethnicity. In the multivariate analysis, we found that Haitian and African-American women were less likely ever to have had mammography as compared to white women. For all ethnic groups, having a regular healthcare provider, more knowledge about breast cancer, greater education and private health insurance significantly increased the odds of mammography use. In this population of women, age, marital status, breast modesty, fatalistic beliefs on cancer and positive attitude toward cancer prevention and treatment were not independently associated with mammogram use.

Unlike other population-based studies,[15,16] Haitian women from our study reported a high prevalence of mammogram use. O'Malley and colleagues showed that Haitian women living in New York had a lower prevalence of mammogram use compared to other U.S.-born blacks and English-speaking Caribbeans.[15] Our study found some difference in those who had ever had a mammogram but no difference in recent

Table 3. Unadjusted Odds Ratios of Ever Had a Mammogram and Had a Mammogram in the Past Two Years*

	Ever Had a Mammogram OR (95% CI)	Had a Mammogram in the Past Two Years OR (95% CI)
Ethnicity		
Haitian	0.23 (0.075–0.69)	1.18 (0.56–2.49
African-American	0.25 (0.66–0.97)	1.94 (0.65–5.75)
Caribbean	0.32 (0.12–0.88)	1.08 (0.6–1.93)
Latina	0.42 (0.093–1.93	2.29 (0.71–7.44)
White	1.0	1.0

The unadjusted odds ratio was calculated using bivariate logistic regression (i.e. with only ethnicity in the model.

use across the five different racial/ethnic groups.

The overall high prevalence observed in the use of mammography among Haitian women in our study might reflect the combined efforts of public health agencies in Massachusetts (the Massachusetts and the Boston Departments of Public Health, major teaching hospitals and community health centers).[20,21] Those agencies targeted neighborhoods in which the women in our study lived, providing intensive outreach and education, as well as free screening services, using a culturally competent approach.[22,23] In addition, many language-specific (Haitian Creole) health promotion radio and television programs regularly targeted Haitian women in the greater Boston area to encourage the use of preventive services. Further, in 1996, the Massachusetts Division of Medical Assistance developed quality measures for interpreter service through the Acute Hospital Request for Application process. As a result, hospitals establish standards of practice to ensure Massachusetts Medicaid subscribers have access to trained medical interpreters at all key points of contact throughout the hospital, including outpatient clinics.[24,25]

The high prevalence of mammography use in our study supports the evidence that the proportion of women getting mammography has substantially increased in the last several years.[2,3,17,26,27] The age-adjusted percent of women who reported having had a mammogram in the past two years in the United States has increased to 76%.[1] In Massachusetts, there is also an increase in mammography use. The Behavioral Risk Factor Surveillance System in Massachusetts shows that the total age-adjusted proportion of women aged ≥40 years who reported having had a mammogram in the past two years increases from 66% in 1992 to 84% in 2000.[1]

As in other investigations of majority populations,[26] our investigation found that women with regular healthcare providers have a higher prevalence of mammogram use. Our study also supports findings from the literature that show an association between mammography use and education.[28] We also found that knowledge about breast cancer screening is associated with mammogram use. Although many studies probed the knowledge variable through specific questions rather than a series of questions summed in a scale, their findings corroborate ours that knowledge about breast cancer screening is significantly related to mammogram use.[10]

Our study has several strengths—chief among them was our response rate. The Haitian community is hard to reach due to language barriers, cultural barriers and, for the large number of undocumented Haitians, fear of both public institutions and deportation. The threat and/or perception of the latter posed by the 1996 immigration laws has greatly complicated efforts to ensure cooperation with public health initiatives and surveys

Table 4. Adjusted Odds Ratios of Ever Had a Mammogram and Had a Mammogram in the Past Two Years*

	Ever Had a Mammogram Adjusted OR (95% CI)	Had a Mammogram in the Past Two Years Adjusted OR (95% CI)
Ethnicity		
Haitian	0.3 (0.1-0.9)	1.3 (0.7 – 2.6)
African- American	0.3 (0.1-0.9)	1.1 (0.5 – 2.7)
Caribbean	0.2 (0.1-1.0)	2.2 (0.7 – 6.9)
Latina	0.3 (0.1-1.4)	2.7 (0.7– 10.2)
White	1.0 (reference)	1.0 (reference)
Have a Regular Provider		
Yes	3.4 (1.1 – 10.1)	3.4 (1.3 – 9.1)
Knowledge about Breast Cancer		
More	2.8 (1.3 – 6.1)	1.9 (1.0 – 3.4)
Education		
High school	3.4(1.5 –7.9)	
Insurance		
None	2.2 (0.3 – 19.7)	4.6 (0.5 – 39.4)
Private	2.4 (1.1 – 5.0)	1.7 (1.1 – 3.0)
Public	1.0 (reference)	1.0 (reference)
C-statistic	0.74	0.66

Odds ratios were calculated by stepwise logistic regression. Ethnicity was forced in the model. Age, education, marital status, fatalistic attitude on cancer, breast modesty, and efficacy, length of stay in the US were rejected by in the multivariate model.

in Haitian and other immigrant neighborhoods, particularly since political instability in Haiti and erratic immigration policies in the United States have frequently led to chaotic patterns of family immigration.[8] Those who do immigrate are often hindered by nontransferable employment skills, economic difficulties secondary to relocation in a new culture and disenfranchisement stemming from limited English and inexperience with psychological strategies of importance in negotiating an American urban environment.[10] We achieved this high response rate (84%) by involving community members in the data collection process, which increased our access to this hard-to-reach population. By using a collaborative paradigm, we were able to increase trust between the community and researchers and thereby increase knowledge by providing culturally appropriate educational materials on breast cancer, as well as access to free mammograms.

Other strengths of our study include our community-based sampling, which allowed us to achieve a random sampling and minimize the potential for selection bias. Further, as all respondents came from the same neighborhoods, the sampling scheme permitted some leveling of socioeconomic status among our subjects across the racial/ethnic groups. Finally, we were able to obtain and compare data from members of five different ethnic groups, white (non-Hispanic), Haitian, African-American, English-speaking Caribbean, and Latina. Having information of multiple ethnic groups avoids a comparison of the prevalence of mammography use only between black and white women, an approach that combines women from diverse ethnic backgrounds into single categories.

Study limitations include the reliance on self-report for obtaining information on mammogram use. Since respondents received care from a wide variety of healthcare settings in the Greater Boston area, a validation of their self-reports on mammogram use through a review of their medical record was not feasible. However, the questions used to assess mammography use are standardized self-report questions and, therefore, comparable to others published in the literature.[29] The distinctiveness of healthcare services in Massachusetts may pose another limitation to generalizing from our findings. The use of mammography among Haitian populations in other areas may be

Appendix 1. Questions and Statements Included on the Knowledge and Beliefs Scales

Items	**Statement**
Knowledge of breast cancer screening	Have you heard of mammogram
	Since I do breast self examination, I don't need a mammogram
	You only need a mammogram if you have symptoms
	You only need a mammogram if a family member had breast cancer
Beliefs	
a) Modesty	I find that having my breast examined by a male doctor is embarrassing
	I find that having my breast examined by a female doctor is embarrassing
b) Fatalism	If doctors find cancer, there's nothing they can do anyway
	If I had cancer, I had rather not know about it
	Getting cancer is a death sentence for most people
	Since no-one knows what causes cancer, there's really nothing you can do to prevent it
	If someone gets cancer, it's just their fate, there's nothing that can be done about it
c) Faith in Medicine	If more people would get check ups regularly, there would be fewer deaths from cancer
	People can reduce their chance of getting cancer by leading a healthy life
	I think they will find a cure for cancer

greatly impacted by the presence or absence of the kind of health outreach efforts that have been undertaken in Massachusetts. Another limitation is the time since the survey was conducted. Overall screening rates for black and white women have increased.[30] Further work is needed to determine if Haitian women have continued to benefit from public health efforts to promote mammography screening.

Trends in screening mammography are increasing in most populations. Our results, which indicate that this increase includes Haitian and other black women, are encouraging. However, breast cancer mortality among black women in the United States has not decreased, in contrast to the decrease in breast cancer mortality observed among white women. Our findings indicate that the use of mammography may not explain the observed ethnic differences in mortality. One possible explanation is that it is too soon yet to observe the effects of increased breast cancer screening practices on breast cancer mortality in black women. Alternatively, black women may not be following up for abnormal breast exam in the same rate as white women. To address these issues, future work should investigate not only mammogram use in this population but also ethnic differences in follow-up care after abnormal findings on mammography.

ACKNOWLEDGEMENTS

This study was made possible by a grant from the Massachusetts Department of Public Health Breast Cancer Research Grants Program (#197101) and the Boston Medical Center/Boston University Interdisciplinary Women Health's Research Scholar's Program (BIRCWH # K12 HDO43444-02), and there are no known financial conflicts of interest. The authors are grateful to Bart Laws for his help in developing the survey instrument We would also like to thank the Haitian Consulate, the Haitian Public Health Initiatives Inc., the Haitian Multicultural Service Center and other members of the Haitian community for their help with this study. Our statistician, Michael Posner, is working at the Data Coordinating Center of Boston University as a statistical analyst. I also want to thank JoAnn Fordham and Mathilda Ruwe for their editorial assistance. Manuscript preparation was funded by the Boston Medical Center/Boston University Interdisciplinary Women Health's Research Scholar's Program (BIRCWH).

REFERENCES

1. 2003 Women's Health in Massachusetts: Highlights from the Massachusetts Behavioral Risk Factor Surveillance System (BRFSS): 2000–2001, January 2003 Health Survey Program, Bureau of Health Statistics, Research and Evaluation, Massachusetts Department of Health, http://www.state.ma.us/dph/bhsre/cdsp/brfss/brfss.htm. Accessed 10/27/03.

2. Anonymous. From the Centers for Disease Control and Prevention. Recent trends in mortality rates for four major cancers, by sex and race/ethnicity—United States, 1990–1998. *JAMA.* 2002;287:1391-1392.

3. Anonymous. Recent trends in mortality rates for four major cancers by sex and race/ethnicity—United States, 1990–1998. *MMWR Morb Mortal Wkly Rep.* 2002;51:49-53.

4. 2002, The Burden of Heart Disease, Stroke, Cancer and Diabetes, United States, http://www.cdc.gov/nccdphp/burdenbook2002/02_breastcancer.htm. Accessed 10/30/03.

5. Age-adjusted SEER incidence and U.S. mortality rates, 1996–2000, http://seer.cancer.gov. Accessed 10/27/03.

6. Eshiett MU, Parry EH. Migrants and health: a cultural dilemma. *Clin Med.* 2003;3:229-231.

7. Hayward R, Shapiro M, Freeman H, et al. Who gets screened for cervical and breast cancer? Results from a new national survey. *Arch Intern Med.* 1988;148.

8. Laguerre M. Haitian Americans, Ethnicity and Medical care. Cambridge, MA: Harvard University Press; 1981.

9. 2003, Census 2000;QT-P13. Ancestry: 2000, http://factfinder.census.gov/servlet/QTTable?ds_name=D&geo_id=D&qr_name=DEC_2000_SF3_U_QTP13&_lang=en. Accessed 10/27/03.

10. Stepick A, Stepick C. People in the Shadows: Survey Research among Haitians in Miami. *Human Organization.* 1990;49:64-77.

11. Farmer P. Sending sickness: sorcery, politics and changing concepts of AIDS in rural Haiti. *Med Anthropol Q.* 1990;4:6-27.

12. Laguerre M. Afro-Caribbean folk medicine. South Hadley, MA: Bergin & Garvey; 1987.

13. Coreil J, Marshall P. Locus of control: a cross-cultural study. *Human Organization.* 1982;41:133-138.

14. Ministere De La Sante Publique et de la Population H, Sante OPDL. *Analyse de la Situation Sanitaire Haiti.* 1998.

15. O'Malley AS, Mandelblatt J, Gold K, et al. Continuity of care and the use of breast and cervical cancer screening services in a multiethnic community. *Arch Intern Med.* 1997;157:1462-1470.

16. Mandelblatt JS, Gold K, O'Malley AS, et al. Breast and cervix cancer screening among multiethnic women: role of age, health and source of care. *Prev Med.* 1999;28:418-425.

17. Laws M, Mayo S. The Latina Breast Cancer Control Study, year one: factors predicting screening mammography utilization by urban Latina women in Massachusetts. *J Community Health.* 1998;23:251-267.

18. Hazuda H, Stern M, Haffner S. Acculturation and assimilation among Mexican Americans: scales and population-based data. *Soc Sci Quarterly.* 1988;69:687-706.

19. Benson V, Marano M. Appendix III, Questionnaires and flashcards. In current estimates from the National Health Interview survey. National Center for Health Statistics. *Vital Health Stat.* 1994;10.

20. Massachusetts Health Policy Forum. The Uncompensated Care Pool: Saving the Safety Net. 10/23/02 . Report No. 16.

21. Ruthardt L. Uncompensated Care Pool: PFY 2001 Annual Report. 2002 August (Massachusetts Division of Health Care Finance and Policy).

22. 2003, Women's Health Network, http://www.state.ma.us/dph/whn/whn.htm. Accessed 10/27/03.

23. Bigby J, Ko LK, Johnson N, et al. A community approach to addressing excess breast and cervical cancer mortality among women of African descent in Boston. *Public Health Rep.* 2003;118:338-347.

24. 2003, New interpreter services law for hospital ER's takes effect July 1, 2001, http://www.state.ma.us/dph/media/2001/pr0626.htm. Accessed 10/27/03.

25. Interpreter Service Requirement. CMR 2001;105:130-1100.

26. Anderson L, May D. Has the use of cervical, breast and colorectal cancer screening increased in the United States? *Am J Public Health.* 1995;85.

27. Blackman D, Bennett E, Miller D. Trends in Self-Report Use of Mammograms (1987–1997) and Pap (1991–1997)—Behavioral Risk Factor Surveillance System. *MMWR Morb Mortal Wkly Rep.* 1999; 48(SS06):1-22.

28. Zapka J, Stoddard A, Costanza M, et al. Breast Cancer Screening by Mammography: Utilization and Associated Factors. *Am J Public Health.*

1989;79:1499-1502.

29. Gordon N, Hiatt R, Hatcher R, et al. Concordance of self-reported data and medical record audit for six cancer screening procedures. *Journal of National Cancer Institute.* 1993; 85:566-570.

30. American Cancer Society, Cancer Prevention and Early Detection Facts and Figures 2003, http://www.cancer.org/downloads/STT/CPED2003PWSecured.pdf. Accessed 09/20/04. ■

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 ktaylor@nmanet.org.

Impact of U.S. Citizenship Status on Cancer Screening Among Immigrant Women

Israel De Alba, MD, MPH,[1,2] F. Allan Hubbell, MD, MSPH,[1,2] Juliet M. McMullin, PhD,[1,2,3] Jamie M. Sweningson, MPH,[2] Richard Saitz, MD, MPH[4,5]

[1]Division of General Internal Medicine and Primary Care, Department of Medicine, [2]Center for Health Policy and Research, and [3]Department of Anthropology, School of Social Ecology, University of California, Irvine, CA, USA; [4]Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA; [5]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA.

OBJECTIVES: We evaluated the relationship between U.S. citizenship status and the receipt of Pap smears and mammograms among immigrant women in California.

DESIGN: Cross-sectional study using data from the 2001 California Health Interview Survey.

PATIENTS/PARTICIPANTS: Noninstitutionalized, civilian women, aged 18 years and older living in California.

MEASUREMENTS AND MAIN RESULTS: We analyzed data from the 2001 California Health Interview Survey and used logistic regression models to adjust for sociodemographic factors and for access and utilization of health services. After adjusting we found that U.S. citizen immigrants were significantly more likely to report receiving a Pap smear ever (adjusted prevalence ratio [aPR], 1.05; 95% confidence interval [CI], 1.01 to 1.08), a recent Pap smear (aPR, 1.07; 95% CI, 1.03 to 1.11), a mammogram ever (aPR, 1.17; 95% CI, 1.12 to 1.21), and a recent mammogram (aPR, 1.38; 95% CI, 1.26 to 1.49) as compared to immigrants who are not U.S. citizens. Also associated with receiving cancer screening were income, having a usual source of care, and having health insurance. Hispanic women were more likely to receive Pap smears as compared to whites and Asians.

CONCLUSIONS: Not being a U.S. citizen is a barrier to receiving cervical and breast cancer screening. Additional research is needed to explore causal factors for differences in cancer screening rates between citizens and noncitizens.

KEYWORDS: citizenship; immigrant; Pap smear; mammogram; women.
DOI: 10.1111/j.1525-1497.2005.40158.x
J GEN INTERN MED 2005; 20:290–296.

Despite a remarkable increase in cervical Papanicolaou (Pap) smear and mammogram use rates in the past decade in the United States,[1] the benefits of cancer screening are not reaching all women. Cervical cancer continues to disproportionately strike low-income, immigrant, and minority women.[2–4] According to data from the National Health Interview Survey (NHIS), only 61% of recent immigrants reported having a Pap smear in the past 3 years as compared to 83% of women born in the United States.[1] From 1985 to 1996, cervical cancer mortality rates increased among foreign-born women in the United States to such a great extent that it contributed substantially to and influenced overall U.S. cervical cancer mortality trends.[5] Similarly, although breast cancer is more common among white women,[6] immigrants are less likely than nonimmigrants to report a mammogram in the past 2 years[1] and to be diagnosed with early stage disease.[7,8]

Numerous barriers to cancer screening have been identified; sociodemographic and health access factors such as older age, low income or educational level, and lack of health insurance or regular source of health care have been extensively documented.[3,9–19] Barriers related to culture, knowledge, and attitudes such as acculturation, fatalism, and low English proficiency also play an important role.[17,20–26] Immigrants who are not U.S. citizens may be disproportionately affected by these barriers and may face additional challenges to access and receive appropriate health care as compared to immigrants who have become U.S. citizens.[3,27–29] In previous studies, citizenship status has been shown to independently affect access to health insurance[27,30] and receipt of medications for diabetes mellitus and hypertension[31] and referrals to mental health services.[31] Previous reports suggest a potential effect of citizenship status on receipt of cancer screening[32,33]; however, one presents only unadjusted data on citizenship status and Pap smear use,[33] and the other is a preliminary report.[32] Furthermore, both of these studies are limited to a single urban location and to one racial/ethnic group.

Immigrants who are not U.S. citizens constitute an important and rapidly growing segment of U.S. and California populations. Currently, close to 18 million non-U.S. citizens live in the United States[29] and only in the last decade their numbers have increased by over 50%.[29] While the number of immigrants who became U.S. citizens has increased by 71% in the last three decades, the number of immigrants who remained non-U.S. citizens has increased by 400% during the same period.[29] California has a higher percentage and number of noncitizens than any other state in the country; approximately 5.5 million noncitizens live in the state.[29] Therefore, identifying factors than prevent immigrant women from receiving appropriate cancer screening continues to be an important public health goal. In this report, we analyze data from the California Health Interview Survey (CHIS) to assess the impact of citizenship status on the receipt of Pap smears and mammograms among immigrant women in California.

METHODS

Data Source and Study Population

We analyzed data from the 2001 CHIS. The CHIS is a telephone survey of the state of California civilian, noninstitutionalized population. The CHIS is a two-stage, geographically stratified random-digit-dial sample conducted for the first time in 2001. Personnel from CHIS interviewed one randomly selected adult in each of the 55,000 households sampled in the state. Major

Accepted for publication July 1, 2004

None of authors affiliated with this manuscript had any conflicts of interest related to the research conducted for this paper.

Address correspondence and requests for reprints to Dr. De Alba: 111 Academy Way, Suite 220, Irvine, CA 92697-5800 (e-mail idealba@uci.edu).

content areas for the survey included health-related behaviors, health status and conditions, health insurance coverage, and access to health care services. The interviews were conducted in 6 languages: English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, Korean, and Khmer (Cambodian). The overall response rate for the 2001 CHIS adult survey was 37.7%.[34]

We examined the impact of citizenship status on the receipt of Pap smears and mammograms among immigrant women living in California. Immigrants were defined as individuals born outside the United States or its territories who currently live in this country. For our analyses on Pap smear use, we included women age 18 or older without a hysterectomy. Women with history of cervical cancer were not excluded because they would need continued screening.[35] For our analyses on mammogram use, we included women age 40 or older. Although there were inconsistencies in the guidelines regarding screening mammograms for women age 40 to 49 at the time the data were collected, citizens and noncitizens in this age group should receive mammograms in the same proportion.

There are considerable cultural differences between immigrant women and those born in the United States. To decrease the confounding effect of cultural factors, we limited our analysis of the impact of citizenship status on Pap smear and mammogram use to immigrants. Therefore, we compared use of Pap smears and mammograms between those immigrants who became U.S. citizens, know as naturalized citizens, and those who remained as noncitizens. Noncitizens are a heterogeneous group and include legal permanent residents, refugees, asylees, undocumented immigrants, and others.

Variables Assessed

The outcomes of interest were receiving a Pap smear in the past 3 years, a Pap smear ever, a mammogram in the past 2 years, and a mammogram ever according to generally accepted preventive guidelines.[35,36] We developed a multivariable logistic regression model for each outcome of interest. Self-reported citizenship status (U.S. citizen or non-U.S. citizen) was the main independent variable.

We considered the following to be potential confounders in all models: age (continuous variable), race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, Asian, and other race), education attainment (<high school, high school, or >high school), annual household income (<200% or ≥200% the federal poverty level), having health insurance (any coverage or no insurance), having a usual source of health care (yes or no), years in the United States (<10 or ≥10 years), and self-reported health status (excellent, very good, or good vs fair or poor). We also included English language proficiency as a controlling variable and it included 2 categories: high (speak English "very well" or "well") and low (speak English "not well" or "not at all"). The Institutional Review Board at the University of California, Irvine approved this research project as exempt of review, because it involves use of publicly available data without personal identifiers.

Statistical Analysis

All analyses were performed with SAS Callable SUDAAN (Version 7.5.6 for Windows; Research Triangle Park, NC) to account for the CHIS's complex sampling design and to obtain proper variance estimations. For the data analysis, we first generated descriptive statistics for each study variable. To characterize factors associated with the outcomes of interest, we then conducted a bivariable analysis using χ^2 tests to compare categorical variables and t tests for continuous variables. Two-tailed P values less than or equal to .05 were considered statistically significant. To assess collinearity, we estimated Pearson correlation coefficients between all pairs of variables; a coefficient score ≥ 0.7 defined collinearity. We also assessed multicollinearity, which we defined as tolerance test scores <0.1 and variance inflation factors of >2. We found no collinearity among independent variables included in the models.

To determine the impact of citizenship status on use of Pap smears and mammograms, we developed the 4 multivariable logistic regression models described, one for each outcome of interest, adjusting for all confounders described previously. Because the outcome of interest is prevalent in our population, the odds ratios may magnify or overstate the risk association. Therefore, in order to produce accurate approximations of the risk ratios, we transformed odds ratios into prevalence ratios following standard procedures;[37] we present the results in both odds ratios and prevalence ratios. The adjusted prevalence of receiving a screening test was compared between naturalized citizens and noncitizen immigrants. We used the Wald F statistic to compare the levels of explanatory variables.

RESULTS

A total of 6,320 women were included in our analysis of cervical cancer screening and 3,828 in the analysis of breast cancer screening; 47% and 65% were naturalized citizens, respectively. Naturalized citizens in both samples were older and more likely to report having health insurance, a usual source of care, more than high school education, and an annual income of 200% of the federal poverty level or higher as compared to noncitizen immigrants (Table 1). As expected, naturalized citizens were also more likely to report speaking English well and 93% in the cervical cancer sample and 97% in the breast cancer sample had been in the United States for 10 years or more as compared to 55% and 75%, respectively, of noncitizens.

Most naturalized citizens were Hispanic or Asian, reflecting the large number of participants from these groups. However, the majority of Hispanics in our sample were noncitizens (Table 1). White women represented only 20% of the naturalized citizens and 9% of the noncitizens in the Pap smear use sample and 26% and 14%, respectively, in the mammogram use sample.

Overall, 82.7% of immigrants reported a Pap smear in the past 3 years and 88.1% a Pap smear ever; 55% reported a mammogram in the past 2 years and 84% a mammogram ever. Naturalized citizens were more likely to report having Pap smears and mammograms recently or ever.

In the multivariable logistic regression models, after adjusting for potential confounders, naturalized citizens were more likely to report receiving a Pap smear in the past 3 years (adjusted prevalence ratio [aPR], 1.07; 95% confidence interval [CI], 1.03 to 1.11) and ever (aPR, 1.05; 95% CI, 1.01 to 1.08) as compared to noncitizens (Table 2). Other factors associated with having a Pap smear ever or Pap smear in the past 3 years

Table 1. Characteristics of Female Immigrant Respondents of the 2001 CHIS Eligible for Pap Smear and Mammogram Testing

	Analysis of Pap Smear Use of Women Without Hysterectomy (N=6,320)			Analysis of Mammogram Use of Women Age 40 or Older (N=3,828)		
Characteristics, %	Naturalized Citizens (n=2,976) 43.1%	Noncitizens (n=3,344) 56.9%	P Value	Naturalized Citizens (n=2,472) 62.7%	Noncitizens (n=1,356) 37.3%	P Value
Had a Pap smear in past 3 years	85.5	80.6	<.001	–	–	–
Ever had a Pap smear	91.3	85.7	<.001	–	–	–
Had a mammogram in past 2 years	–	–	–	58.8	48.4	<.001
Ever had a mammogram	–	–	–	90.2	73.5	<.001
Race/ethnicity			<.001			<.001
White	20.1	9.1		26.3	13.9	
Hispanic	38.8	69.6		36.5	60.7	
Asian	37.9	19.5		34.6	23.4	
African American	0.8	0.8		0.8	0.6	
Other	2.6	1.1		1.7	1.4	
Years in the U.S.			<.001			<.001
<10	6.1	44.8		2.9	25.0	
>10	93.9	55.2		97.1	75.0	
Age, y			<.001			<.001
18–29	16.9	35.1		–	–	
30–39	26.0	36.0		–	–	
40–49	25.8	17.2		39.7	56.0	
50–64	21.0	8.7		37.0	31.8	
65+	10.3	3.1		23.3	12.3	
How well English is spoken			<.001			<.001
Not well	25.2	63.2		31.2	67.8	
Well	74.8	36.8		68.8	32.2	
Health status			<.001			<.001
Fair to poor	20.8	30.2		28.5	42.9	
Excellent to good	79.3	69.8		71.5	57.2	
Currently insured	84.9	60.2	<.001	87.9	66.5	<.001
Has a usual source of care	90.6	76.1	<.001	93.3	81.5	<.001
Federal poverty level			<.001			<.001
<199%	39.7	70.1		42.7	66.5	
>200%	60.3	29.9		57.3	33.5	
Education level			<.001			<.001
Less than high school	19.0	48.1		24.6	52.1	
High school diploma	22.5	21.2		21.2	18.3	
More than high school	58.5	30.8		54.2	29.6	

Percentages are weighted to yield California population estimates.

were Hispanic ethnicity, having a usual source of care, higher income, and having health insurance. Those in the United States for 10 years or more were more likely to report ever having a Pap smear, as were individuals over the age of 30. Compared to the youngest age group (18–29 years), women between 30 and 64 years old were most likely to get a recent Pap smear and women age 65 and older were the least likely (Table 2). As compared to white women, Hispanics were significantly more likely to report having a Pap smear ever or in the past 3 years and Asians were less likely to report any of these outcomes as compared to any other racial/ethnic group.

Naturalized citizens were more likely to report a mammogram ever (aPR, 1.17; 95% CI, 1.12 to 1.21) or in the past 2 years (aPR, 1.38; 95% CI, 1.26 to 1.49) as compared to noncitizens (Table 3). Other factors significantly associated with both of these outcomes were having a usual source of care, having health insurance, being in the United States over 10 years, and women in the 50–64 years of age group. As compared to white women, Asians were less likely to report a mammogram in the past 2 years or ever. There was no difference between white women and Hispanics. Being in the United States for 10 years or more was significantly associated with having a mammogram ever but not with having one in the past 2 years.

DISCUSSION

We found that not being a U.S. citizen is a barrier to receiving cervical and breast cancer screening among immigrants in California. Even after adjusting for sociodemographics, access to health care, English proficiency, and years in the United States, noncitizens were less likely to receive cervical or breast cancer screening as compared to immigrants who were U.S. citizens.

Consideration must be given to explanations for our findings. Noncitizen immigrants in our study were disproportionately affected by factors that have a negative impact on health care access and utilization. They had lower income, education, and English proficiency levels as compared to immigrants who had become U.S. citizens. However, as in previous reports, the impact of citizenship status persisted after taking these and health access factors into account,[30–32] suggesting a possible role for additional cultural, attitudinal, or social factors. Although we adjusted for acculturation by including language proficiency and years in the United States in the regression models, differences in cultural integration between naturalized citizens and noncitizens may have persisted, explaining some of the observed differences.[38] In addition, preventive cancer care is uncommon in many countries and immigrants

Table 2. Adjusted Odds Ratios and Prevalence Ratios from Logistic Regression Models Predicting Pap Smear Usage Among Immigrant Women Aged 18 and Older Without a Hysterectomy

Independent Variables	Had a Pap Smear Ever		Had a Pap Smear in the Past 3 Years	
	AOR (95% CI)	APR (95% CI)	AOR (95% CI)	APR (95% CI)
Race/ethnicity				
White	1.00	1.00	1.00	1.00
Hispanic	2.23 (1.27 to 3.92)	1.04 (1.02 to 1.06)	2.40 (1.57 to 3.66)	1.09 (1.06 to 1.12)
Asian	0.31 (0.20 to 0.48)	0.86 (0.77 to 0.93)	0.45 (0.33 to 0.63)	0.85 (0.77 to 0.92)
African American	2.55 (0.23 to28.55)	1.05 (0.80 to 1.08)	3.16 (0.52 to 19.22)	1.11 (0.88 to 1.16)
Other	0.57 (0.20 to 1.59)	0.95 (0.77 to 1.03)	0.59 (0.25 to 1.36)	0.91 (0.69 to 1.04)
Citizenship status				
Noncitizen	1.00	1.00	1.00	1.00
Naturalized citizen	1.54 (1.10 to 2.15)	1.05 (1.01 to 1.08)	1.51 (1.15 to 1.99)	1.07 (1.03 to 1.11)
Years in the U.S.				
<10	1.00	1.00	1.00	1.00
>10	1.40 (1.03 to 1.90)	1.06 (1.01 to 1.11)	1.15 (0.90 to 1.47)	1.03 (0.97 to 1.09)
Age, y				
18–29	1.00	1.00	1.00	1.00
30–39	3.94 (2.89 to 5.36)	1.21 (1.18 to 1.24)	2.51 (1.88 to 3.35)	1.18 (1.14 to 1.22)
40–49	3.91 (2.69 to 5.70)	1.21 (1.17 to 1.24)	1.97 (1.44 to 2.71)	1.15 (1.09 to 1.19)
50–64	4.52 (2.88 to 7.08)	1.23 (1.18 to 1.25)	1.91 (1.35 to 2.71)	1.14 (1.07 to 1.19)
65+	1.69 (1.09 to 2.62)	1.11 (1.02 to 1.17)	0.63 (0.43 to 0.93)	0.87 (0.75 to 0.98)
How well English is spoken				
Not well	1.00	1.00	1.00	1.00
Well	0.97 (0.70 to 1.35)	0.996 (0.95 to 1.03)	0.87 (0.66 to 1.14)	0.98 (0.92 to 1.02)
Health status				
Fair to poor	1.00	1.00	1.00	1.00
Excellent to good	1.14 (0.85 to 1.52)	1.01 (0.98 to 1.04)	1.21 (0.95 to 1.55)	1.03 (0.99 to 1.07)
Currently insured				
No	1.00	1.00	1.00	1.00
Yes	1.63 (1.23 to 2.17)	1.07 (1.03 to 1.10)	1.64 (1.29 to 2.08)	1.10 (1.05 to 1.13)
Has a usual source of care				
No	1.00	1.00	1.00	1.00
Yes	1.97 (1.46 to 2.67)	1.12 (1.08 to 1.16)	2.45 (1.90 to 3.17)	1.23 (1.18 to 1.27)
Federal poverty level				
<199%	1.00	1.00	1.00	1.00
>200%	1.74 (1.32 to 2.28)	1.05 (1.03 to 1.07)	1.59 (1.25 to 2.03)	1.07 (1.04 to 1.10)
Education level				
Less than high school	1.00	1.00	1.00	1.00
High school diploma	0.74 (0.53 to 1.03)	0.97 (0.95 to 1.03)	0.84 (0.63 to 1.12)	0.97 (0.92 to 1.02)
More than high school	0.81 (0.56 to 1.18)	0.98 (0.93 to 1.02)	0.94 (0.69 to 1.28)	0.99 (0.94 to 1.03)

CI, confidence interval; aOR, adjusted odds ratio; aPR, adjusted prevalence ratio; CI, confidence interval.

may perceive it as an element of the local culture. When they formally become U.S. citizens, they may embrace cancer screening as an expected behavior of citizens of this country. Acquiring U.S. citizenship may change attitudes and behaviors among many immigrants, including those related to health such as cancer screening.

Noncitizens are a heterogeneous group with diverse sociodemographic characteristics and health services use rates.[39–44] In the case of California, groups such as refugees and undocumented immigrants constitute an important proportion of noncitizen immigrants and their cancer screening rates may influence those of noncitizens in general.[43] Refugees and undocumented immigrants may be disproportionately affected by cultural, knowledge, and attitudinal barriers to cancer screening, such as fatalism, fear, and lack of knowledge of preventive health interventions or their benefits.[17,20–24] The legal status of some noncitizens may be an additional barrier, perceived by the patient or imposed by care sites intentionally or unintentionally, to access and use health services, including preventive care.[45,46] Legal and citizenship status have become progressively more important after California's ballot Proposition 187 and the Welfare reform of 1996. Proposition 187 intended to discontinue undocumented immigrants' eligibility for most health services while mandating that health care professionals report suspected undocumented patients to authorities.[47–55] The Personal Responsibility and Work Opportunity Reconciliation Act of 1996, the Welfare reform, greatly restricted the provision of public services to undocumented immigrants and based eligibility on citizenship status.[56–58] This adverse social climate for immigrants has been progressively exacerbating.[59] More inclusive law initiatives may assist in the implementation of outreach interventions and may contribute to reduce self-imposed barriers limiting access to cancer screening programs among immigrants.

We also found that having a usual source of care and health insurance were strong predictors of recent or ever screening for cervical and breast cancer among immigrants in California, supporting findings in previous studies.[1,3] Unexpectedly, education and language proficiency were not associated with the outcomes assessed in our sample of immigrants, with the exception of those with a high school diploma being more likely to report ever having a mammogram. In addition, in our study, Hispanic immigrants were more likely to receive cervical cancer screening as compared to whites and Asians. These findings may reflect the success of statewide campaigns targeting low-acculturated Hispanics. In

Table 3. Adjusted Odds Ratios and Prevalence Ratios from Logistic Regression Models Predicting Recent Mammograms Among Immigrant Women Aged 40 and Older

Independent Variables	Had a Mammogram Ever		Had a Mammogram in the Past 2 Years	
	AOR (95% CI)	APR (95% CI)	AOR (95% CI)	APR (95% CI)
Race/ethnicity				
White	1.00	1.00	1.00	1.00
Hispanic	0.69 (0.45 to 1.06)	0.97 (0.92 to 1.004)	0.79 (0.52 to 1.20)	0.90 (0.73 to 1.07)
Asian	0.46 (0.32 to 0.66)	0.92 (0.86 to 0.96)	0.50 (0.35 to 0.72)	0.71 (0.57 to 0.87)
African American	4.49 (0.89 to 22.70)	1.06 (0.99 to 1.08)	4.58 (0.91 to 23.14)	1.45 (0.96 to 1.62)
Other	0.51 (0.21 to 1.26)	0.93 (0.78 to 1.02)	0.57 (0.23 to 1.41)	0.77 (0.43 to 1.13)
Citizenship status				
Noncitizen	1.00	1.00	1.00	1.00
Naturalized citizen	2.15 (1.65 to 2.81)	1.17 (1.12 to 1.21)	2.15 (1.65 to 2.78)	1.38 (1.26 to 1.49)
Years in the U.S.				
<10 years	1.00	1.00	1.00	1.00
>10 years	1.60 (1.16 to 2.19)	1.15 (1.05 to 1.23)	1.58 (1.16 to 2.15)	1.27 (1.09 to 1.45)
Age, y				
40–49	1.00	1.00	1.00	1.00
50–64	2.80 (2.16 to 3.63)	1.17 (1.14 to 1.20)	2.09 (1.77 to 2.47)	1.38 (1.30 to 1.46)
65+	1.30 (0.89 to 1.91)	1.06 (0.97 to 1.12)	1.15 (0.89 to 1.48)	1.07 (0.94 to 1.21)
How well English is spoken				
Not well	1.00	1.00	1.00	1.00
Well	0.87 (0.60 to 1.25)	0.97 (0.87 to 1.04)	0.85 (0.63 to 1.14)	0.92 (0.78 to 1.06)
Health status				
Fair to poor	1.00	1.00	1.00	1.00
Excellent to good	0.92 (0.70 to 1.21)	0.98 (0.93 to 1.03)	0.93 (0.77 to 1.14)	0.97 (0.88 to 1.05)
Currently insured				
No	1.00	1.00	1.00	1.00
Yes	1.66 (1.21 to 2.26)	1.13 (1.05 to 1.20)	1.87 (1.47 to 2.39)	1.39 (1.24 to 1.55)
Has a usual source of care				
No	1.00	1.00	1.00	1.00
Yes	2.36 (1.68 to 3.32)	1.27 (1.18 to 1.35)	2.16 (1.61 to 2.90)	1.58 (1.35 to 1.80)
Federal poverty level				
<199%	1.00	1.00	1.00	1.00
>200%	1.18 (0.89 to 1.55)	1.03 (0.98 to 1.08)	1.13 (0.94 to 1.35)	1.06 (0.97 to 1.14)
Education level				
Less than high school	1.00	1.00	1.00	1.00
High school diploma	1.44 (1.03 to 2.00)	1.07 (1.01 to 1.12)	1.04 (0.80 to 1.35)	1.02 (0.89 to 1.14)
More than high school	1.39 (0.98 to 1.96)	1.07 (0.996 to 1.12)	1.10 (0.83 to 1.44)	1.05 (0.91 to 1.17)

CI, confidence interval; aOR, adjusted odds ratio; aPR, adjusted prevalence ratio; CI, confidence interval.

contrast, Asian immigrants were the least likely to report recent or ever screening for cervical or breast cancer. This result supports findings of previous studies[30,33,39,60–62] and highlights the need for culturally sensitive interventions targeting this relevant and numerous group.

Our study has several limitations. The CHIS data are based on self-report, which involves recall and social desirability biases. Previous research shows that self-report overestimates screening rates, in particular among low-income ethnic women such as noncitizens.[63–65] In this case, self-report would bias our results toward a null finding; it would decrease the cancer screening gap between naturalized citizens and noncitizens. Because of the relatively low response rate in the CHIS, the possibility of nonresponse bias exists. In an attempt to adjust for this, the racial/ethnic and sociodemographic profiles in our sample are weighted to match those of the 2000 U.S. Census for the state. Furthermore, the unweighted racial/ethnic profiles are very similar to those in the census data, suggesting that the sample is representative of the population of California. Due to the sensitive nature of legal status in the United States, some undocumented immigrants may have refused to participate in the survey. The exclusion of a group of noncitizens with presumably lower screening rates, such as undocumented immigrants, would increase the observed noncitizens' overall screening rates and decrease the disparities with citizens, again biasing the findings to the null hypothesis. Additionally, some noncitizens may have provided inaccurate representation of their citizenship or legal status. However, cross-contamination would also bias our results toward the null. Because legal status of CHIS participants is not provided in the publicly available database, we could not estimate the contribution of undocumented immigrants or refugees to the screening rates of noncitizens. Finally, the magnitude of effects was relatively small; however, a small increase in the prevalence ratio remains important because it represents large numbers of screening tests done.

Understanding the impact of citizenship status on Pap smear use may have public health policy implications at a national level but in particular to states such as California, Texas, New York, New Jersey, and Florida, where almost 70% of non-U.S. citizens live.[29] The findings of our study may help policy makers design more effective interventions aimed at eliminating barriers to cancer screening and help guide implementation of community-based educational programs and outreach initiatives among immigrants. Community-based educational interventions should highlight the need for screening among all women regardless of citizenship or legal residency status in a culturally sensitive manner. They should

also address potential psychological and attitudinal barriers among noncitizens such as the perception that cancer screening is a privilege or duty limited to citizens of this country. Future research should explore the nature of psychological and attitudinal barriers that prevent cancer screening among noncitizen immigrant women and look at additional cultural and social factors. Future studies should also explore the impact of legal status on use of Pap smears, mammograms, and other health services among subgroups of noncitizens.

Not being a U.S. citizen is a barrier to receiving cervical and breast cancer screening. Additional research is needed to explore causal factors for differences in cancer screening rates between citizens and noncitizens. For the time being, immigrants, especially those who are not U.S. citizens, should be targeted for improved health care access and appropriate cancer screening.

Supported by a grant from the American Cancer Society (CCCDA-03-197-01-CCCDA).

REFERENCES

1. **Swan J, Breen N, Coates RJ, Rimer BK, Lee NC.** Progress in cancer screening practices in the United States: results from the 2000 National Health Interview Survey. Cancer. 2003;97:1528–40.
2. **Skaer TL, Robison LM, Sclar DA, Harding GH.** Cancer-screening determinants among Hispanic women using migrant health clinics. J Health Care Poor Underserved. 1996;7:338–54.
3. **Calle EE, Flanders WD, Thun MJ, Martin LM.** Demographic predictors of mammography and Pap smear screening in US women. Am J Public Health. 1993;83:53–60.
4. **Trapido EJ, Burciaga Valdez R, Obeso JL, Strickman-Stein N, Rotger A, Perez-Stable EJ.** Epidemiology of cancer among Hispanics in the United States. J Natl Cancer Inst Monogr. 1995;18:17–28.
5. **Seeff LC, McKenna MT.** Cervical cancer mortality among foreign-born women living in the United States, 1985 to 1996. Cancer Detect Prev. 2003;27:203–8.
6. **Ries LAG, Eisner MP, Kosary CL, et al.** SEER Cancer Statistics Review, 1975–2000. Bethesda, MD: National Cancer Institute. Available at: http://seer.cancer.gov/csr/1975_2000. Accessed May 2003.
7. **Hedeen AN, White E.** Breast cancer size and stage in Hispanic American women, by birthplace: 1992–1995. Am J Public Health. 2001;91:122–5.
8. **Hedeen AN, White E, Taylor V.** Ethnicity and birthplace in relation to tumor size and stage in Asian American women with breast cancer. Am J Public Health. 1999;89:1248–52.
9. **Wu ZH, Black SA, Markides KS.** Prevalence and associated factors of cancer screening: why are so many older Mexican American women never screened? Prev Med. 2001;33:268–73.
10. **Howe HL, Wingo PA, Thun MJ, et al.** Annual report to the nation on the status of cancer (1973 through 1998), featuring cancers with recent increasing trends. J Natl Cancer Inst. 2001;93:824–42.
11. **Roetzheim RG, Pal N, Tennant C, et al.** Effects of health insurance and race on early detection of cancer. J Natl Cancer Inst. 1999;91:1409–15.
12. **Powell-Griner E, Bolen J, Bland S.** Health care coverage and use of preventive services among the near elderly in the United States. Am J Public Health. 1999;89:882–6.
13. **O'Malley AS, Mandelblatt J.** Delivery of preventive services for low-income persons over age 50: A comparison of community health clinics to private doctors' offices. J Community Health. 2003;28:185–97.
14. **O'Malley AS, Forrest CB, Mandelblatt J.** Adherence of low-income women to cancer screening recommendations. J Gen Intern Med. 2002;17:144–54.
15. **Mandelblatt JS, Gold K, O'Malley AS, et al.** Breast and cervix cancer screening among multiethnic women: role of age, health, and source of care. Prev Med. 1999;28:418–25.
16. **O'Malley AS, Mandelblatt J, Gold K, Cagney KA, Kerner J.** Continuity of care and the use of breast and cervical cancer screening services in a multiethnic community. Arch Intern Med. 1997;157:1462–70.
17. **O'Malley AS, Kerner J, Johnson AE, Mandelblatt J.** Acculturation and breast cancer screening among Hispanic women in New York City. Am J Public Health. 1999;89:219–27.
18. **Schur CL, Albers LA, Berk ML.** Health care use by Hispanic adults: financial vs. non-financial determinants. Health Care Financ Rev. 1995;17:71–88.
19. **Janes GR, Blackman DK, Bolen JC, et al.** Surveillance for use of preventive health-care services by older adults, 1995–1997. MMWR CDC Surveill Summ. 1999;48:51–88.
20. **Hubbell FA, Chavez LR, Mishra SI, Valdez RB.** Beliefs about sexual behavior and other predictors of Papanicolaou smear screening among Latinas and Anglo women. Arch Intern Med. 1996;156:2353–8.
21. **Chavez L, McMullin J, Mishra S, Hubbell F.** Beliefs matter: cultural beliefs and the use of cervical cancer screening tests. Am Anthropol. 2001;103:1–16.
22. **Ramirez AG, Suarez L, Laufman L, Barroso C, Chalela P.** Hispanic women's breast and cervical cancer knowledge, attitudes, and screening behaviors. Am J Health Promot. 2000;14:292–300.
23. **Tortolero-Luna G, Glober GA, Villarreal R, Palos G, Linares A.** Screening practices and knowledge, attitudes, and beliefs about cancer among Hispanic and non-Hispanic white women 35 years old or older in Nueces County, Texas. J Natl Cancer Inst Monogr. 1995;18:49–56.
24. **Perez-Stable EJ, Otero-Sabogal R, Sabogal F, McPhee SJ, Hiatt RA.** Self-reported use of cancer screening tests among Latinos and Anglos in a prepaid health plan. Arch Intern Med. 1994;154:1073–81.
25. **De Alba I, Sweningson JM, Chandy C, Hubbell FA.** Impact of English language proficiency on receipt of Pap smears among Hispanics. J Gen Intern Med. 2004;19:967–70.
26. **Goel M, Wee C, McCarthy E, Davis R, Ngo-Metzger Q, Phillips R.** Racial and ethnic disparities in cancer screening: the importance of foreign-birth as a barrier of care. J Gen Intern Med. 2003;18:1028–35.
27. **Carrasquillo O, Carrasquillo AI, Shea S.** Health insurance coverage of immigrants living in the United States: differences by citizenship status and country of origin. Am J Public Health. 2000;90:917–23.
28. **Jang M, Lee E, Woo K.** Income, language, and citizenship status: factors affecting the health care access and utilization of Chinese Americans. Health Soc Work. 1998;23:136–45.
29. **Schmidley A.** U.S. Census Bureau, Current Population Reports, Series P23-206, Profile of the Foreign-born Population in the United States: 2000. Washington, DC: U.S. Census Bureau, U.S. Government Printing Offic; 2001.
30. **Jang M, Lee E, Woo K.** Income, language, and citizenship status: factors affecting the health care access and utilization of Chinese Americans. Health Soc Work. 1998;23:136–45.
31. **Li A, Covinski K.** Citizenship status is an important determinant of health care disparity. J Gen Intern Med. 2003;18(suppl. 1):179.
32. **Katzburg J.** Impact of citizenship on access to preventive services among Latina women. Paper presented at the 128th annual meeting of the American Public Health Association, November 4, 2000, Boston, MA.
33. **Chaudhry S, Fink A, Gelberg L, Brook R.** Utilization of Papanicolaou smears by South Asian women living in the United States. J Gen Intern Med. 2003;18:377–84.
34. **California Health Interview Survey.** CHIS 2001 Methodology Series: Report 4—Response Rates. Los Angeles, CA: UCLA Center for Health Policy Research; 2002.
35. **United States Preventive Services Task Force.** Guide to Clinical Preventive Services. 2nd ed. Rockville, MD: Agency for Healthcare Research and Quality; 1996.
36. **Smith RA, Saslow D, Sawyer KA, et al.** American Cancer Society guidelines for breast cancer screening: update 2003. CA Cancer J Clin. 2003;53:141–69.
37. **Zhang J, Yu KF.** What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA. 1998;280:1690–1.
38. **Yang PQ.** Explaining immigrant naturalization. Int Migr Rev. 1994; 28:449–77.
39. **Taylor VM, Jackson JC, Yasui Y, et al.** Pap testing stages of adoption among Cambodian immigrants. Asian Am Pac Isl J Health. 2000;8: 58–68.
40. **Hubbell FA, Waitzkin H, Mishra SI, Dombrink J, Chavez LR.** Access to medical care for documented and undocumented Latinos in a southern California county. West J Med. 1991;154:414–7.
41. **Chavez LR, Hubbell FA, Mishra SI, Valdez RB.** Undocumented Latina immigrants in Orange County, CA: A comparative analysis. Int Migr Rev. 1997;31:88–107.

42. **Chavez LR, Cornelius WA, Jones OW.** Utilization of health services by Mexican immigrant women in San Diego. Women Health. 1986;11:3–20.
43. **U.S. Department of Homeland Security.** Yearbook of Immigration Statistics 2002. Washington, DC: U.S. Department of Homeland Security; 2003.
44. **Fix M, Passel JS, Sucher K.** Trends in naturalization. The Urban Institute. September 2003. Available at: http://www.urban.org/url.cfm?ID=310847. Accessed March 1, 2004.
45. **Asch S, Leake B, Gelberg L.** Does fear of immigration authorities deter tuberculosis patients from seeking care? West J Med. 1994;161:373–6.
46. **Crane LA, Kaplan CP, Bastani R, Scrimshaw SC.** Determinants of adherence among health department patients referred for a mammogram. Women Health. 1996;24:43–64.
47. **Clark WA.** Large-scale immigration and political response: popular reaction in California. Int J Popul Geogr. 1998;4:1–10.
48. **Fenton JJ, Catalano R, Hargreaves WA.** Effect of Proposition 187 on mental health service use in California: A case study. Health Aff (Millwood). 1996;15:182–90.
49. **Fenton JJ, Moss N, Khalil HG, Asch S.** Effect of California's Proposition 187 on the use of primary care clinics. West J Med. 1997;166:16–20.
50. **Freeman GP.** Reform and retreat in United States immigration policy. People Place. 1998;6:1–11.
51. **Marx JL, Thach AB, Grayson G, Lowry LP, Lopez PF, Lee PP.** The effects of California Proposition 187 on ophthalmology clinic utilization at an inner-city urban hospital. Ophthalmology. 1996;103:847–51.
52. **McBride MJ.** Migrants and asylum seekers: policy responses in the United States to immigrants and refugees from Central America and the Caribbean. Int Migr. 1999;37:289–317.
53. **Ziv TA, Lo B.** Denial of care to illegal immigrants. Proposition 187 in California. N Engl J Med. 1995;332:1095–8.
54. **Lee YT, Ottati V.** Attitudes toward U.S. immigration policy: the roles of in-group-out-group bias, economic concern, and obedience to law. J Soc Psychol. 2002;142:617–34.
55. **Halfon N, Wood DL, Valdez RB, Pereyra M, Duan N.** Medicaid enrollment and health services access by Latino children in inner-city Los Angeles. JAMA. 1997;277:636–41.
56. **Kullgren JT.** Restrictions on undocumented immigrants' access to health services: the public health implications of welfare reform. Am J Public Health. 2003;93:1630–3.
57. **Fix M, Passel JS.** Trends in noncitizens' and citizen' use of public benefits following Welfare reform: 1994–97. The Urban Institute, Available at: http://urban.org/url.cfm?ID=408086. Accessed October 12, 2003.
58. **Ellwood MR, Ku L.** Welfare and immigration reforms: unintended side effects for Medicaid. Health Aff (Millwood). 1998;17:137–51.
59. **The Gallup Organization.** Americans worried about immigration, oppose Bush plan; majority says immigrants hurt economy. The Gallup Organization. Available at: http://www.gallup.com/poll/content/login.aspx?ci=10195. Accessed March 1, 2004.
60. **McPhee SJ, Stewart S, Brock KC, Bird JA, Jenkins CN, Pham GQ.** Factors associated with breast and cervical cancer screening practices among Vietnamese American women. Cancer Detect Prev. 1997;21:510–21.
61. **Tu SP, Yasui Y, Kuniyuki A, et al.** Breast cancer screening among Cambodian American women. Cancer Detect Prev. 2000;24:549–63.
62. **Yu SM, Huang ZJ, Singh GK.** Health status and health services utilization among US Chinese, Asian Indian, Filipino, and other Asian/Pacific Islander children. Pediatrics. 2004;113(pt 1):101–7.
63. **Hiatt RA, Perez-Stable EJ, Quesenberry C Jr, Sabogal F, Otero-Sabogal R, McPhee SJ.** Agreement between self-reported early cancer detection practices and medical audits among Hispanic and non-Hispanic white health plan members in northern California. Prev Med. 1995;24:278–85.
64. **Paskett ED, Tatum CM, Mack DW, Hoen H, Case LD, Velez R.** Validation of self-reported breast and cervical cancer screening tests among low-income minority women. Cancer Epidemiol Biomarkers Prev. 1996;5:721–6.
65. **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;35:463–73.

Epilepsia, 46(9):1418–1422, 2005
Blackwell Publishing, Inc.

Characteristics of Male Veterans with Psychogenic Nonepileptic Seizures

*Barbara A. Dworetzky, †Andreja Strahonja-Packard, ‡Christopher W. Shanahan, §Jeanette Paz, †Barbara Schauble, and *Edward B. Bromfield

*Brigham and Women's Hospital, Harvard Medical School; †Boston University School of Medicine; ‡Boston Medical Center, Boston University School of Medicine, and §Boston VA Healthcare System, Boston, Massachusetts, U.S.A.

Summary: *Purpose:* To describe male patients (pts) with psychogenic nonepileptic seizures (PNESs) followed up in a Veteran's Administration (VA) seizure clinic and to compare them with those with epileptic seizures (ESs) by using clinical, and psychosocial variables.

Methods: Adult male veterans seen between 1997 and 2000 with ESs were compared with those with PNESs with respect to clinical history (head trauma, antiepileptic drug exposure, depression, anxiety, substance abuse, seizure description), documented chronic pain, posttraumatic stress disorder (PTSD), compensation for diagnosis, neurologic examination, and test results including imaging and EEG data.

Results: Men with PNESs were younger and reported more frequent events, and diagnoses of chronic pain, anxiety, and PTSD were significantly greater. Neuroimaging [computed tomography (CT) or magnetic resonance imaging (MRI) of the brain] and neurologic examination were significantly more likely to be normal or nonspecific in pts with PNESs, although history of ictal urinary incontinence or service-connected compensation for diagnosis did not distinguish the groups.

Conclusions: Male veterans with PNESs have characteristics similar to those reported in the literature, even though younger women have dominated previously studied populations. Compared with men with ESs, those with PNESs are more likely to have chronic pain, anxiety, and PTSD, as well as normal examinations and brain imaging. **Key Words:** Gender—Men—Veterans—Psychogenic nonepileptic seizures—Video-EEG monitoring.

Psychogenic nonepileptic seizures (PNESs) are common in patients with seizure-like episodes that do not respond to antiepileptic drugs (AEDs) (1,2). Consequently, patients with PNESs are at risk of receiving inappropriate medications for prolonged treatment periods or invasive intervention such as vagal nerve stimulation (3). These treatments are costly and potentially dangerous. Diagnosis is typically delayed until referral for video-EEG monitoring (4,5). Presence of PNESs has been shown to be associated with a lower quality of life and higher level of stress compared with the presence of epileptic seizures (ESs) (6–8). This further highlights the importance of accurate diagnosis.

Most of what is understood about PNESs is derived from studies conducted at tertiary epilepsy centers where 10 to 58% of adult patients with intractable spells are ultimately diagnosed with PNESs (9,10). In these studies, 65–80% of patients are young and female, which has led to the commonly held opinion that PNESs are less important in men. Furthermore, a small number of studies have demonstrated that significant gender differences may exist in background and clinical manifestations (11,12). For example, compared with women, men with PNESs are more likely to be substance abusers, report minor head injury, have a documented financial gain from having seizures, and worse emotional adjustment by MMPI-II testing (12–14). Having a history of physical or sexual abuse appears to be more common in women with PNES, but is still more likely in men with PNESs than in the general population of men. Abuse is correlated with conversion disorder, the primary psychopathologic correlate of PNESs; nonconversion etiologies such as anxiety, psychosis, and impulse-control disorders, interestingly, have shown similar frequencies in men and women (15,16) Other studies suggest that PNESs may appear clinically different in men than in women (12,16), with more dramatic-appearing motor movements and less affective change in men (17). Gender also may be an important prognostic factor, because the probability of a favorable outcome of PNESs is higher in women than in men (18). Gender differences in the use of health care services and the widespread belief that

Accepted April 24, 2005.
Address correspondence and reprint requests to Dr. B.A. Dworetzky at Brigham & Women's Hospital, 75 Francis Street, Boston, MA 02115, U.S.A. E-mail: bdworetzky@partners.org

PNESs is primarily a woman's diagnosis may impede referral to epilepsy centers and lead to delayed diagnosis of PNESs in men (19,20).

The Veteran's Administration (VA) population consists largely of men older than 50 years. Three years of video-EEG monitoring experience has convinced us that PNESs are not rare in the VA population. We describe the characteristics of male veterans with probable or definite PNESs evaluated in a VA Hospital Seizure Clinic, and compare them with male veterans with ESs with respect to clinical, psychiatric, and psychosocial variables.

METHODS

Subjects were consecutive adult male veterans evaluated for spells in a VA seizure clinic or admitted to the VA neurology ward for long-term video-EEG monitoring (LTM) between 1997 and 2000. This study was approved by the Institutional Review Board of the Boston VA Healthcare System. Of 2,443 total patient encounters, 288 unique patients were seen. Female patients were excluded (n = 37); 140 male patients were excluded on the basis of incomplete data, limited mental status examination, or lack of evidence for an episodic disorder. Of 111 eligible patients, we excluded those with probable PNESs but definite interictal epileptiform discharges (n = 3), no clinical events captured during the monitored period (n = 45), only subjective events during monitoring (n = 4), or purely physiological spells [vasovagal syncope, paroxysmal coughing, hypnic (sleep onset) jerks; n = 3]. Of the remaining patients, 22 had PNESs based on a history of paroxysmal behavioral events to suggest epileptic seizures, and at least one typical event captured on monitoring either spontaneously or with suggestion strongly suggesting PNESs without ictal, interictal, or postictal EEG change, no response to AEDs, and no events suggestive of ESs, as defined by Vanderzant et al. (21). The ESs group totaled 34 and was defined as patients with an unequivocal history of epilepsy, no events to suggest PNESs, and definite epileptiform discharges on EEG, ictally (n = 12) or interictally (n = 22) or both, interpreted by a certified electroencephalographer (B.D.). Duration of disease for this group on average was 18.3 years, with a range from 1 to 56 years. Classification of seizure types and epilepsy syndromes are shown in Table 1. Clinical history and neuroimaging data extracted from the electronic medical record were used to compare PNES and ES patients. A registered nurse (J.P.) interviewed all patients and recorded on a standardized intake form information regarding history of chronic pain, sleep problems, substance abuse, prior head trauma, family history of seizures, history of childhood epilepsy, febrile convulsions, and physical or sexual abuse. Neuropsychological assessment was reviewed where available but was too incomplete for formal analysis. AED therapy and service-connected compensation was ascertained from the medical record. Psychiatric diagnoses, including posttraumatic stress disorder (PTSD), anxiety, and depression, were included if they were documented in the medical record by a psychiatrist, psychologist, or social worker. All data were analyzed by using two-tailed *t* tests for continuous variables and Fisher's exact tests for categorical variables.

TABLE 1. *Classification and etiologies of seizures and epilepsy in male veterans*

	No.
Simple partial seizures	3
Complex partial seizures	18
Secondarily generalized seizures	8
Idiopathic generalized epilepsy	4
Idiopathic partial epilepsy	3
Symptomatic partial epilepsy	27
Traumatic brain injury	14
Neoplasm	4
Vascular	7

RESULTS

Compared to male veterans with ESs, those with PNESs were ~10 years younger (Table 2). Handedness, education level, employment status, and race did not significantly differ between groups (see Table 2). PNES patients were more likely to have a normal neurologic examination and a normal magnetic resonance imaging (MRI) scan of the brain (see Table 2). PNES patients had more-frequent events than did ES patients, with the majority complaining of daily episodes (Table 3). The groups reported similar clinical symptoms, with the exception of more staring spells for the ES group (see Table 3). Furthermore, report of incontinence by history was similar in both groups. Both groups had similar reports of substance abuse, sleep-related problems, and depressive symptoms, whereas PNES patients were significantly more likely to report chronic pain, anxiety, and PTSD (Table 4). More than 95% of PNES patients were treated with AEDs before receiving their diagnosis. Of these, 77% received monotherapy. ES patients were more likely to be receiving polytherapy (see Table 2). PNES and ES groups did not significantly differ with respect to the proportion of patients who received compensation for their diagnosis or in the number of years that a patient experienced the seizures before entry into this study (see Table 2).

DISCUSSION

This study offers a unique opportunity to focus on differences between individuals with ESs and PNESs in a relatively older male population, a group that has been underrepresented in the epilepsy literature as well as in practice in epilepsy monitoring units. McBride and colleagues

TABLE 2. *Patient characteristics of male veterans, obtained through medical record and self-report*

	ESs (n = 34)	PNESs (n = 22)	Fisher's Exact
Handedness			
Right	31	20	p = NS
Age			
yr (SD)	61.2 (14)	50.1 (13)	p < 0.01[a]
Race			
White	26	21	p = NS
Nonwhite	8	1	
Education			
yr (SD)	11.4 (2.5)	12.4 (3)	p = NS[a]
Time with symptoms before diagnosis in study (mean no. yr)			
≤1 yr	10	6	p = NS
>1 yr	24	16	
AED therapy			
No therapy	1	1	p < 0.002
Monotherapy	13	17	
>1 drug	20	4	
History of febrile seizure			
Yes	0	1	p = NS
Compensation			
Yes	13	11	p = NS
Employment history	n = 32	n = 21	
Full time/Part time	5	4	
Retired	5	2	p = NS
Disabled	22	15	
Neurologic examination	n = 32		
Abnormal	19	1	p < 0.001
Normal	13	15	
Brain imaging (MRI/CT)	n = 31		
Abnormal	26	11	p < 0.01
Normal	5	11	

PNES, psychogenic nonepileptic seizure; ES, epileptic seizure; MRI, magnetic resonance imaging; CT, computed tomography; NS, not significant.

[a]Two-sample t test (two tailed).

TABLE 3. *Seizure characteristics, from clinical history of male veterans*

	ESs (n = 34)	PNESs (n = 22)	Fisher's Exact
Seizure frequency	n = 29	n = 19	
None/yr	8	0	
≤2 /mo	12	5	p < 0.0001
Weekly	7	3	
Daily	2	11	
Limb movements	n = 30		
No	8	11	p = NS
Yes	22	11	
Staring	n = 29		
Yes	18	8	p < 0.05
Lip smacking	n = 30		
Yes	5	1	p = NS
Sensory symptoms	n = 27		
Yes	4	3	p = NS
Aura	n = 33		
Yes	15	13	p = NS
Incontinence	n = 31		
Yes	8	5	p = NS

ES, epileptic seizure; PNES, psychogenic nonepileptic seizure; NS, nonsignificant.

(22), publishing retrospective data on PNES in older populations, reported that 27% of patients older than 60 years who undergo diagnostic video-EEG monitoring are found to have NES; 13%, PNES; and 14%, of physiologic cause (i.e., cataplexy, hypotension). This and other recent studies suggest that PNESs, either psychogenic or physiologic, should be strongly considered in older patients with episodic seizure-like events (23,24).

Despite reported gender differences, the clinical characteristics of male veterans with NESs are similar to what has been shown in the literature for women. When compared with patients with ESs, male veterans with PNESs in this investigation were, on average, 10 years younger and reported a higher frequency of episodes. However, the groups did not differ with respect to education or employment history. Incidence of PTSD, a clear risk factor for PNESs in women, has been found in a significantly higher proportion of the male PNES than ES patients as well (25–27). The association of chronic pain, painful "auras," especially headaches, and posttraumatic stress with PNESs has been demonstrated for younger female populations; this same association was found with our male veterans. This suggests that chronic pain and PTSD may be general predictors of PNESs, regardless of patients' gender or age (28). Urinary incontinence, traditionally used as a predictor to support a diagnosis of ESs, was found not to be associated with ESs in our study, because no difference was found between groups. This is similar to what Peguero et al. (29) found by telephone survey. As in previous studies (16–18,30), we found a similarly elevated rate of occurrence of sleep problems and depression in both groups.

We believe that the majority of PNES patients in our study were overtreated, with 77% taking one AED and

TABLE 4. *Behavioral and neuropsychiatric profile of male veterans, obtained from medical records and self-report*

	ESs (n = 34)	PNESs (n = 22)	Fisher's Exact
Substance abuse	n = 33		
Yes (any, incl. EtOH)	18	14	p = NS
Sleep problems	n = 27		
Yes	14	10	p = NS
Chronic pain	n = 26		
Yes	13	19	p < 0.008
Anxiety	n = 32		
Yes	12	16	p < 0.009
PTSD	n = 28		
Yes	3	14	p < 0.0001
Depression	n = 33		
Yes	18	16	p = NS

ES, epileptic seizure; PNES, psychogenic nonepileptic seizure; NS, nonsignificant.

18% taking two or more. We also found 96% were receiving compensation, with nearly 42% of the total receiving it specifically for their diagnosis. This compensation rate was similar for both groups, which on first glance may seem surprising, but might imply a similar disability. Unnecessary treatments are concerning because of potential unwanted side effects from the drugs themselves or from drug–drug interactions, which can cause serious consequences, especially in the elderly (31). These can lead to unnecessary medical costs for visits, blood draws, and treatment of side effects. In a small study of involving 20 patients with diagnostic LTM, an 84% average reduction was found in total seizure-related medical charges within 6 months of making the diagnosis (5). Significantly fewer young female patients with PNESs are treated with AEDs (32,33), perhaps because young women's seizures are perceived as more likely to have a psychological etiology or because of possible risk of the drug should they become pregnant. Additionally, men with PNESs have been shown in the literature to exhibit more tonic–clonic-like events than do women (12), and perhaps this display leads to a higher likelihood of therapy, in that events appear more severe.

Our study supports the observations of others that PNES patients are less likely than ES patients to have abnormal neuroimaging or neurologic examinationss, despite the confirmation of a previously reported high incidence of head trauma in PNES (range, 32–70%) (14,34). Additionally, in individuals with PNESs, head injury has been found to be associated with poor long-term outcome, including long-lasting disability (14), despite being relatively mild.

Substance abuse was common in both groups. Previous studies have shown this to be more common in men with PNESs (18); however, substance abuse, in particular with alcohol, is elevated in the population of male veterans in general (35). Studies also have suggested that men may use PNESs to avoid responsibilities (16). It is possible that service-connected compensation for illness may increase the likelihood for treatment in patients attending VA clinics. However, in our study, the proportion of patients with possible secondary gain is not different between groups. Compensation may have other effects, such as perpetuating the misdiagnosis of ongoing seizures. However, it is important to note that reimbursement does not imply conscious malingering, and it is likely that PNESs represent a form of conversion disorder in our population, as in most women (36). It is notable that a relatively high percentage of VA PNES patients are categorized as disabled, based on the diagnosis of seizure disorder; as has been reported, an individual can be similarly or even more disabled by PNESs than by ESs (6–8,37). Higher ratings of disability, lower self-reported quality-of-life scores, greater impact on families, and overall misdiagnosis and mistreatment have been reported in patients with PNESs (6,8,38,39).

Our study demonstrates that male veterans with PNESs have risk factors similar to those reported in the literature, even though younger females have been the predominant group studied to date. Although this study uses limited neuropsychiatric assessments, it nevertheless introduces data that suggest important distinguishing characteristics for men with PNESs. Thus we believe that screening male patients with frequent uncontrolled seizures should include questions designed to detect various stressors, including a history of PTSD and chronic pain. Further studies are needed to confirm that our results are generalizable to the population of civilian men and to investigate the influence of gender and age on the expression, treatment, and outcome of PNESs.

REFERENCES

1. Gulick TA, Spinks IP, King DW. Pseudoseizures: ictal phenomena. *Neurology* 1982;32:24–30.
2. Boon PA, Williamson PD. The diagnosis of pseudoseizures. *Clin Neurol Neurosurg* 1993;95:1–8.
3. Britton JW, Rathke KM, Cascino GD, et al. Vagus nerve stimulator implantation in patients with nonepileptic events: a costly result of misdiagnosis. *Epilepsia* 2002;43(suppl 7):161.
4. de Timary P, Fouchet P, Sylin M, et al. Nonepileptic seizures delayed diagnosis in patients presenting with electrographic (EEG) or clinical signs of epileptic seizures. *Seizure* 2002;11:193–7.
5. Martin R, Gilliam F, Kilgore M, ct al. Improved health care resource utilization following video-EEG-confirmed diagnosis of nonepileptic psychogenic seizures. *Seizure* 1998;7:385–90.
6. Szaflarski JP, Hughes C, Szaflarski M, et al. Quality of life in psychogenic nonepileptic seizures. *Epilepsia* 2003;44:236–42.
7. Tojek TM, Lumley M, Barkley G, et al. Stress and other psychosocial characteristics of patients with psychogenic nonepileptic seizures. *Psychosomatics* 2000;41:221–6.
8. Breier JI, Fuchs KL, Brookshire BL, et al. Quality of life perception in patients with intractable epilepsy or pseudoseizures. *Arch Neurol* 1998;55:660–5.
9. Sigurdardottir KR, Olafsson E. Incidence of psychogenic seizures in adults: a population-based study in Iceland. *Epilepsia* 1998;39:749–52.
10. Krumholz A. Nonepileptic seizures: diagnosis and management. *Neurology* 1999;53:S76–83.
11. Sirven JI, Glosser DS. Psychogenic nonepileptic seizures: theoretic and clinical considerations. *Neuropsychiatry Neuropsychol Behav Neurol* 1998;11:225–35.
12. van Merode T, de Krom MC, Knottnerus JA. Gender-related differences in non-epileptic attacks: a study of patients' cases in the literature. *Seizure* 1997;6:311–6.
13. Holmes MD, Dodrill CB, Bachtler S, et al. Evidence that emotional maladjustment is worse in men than in women with psychogenic nonepileptic seizures. *Epilepsy Behav* 2001;2:568–73.
14. Westbrook LE, Devinsky O, Geocadin R. Nonepileptic seizures after head injury. *Epilepsia* 1998;39:978–82.
15. Alper K, Devinsky O, Perrine K, et al. Psychiatric classification of nonconversion nonepileptic seizures. *Arch Neurol* 1995;52:199–201.
16. Bowman ES, Markland OM. Psychodynamics and psychiatric diagnoses of pseudoseizure subjects. *Am J Psychiatry* 1996;153(suppl 1):57–63.
17. Gratama CA, Wendt L, Wood CR, et al. Gender differences in psychogenic nonepileptic seizures: motor and affectual features. *Epilepsia* 2002;43(suppl 7).
18. Bowman ES. Nonepileptic seizures: psychiatric framework, treatment, and outcome. *Neurology* 1999;53:S84–8.
19. Mustard C, Kaufert P, Kozyrsky A, et al. Sex differences in the use of health care services. *N Engl J Med* 1998;338:1678–83.

20. Bortz J, Blum D. Nonepileptic seizures may be underappreciated in men. *Neurol Rev* 2000;10:31.
21. Vanderzant CW, Giordani B, Berent S, et al. Personality of patients with pseudoseizures. *Neurology* 1986;36:664–8.
22. McBride AE, Shih TT, Hirsch LJ. Video-EEG monitoring in the elderly: a review of 94 patients. *Epilepsia* 2002;43:165–9.
23. Lancman ME, O'Donovan CO, Dinner D, et al. Usefulness of prolonged video-EEG monitoring in the elderly. *J Neurol Sci* 1996;142:54–8.
24. Drury I, Selwa LM, Schuh LA, et al. Value of inpatient diagnostic CCTV-EEG monitoring in the elderly. *Epilepsia* 1999;40:1100–2.
25. Rosenberg HJ, Rosenberg SD, Williamson PD, et al. A comparative study of trauma and posttraumatic stress disorder prevalence in epilepsy patients and psychogenic nonepileptic seizure patients. *Epilepsia* 2000;41:447–52.
26. Bowman E. Etiology and clinical course of pseudoseizures: relationship to trauma, depression, and dissociation. *Psychosomatics* 1993;34:333–42.
27. Kanner A, Parra J, Frey M, et al. Psychiatric and neurologic predictors of psychogenic pseudoseizure outcome. *Neurology* 1999;53:933–8.
28. Geyer JD, Payne TA, Drury I. The value of pelvic thrusting in the diagnosis of seizures and pseudoseizures. *Neurology* 2000;54:227–9.
29. Peguero E, Abou-Khalil B, Fakhoury T, et al. Self-injury and incontinence in psychogenic seizures. *Epilepsia* 1995;36:586–91.
30. Brunsen R, Ehsan T, Dellabadia J. Pseudoseizures in men: clinical features and associated psychiatric diagnosis. *Epilepsia* 2000(AES Proceedings).
31. Bourdet SV, Gidal BE, Alldredge BK. Pharmacologic management of epilepsy in the elderly. *J Am Pharm Assoc* 2001;41:421–36.
32. Benbadis SR. How many patients with pseudoseizures receive antiepileptic drugs prior to diagnosis? *Eur Neurol* 1999;41:114–5.
33. Krahn LE, Reese MM, Rummans TA, et al. Health care utilization of patients with psychogenic nonepileptic seizures. *Psychosomatics* 1997;38:535–42.
34. Barry E, Krumholz A, Bergey GK, et al. Nonepileptic posttraumatic seizures. *Epilepsia* 1998;39:427–31.
35. Office of Applied Studies. *The NHSDA Report: Alcohol Use among Veterans*. Rockville, MD: Substance Abuse and Mental Health Sources Administration, 2001.
36. Orbach D, Ritaccio A, Devinsky O. Psychogenic, nonepileptic seizures associated with video-EEG-verified sleep. *Epilepsia* 2003;44:64–8.
37. Chabolla DR, Krahn LE, So EL, et al. Psychogenic nonepileptic seizures. *Mayo Clin Proc* 1996;711:493–500.
38. Meierkord H, Will B, Fish D, et al. The clinical features and prognosis of pseudoseizures diagnosed using video-EEG telemetry. *Neurology* 1991;41:1643–6.
39. Krawetz P, Fleisher W, Pillay N, et al. Family functioning in subjects with pseudoseizures and epilepsy. *J Nerv Ment Dis* 2001;189:38–43.

Gender Differences in the Risk of Ischemic Stroke and Peripheral Embolism in Atrial Fibrillation

The AnTicoagulation and Risk factors In Atrial fibrillation (ATRIA) Study

Margaret C. Fang, MD, MPH; Daniel E. Singer, MD; Yuchiao Chang, PhD; Elaine M. Hylek, MD, MPH; Lori E. Henault, MPH; Nancy G. Jensvold, MPH; Alan S. Go, MD

Background—Previous studies provide conflicting results about whether women are at higher risk than men for thromboembolism in the setting of atrial fibrillation (AF). We examined data from a large contemporary cohort of AF patients to address this question.

Methods and Results—We prospectively studied 13 559 adults with AF and recorded data on patients' clinical characteristics and the occurrence of incident hospitalizations for ischemic stroke, peripheral embolism, and major hemorrhagic events through searching validated computerized databases and medical record review. We compared event rates by patient sex using multivariable log-linear regression, adjusting for clinical risk factors for stroke, and stratifying by warfarin use. We identified 394 ischemic stroke and peripheral embolic events during 15 494 person-years of follow-up off warfarin. After multivariable analysis, women had higher annual rates of thromboembolism off warfarin than did men (3.5% versus 1.8%; adjusted rate ratio [RR], 1.6; 95% CI, 1.3 to 1.9). There was no significant difference by sex in 30-day mortality after thromboembolism (23% for both). Warfarin use was associated with significantly lower adjusted thromboembolism rates for both women and men (RR, 0.4; 95% CI, 0.3 to 0.5; and RR, 0.6; 95% CI, 0.5 to 0.8, respectively), with similar annual rates of major hemorrhage (1.0% and 1.1%, respectively).

Conclusions—Women are at higher risk than men for AF-related thromboembolism off warfarin. Warfarin therapy appears be as effective in women, if not more so, than in men, with similar rates of major hemorrhage. Female sex is an independent risk factor for thromboembolism and should influence the decision to use anticoagulant therapy in persons with AF. (***Circulation.* 2005;112:1687-1691.**)

Key Words: anticoagulants ■ atrial fibrillation ■ risk factors ■ stroke ■ women

Atrial fibrillation is the most common clinically significant cardiac arrhythmia and a major risk factor for ischemic stroke and peripheral embolism.[1] Warfarin therapy substantially reduces the risk of atrial fibrillation–related thromboembolism but also increases the risk for hemorrhage.[2] Optimal administration of warfarin requires appropriate risk stratification.

Several prominent schemes are available to facilitate identification of patients at high-enough risk of thromboembolism to merit anticoagulant therapy.[2–5] These schemes, however, provide conflicting recommendations as to whether women with atrial fibrillation are at higher risk for stroke independently of other known risk factors. The Stroke Prevention in Atrial Fibrillation (SPAF)[3] and Framingham risk scores[5] consider women to be at higher risk for ischemic stroke, whereas other studies do not (eg, Atrial Fibrillation Investigators [AFI][2] and $CHADS_2$[4] risk indexes). Notably, the SPAF investigators found only the subset of women >75 years to be at higher risk for stroke.

Variations in the risk assessment for stroke can lead to significant differences in the use of warfarin therapy for atrial fibrillation.[6] To test the hypothesis of whether women are at higher risk for atrial fibrillation–related thromboembolism, we analyzed data from the large AnTicoagulation and Risk factors In Atrial fibrillation (ATRIA) study cohort, comparing rates of ischemic stroke and peripheral embolism between male and female patients not taking anticoagulants while controlling for other known risk factors for thromboembolism.

Methods

ATRIA is a cohort study of 13 559 adults with diagnosed nonvalvular atrial fibrillation who received care within Kaiser Permanente of Northern California, a large integrated healthcare delivery system. Details of the cohort assembly and validation have been described

Received April 4, 2005; revision received May 24, 2005; accepted June 2, 2005.

From the University of California, San Francisco (M.C.F., A.S.G.); Clinical Epidemiology Unit, Massachusetts General Hospital, Boston (D.E.S., Y.C., L.E.H.); General Internal Medicine and Clinical Epidemiology Unit, Boston University School of Medicine, Boston, Mass (E.M.H.); and Division of Research, Kaiser Permanente of Northern California, Oakland (N.G.J., A.S.G.).

Guest Editor for this article was Burton E. Sobel, MD.

Correspondence 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 mfang@medicine.ucsf.edu

Circulation is available at http://www.circulationaha.org

DOI: 10.1161/CIRCULATIONAHA.105.553438

previously.[7] Cohort members were assembled between July 1, 1996, and December 31, 1997, by searching automated inpatient, outpatient, and ECG databases for physician-assigned *International Classification of Diseases, Ninth Revisions, Clinical Modification* (ICD-9) diagnosis of atrial fibrillation (427.31). Patients with diagnosed mitral stenosis, valvular repair or replacement, transient postoperative atrial fibrillation, or concurrent hyperthyroidism were excluded. The cohort was followed up through August 31, 1999, providing a median follow-up of 2.4 years. Follow-up was censored at the time of an outcome event, death, or disenrollment from the health plan.

Thromboembolic and Hemorrhagic Events

We searched hospitalization and billing claims databases for primary discharge diagnoses of thromboembolic events (ischemic stroke and peripheral embolism) using validated algorithms.[8] All potential events were individually validated through medical records review by 2 members of a 3-physician outcomes committee, with discordances resolved by consensus of the 3-member committee, including outside consultant review in selected circumstances. A validated ischemic stroke was defined as the sudden onset of a neurological deficit persisting >24 hours and not explained by other origins. A validated peripheral embolism required confirmation by radiographic imaging, intraoperative examination, or pathological findings and the absence of underlying atherosclerotic disease in the affected artery. We excluded patients who developed events during inpatient hospitalization or from periprocedural complications. Mortality at 30 days after outcome events was based on review of medical records, health plan databases, and California State death files.[9]

Using previously described and validated methods,[8] we identified hemorrhagic events by searching for primary and secondary diagnoses of intracranial hemorrhage and primary diagnoses of extracranial hemorrhage. We excluded intracranial hemorrhages from major trauma. We defined major hemorrhage as fatal, requiring transfusion of ≥2 U packed blood cells, or hemorrhage into a critical anatomic site.

Clinical Characteristics

Data on patient age and sex were obtained from administrative databases. Medical diagnoses related to stroke and hemorrhage risk were obtained by searching hospital discharge and ambulatory visit databases for specific ICD-9–coded diagnoses using previously described and validated methods.[7] Warfarin exposure was determined using a combination of pharmacy, laboratory, and ambulatory visit databases.[8] Anticoagulation intensity was measured using outpatient assessments of the international normalized ratio (INR) obtained from health plan laboratory databases. For patients receiving warfarin, we calculated the proportion of person-time at different INR intervals using an adapted linear interpolation method.[10] If a person was on warfarin by pharmacy records but the interval between INR measurements was >8 weeks, we did not interpolate INR values for this extended period and categorized these INR periods as "not available"; 18% of total person-time fell into this category. Finally, we assessed for longitudinal exposure to oral estrogens (either alone or in combination therapy with progesterone) on the basis of filled prescriptions found in health plan pharmacy databases and a previously validated algorithm.[11]

Statistical Analyses

We compared clinical characteristics of men and women during periods off warfarin using χ^2 tests and compared thromboembolism rates using log-linear models with generalized estimating equations. Multivariable log-linear regression was then used to adjust for previously identified stroke risk factors: age (as a continuous variable by decades), prior ischemic stroke, hypertension, congestive heart failure, coronary artery disease, diabetes mellitus, and estrogen replacement therapy,[12,13] with time-dependent covariates as appropriate. We also tested whether an interaction existed between patient sex and age in the rate of thromboembolism off warfarin, first testing whether an interaction existed when age was dichotomized at 75 years (as reported in SPAF)[3] and then with age as a continuous variable. To assess whether the effectiveness of warfarin varied by sex, we tested the interaction term of warfarin and patient sex in models of thromboembolism that included patients both on and off warfarin therapy. We also compared rates of major hemorrhage by sex, adjusting for risk factors for extracranial and intracranial hemorrhage (age, prior gastrointestinal hemorrhage, hematuria, or other prior hemorrhage, cirrhosis, dementia, mechanical fall during a prior hospitalization, prior stroke, hypertension, and anticoagulation intensity)[14] and tested the interaction term for warfarin and patient sex in models predicting hemorrhagic events.

TABLE 1. Clinical Characteristics and Proportion of Person-Time in $CHADS_2$ Risk Categories Between Women and Men With Atrial Fibrillation Who Were Not Taking Warfarin

	Person-Years, n (%)		
	Women	Men	*P*
Total follow-up, person-years	6828	8850	
Age, y			<0.0001
<60	651 (9.5)	1844 (20.8)	
60–69	1134 (16.6)	1819 (20.6)	
70–80	2381 (34.9)	2883 (32.6)	
≥80	2662 (39.0)	2303 (26.0)	
Prior ischemic stroke	435 (7.1)	528 (6.0)	0.02
Diagnosed hypertension	3938 (57.7)	4235 (47.9)	<0.0001
Diagnosed congestive heart failure	1825 (26.7)	2350 (26.6)	0.85
Diagnosed coronary artery disease	1635 (23.9)	2741 (31.0)	<0.0001
Diabetes mellitus	968 (14.2)	1416 (16.0)	0.01
Prior gastrointestinal bleed	416 (6.1)	635 (7.2)	0.03
Prior hematuria	48 (0.7)	223 (2.5)	<0.0001
Prior other bleed	65 (0.9)	103 (1.2)	0.30
History of cirrhosis	52 (0.8)	154 (1.7)	<0.0001
Diagnosed dementia	465 (6.8)	409 (4.6)	<0.0001
Previous mechanical fall during prior hospitalization	571 (8.4)	374 (4.2)	<0.0001
Oral estrogen replacement therapy	1464 (21.4)	...	...
$CHADS_2$ score			<0.0001
0	1078 (15.8)	2102 (23.8)	
1	2090 (30.6)	2892 (32.7)	
2	2084 (30.5)	2327 (26.3)	
3	1086 (15.9)	1005 (11.4)	
4	349 (5.1)	351 (4.0)	
5	116 (1.7)	133 (1.5)	
6	24 (0.4)	39 (0.4)	

Results

The cohort included 5795 women and 7764 men. Women were generally older and more likely to have a history of stroke or hypertension but were less likely to have diagnosed coronary disease or diabetes mellitus than men (Table 1). Most men and women had $CHADS_2$ risk scores between 0 and 2; only a small proportion of the cohort was categorized in the highest-risk group (Table 1). Among women not taking warfarin, the proportion of person-time on oral estrogen therapy was 21.4% compared with 22.5% among women taking warfarin.

TABLE 2. Annual Unadjusted Incidence Rates of Thromboembolism Among Men and Women With Atrial Fibrillation Not Taking Warfarin Stratified by Known Risk Factors for Stroke and $CHADS_2$ Score*

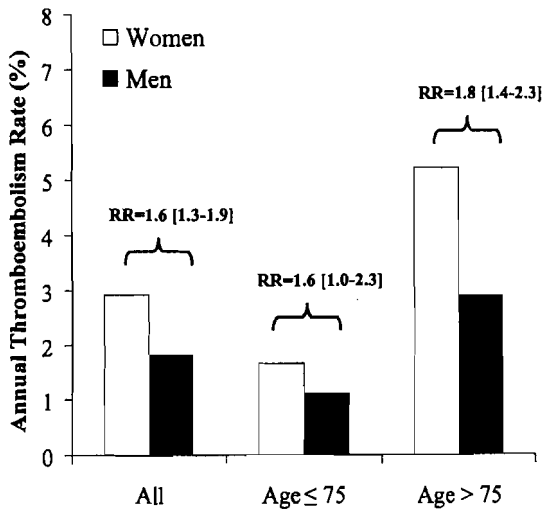
	Annual Thromboembolism Rate (95% CI)	
Risk Factor	Women	Men
Age ≥75 y	5.0 (4.3–5.7)	2.8 (2.3–3.4)
Prior ischemic stroke	9.7 (7.0–13.6)	7.3 (5.2–10.3)
Diagnosed hypertension	4.0 (3.4–4.7)	2.4 (2.0–3.0)
Diagnosed congestive heart failure	5.7 (4.7–6.9)	2.5 (1.9–3.2)
Diagnosed coronary artery disease	4.7 (3.8–6.0)	2.4 (1.9–3.1)
Diabetes mellitus	5.0 (3.7–6.6)	3.1 (2.3–4.2)
$CHADS_2$ score		
0	0.6 (0.2–1.2)	0.5 (0.3–0.9)
1	1.8 (1.3–2.4)	1.2 (0.9–1.7)
2	4.4 (3.6–5.4)	1.9 (1.4–2.6)
3	6.1 (4.8–7.8)	3.9 (2.8–5.3)
4	9.1 (6.2–13.3)	6.5 (4.2–10.0)
5	7.7 (3.6–16.5)	2.6 (0.8–8.1)
6	11.4 (2.5–51.9)	16.2 (7.4–35.6)

*$CHADS_2$ score calculated by assigning 2 points to prior stroke or transient ischemic attack and 1 point to any of the following risk factors: congestive heart failure, hypertension, age ≥75 years, and diabetes mellitus.

Risk of Thromboembolism Off Warfarin Therapy

During periods off warfarin therapy, we identified 394 valid thromboembolic events (369 ischemic strokes) over 15 494 person-years of follow-up. Women had higher annual incidence rates of thromboembolism off warfarin than did men (3.5% versus 1.8%), with an unadjusted rate ratio (RR) of 1.9 (95% CI, 1.6 to 2.4). These higher rates of thromboembolism among women were observed across various stroke risk factors and categories of the $CHADS_2$ Index[4] (Table 2). In a multivariable model controlling for stroke risk factors (age, prior stroke, diagnosed hypertension, congestive heart failure, coronary artery disease, diabetes mellitus, and estrogen replacement therapy), women had a greater independent risk of thromboembolism than men, with an adjusted RR of 1.6 (95% CI, 1.3 to 1.9). Results were similar when analyses were restricted to only ischemic strokes (adjusted RR, 1.5; 95% CI, 1.2 to 1.8). Of note, exposure to oral estrogen replacement therapy in multivariable analyses was not associated with a significantly increased risk of thromboembolism in women (adjusted RR, 1.0; 95% CI, 0.7 to 1.4). Thirty-day mortality after ischemic stroke did not differ significantly by patient sex (23.4% for men and 23.7% for women; $P=0.94$).

Women were at higher risk for incident thromboembolism than men at both younger and older ages (the Figure). The adjusted RR for women versus men was 1.6 (95% CI, 1.0 to 2.3) for those ≤75 years of age and 1.8 (95% CI, 1.4 to 2.3) for those >75 years of age. The difference between these 2 rate ratios was not statistically different ($P=0.38$ for the interaction of sex and age, dichotomized as >75 versus ≤75 years). The interaction between sex and age was also not statistically significant when age was coded as a continuous variable ($P=0.11$).



Annualized adjusted rate of thromboembolism (ischemic stroke and peripheral embolism) during off-warfarin periods among women and men with atrial fibrillation with age, prior stroke, hypertension, congestive heart failure, coronary artery disease, diabetes mellitus, and estrogen use controlled for. Age cutoffs of ≤75 and >75 years used as in the SPAF analysis.[3] RR indicates adjusted RR and 95% CI.

Effect of Warfarin Therapy in Women Compared With Men

The distribution of INR intensities was similar between men and women. In men taking warfarin, 26.8% of the person-time was spent at INR levels <2.0 and 62.7% between 2.0 and 3.0; in women, the proportions were 27.9% and 61.3% (These proportions were calculated excluding the 18% of total person-time for which INR was not available).

Among warfarin users, there were 204 thromboembolic and 167 major hemorrhagic events over ≈15 000 person-years of follow-up. Rates of thromboembolism on warfarin were lower than rates observed in patients off warfarin: annual unadjusted rates of thromboembolism on warfarin were 1.5% in women and 1.2% in men. After multivariable adjustment for other risk factors for stroke, warfarin therapy continued to be associated with a significant reduction in the rate of thromboembolism, with an adjusted RR of 0.4 (95% CI, 0.3 to 0.5) in women and 0.6 (95% CI, 0.5 to 0.8) in men. In multivariable models including patients both on and off warfarin therapy, the reduction in rates of thromboembolism with warfarin was larger in women than in men ($P=0.01$ for the interaction of sex and warfarin).

Rates of Major Hemorrhage on Warfarin Therapy

On warfarin, women had similar rates of all major hemorrhage compared with men (1.0% versus 1.1%; adjusted RR, 0.8; 95% CI, 0.6 to 1.1). Women were less likely than men to develop intracranial hemorrhage while on warfarin (0.36% versus 0.55%; adjusted RR, 0.5; 95% CI, 0.3 to 0.9). In multivariable models assessing predictors of intracranial hemorrhage that included patients both on and off warfarin therapy, warfarin therapy was associated with an increased risk for intracranial hemorrhage (adjusted RR, 1.6; 95% CI, 1.1 to 2.4), but women were not at greater risk for developing intracranial hemorrhage

TABLE 3. Relative Risk of Thromboembolism in Atrial Fibrillation Patients Not Taking Warfarin Comparing Women and Men Across Various Risk Stratification Schemes

	Time Off Warfarin, person-years	Women, %	Events, n	Relative Risk, Women vs Men
ATRIA cohort	15 494	44	369 Ischemic strokes, 25 peripheral emboli	1.6 (1.3–2.0)*
AFI[2]	3432	34	81 Ischemic strokes, 10 peripheral emboli	1.2 (0.8–1.8)†
SPAF[3]	3977	28	130 Ischemic strokes	1.6 (P=0.01)†
Framingham cohort[5]	2844	48	83 Strokes (both ischemic and hemorrhagic)	1.9 (1.2–3.1)†

*Reported as rate ratio.
†Reported as hazard ratio.

with warfarin therapy than were men (P=0.10 for the interaction term between sex and warfarin use).

Discussion

In this large cohort of patients with atrial fibrillation, women had higher rates of ischemic stroke and peripheral embolism while not taking warfarin than did men, even after adjustment for established clinical risk factors for stroke. Higher rates of thromboembolism among women were observed at both younger and older ages and across all stroke risk factor categories. The 30-day mortality rate following an ischemic stroke did not differ by sex, indicating that the increased risk of stroke faced by women was not due to the occurrence of less severe strokes.[15]

Warfarin appears to be at least as effective for women in reducing the risk of thromboembolism, if not more so, than in men. This observation in our cohort was also reported in the pooled analyses of 5 randomized trials of warfarin for atrial fibrillation.[2] Warfarin therapy did not pose a greater risk of major hemorrhagic complications in women. This was particularly true for the most important hemorrhagic complication, namely intracranial hemorrhage.

Some studies have shown that women, especially older women, are less likely to receive warfarin for atrial fibrillation.[7,16,17] Our findings indicate that women with atrial fibrillation face a higher absolute risk for thromboembolism independently of other risk factors and should gain more from anticoagulant therapy.

Available risk stratification schemes differ on whether female sex is a risk factor for atrial fibrillation–related thromboembolism (Table 3). We found that women had consistently higher rates of thromboembolism across all stroke risk factor strata and after multivariable adjustment. Our cohort analysis offers several advantages over prior studies. We had substantially greater numbers of person-years of follow-up and outcome events, providing a more powerful assessment. Our cohort is also more contemporary and based in a usual clinical practice setting, potentially yielding more generalizability. In comparison, the AFI and SPAF risk schemes were based on participants in randomized trials completed 10 to 15 years ago.[2,3] In contrast to the SPAF analysis, we did not find a significant interaction between patient sex and age >75 years. Another advantage of the large size of the cohort is that it allowed us to assess whether women faced an increased risk of warfarin-associated hemorrhage, particularly intracranial hemorrhage. Prior studies did not observe sufficient numbers of intracranial hemorrhages to assess warfarin-sex interactions. This is especially important because the health consequences of intracranial hemorrhage are worse than those resulting from the ischemic strokes we seek to prevent through anticoagulation.[18,19] In our ATRIA cohort, warfarin was not more dangerous in women than in men.

The mechanism behind the observed difference in atrial fibrillation-related thromboembolism risk between men and women is unclear. Atrial fibrillation is associated with higher levels of prothrombotic factors, endothelial dysfunction, and markers of platelet activation,[20–23] but sex-related differences in these factors have not been well characterized. Interestingly, women with atrial fibrillation may have higher levels of prothrombin fragment F1.2,[24] von Willebrand factor,[22] and tissue plasminogen activator antigen,[25] but studies have not clearly linked these factors to an increased risk of stroke in atrial fibrillation. It also remains to be seen whether differences in left atrial structure and function[26,27] contribute to differential thromboembolism risk by sex. Although estrogen replacement therapy has been reported to increase risk of ischemic stroke among postmenopausal women,[13] it was not a significant risk factor in our study.

Our study has several limitations. We lacked data on potential differences in left ventricular systolic function and blood pressure control between men and women, factors shown to affect stroke risk.[5,28] Although we controlled for the diagnosis of hypertension, we could not adjust for individual patients' blood pressure levels. It is noteworthy, however, that these factors were also not used in the AFI and $CHADS_2$ risk schemes. We did not have comprehensive information on the use of aspirin in our pharmacy database because many patients used nonprescription forms of aspirin. We addressed this shortcoming in a previous review of the medical charts of 232 randomly sampled patients in our cohort who were not taking warfarin.[8] In these nonusers of warfarin, 38% of women and 56% of men were recorded as taking aspirin. Assuming that aspirin reduces thromboembolic event rates by 21%,[29] differential use of aspirin between men and women would not materially change our original estimate. Thus, it is unlikely that variation in aspirin use between men and women explains their differing rates of thromboembolism. Finally, we note that stroke rates in our cohort were generally lower than those reported in other earlier studies. The reason for these lower rates is not clear but may reflect a somewhat healthier, insured population of patients, Because we lacked information on individual patients' blood pressure measurements, we were unable to determine whether control of hypertension was better in our cohort than in other populations. In addition, we required that each stroke event be validated by chart review. It is possible that our search strategy

missed some stroke events because searching for cerebrovascular diseases using ICD-9 codes may not be highly sensitive.[30,31]

In conclusion, women have a higher risk than men for atrial fibrillation–related thromboembolism at both younger and older ages that is independent of the presence of other risk factors for stroke. Furthermore, warfarin therapy appears to be at least as effective in women as in men in preventing thromboembolism. Finally, women do not have a higher risk than men for intracranial or other major bleeding events associated with warfarin. On balance, the overall net benefit of warfarin therapy for atrial fibrillation appears to be greater in women compared with men. Our findings indicate that female sex is an important factor supporting the use of anticoagulant therapy in patients with atrial fibrillation.

Acknowledgments

This work was supported by Public Health Services research grant AG15478 from the National Institute on Aging and the Eliot B. and Edith C. Shoolman Fund of Massachusetts General Hospital.

Disclosure

Dr Hylek is currently a principal investigator on 2 industry-sponsored research grants limited to analyses of completed databases. Bristol-Myers Squibb, the source of one of these grants, makes the brand-name warfarin (Coumadin). AstraZeneca, the source of the other grant, manufactures the brand-name ximelagatran Exanta. Dr Hylek also served as a panel participant at a symposium sponsored by AstraZeneca.

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. Hart R, Pearce L, McBride R, Rothbart R, Asinger R. 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.
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;285:2864–2870.
5. Wang TJ, Massaro JM, Levy D, Vasan RS, Wolf PA, D'Agostino RB, Larson MG, Kannel WB, Benjamin EJ. 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.
6. Go AS, Hylek EM, Phillips KA, Borowsky LH, Henault LE, Chang Y, Selby JV, Singer DE. Implications of stroke risk criteria on the anticoagulation decision in nonvalvular atrial fibrillation: the Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. *Circulation*. 2000;102:11–13.
7. Go AS, Hylek EM, Borowsky LH, Phillips KA, Selby JV, Singer DE. Warfarin use among ambulatory patients with nonvalvular atrial fibrillation: the Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. *Ann Intern Med*. 1999;131:927–934.
8. 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.
9. Arellano MG, Petersen GR, Petitti DB, Smith RE. The California Automated Mortality Linkage System (CAMLIS). *Am J Public Health*. 1984;74:1324–1330.
10. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thrombos Haemostas*. 1993;69:236–239.
11. Ettinger B, Pressman A. Continuation of postmenopausal hormone replacement therapy in a large health maintenance organization: transdermal matrix patch versus oral estrogen therapy. *Am J Managed Care*. 1999;5:779–785.
12. Singer DE, Albers GW, Dalen JE, Go AS, Halperin JL, Manning WJ. Antithrombotic therapy in atrial fibrillation: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;126:429S–456S.
13. Wassertheil-Smoller S, Hendrix SL, Limacher M, Heiss G, Kooperberg C, Baird A, Kotchen T, Curb JD, Black H, Rossouw JE, Aragaki A, Safford M, Stein E, Laowattana S, Mysiw WJ, Investigators WHI. Effect of estrogen plus progestin on stroke in postmenopausal women: the Women's Health Initiative: a randomized trial. *JAMA*. 2003;289:2673–2684.
14. Levine MN, Raskob G, Beyth RJ, Kearon C, Schulman S. Hemorrhagic complications of anticoagulant treatment: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;126(suppl):287S–310S.
15. Longstreth WT Jr, Bernick C, Fitzpatrick A, Cushman M, Knepper L, Lima J, Furberg CD. Frequency and predictors of stroke death in 5,888 participants in the Cardiovascular Health Study. *Neurology*. 2001;56:368–375.
16. Humphries KH, Kerr CR, Connolly SJ, Klein G, Boone JA, Green M, Sheldon R, Talajic M, Dorian P, Newman D. New-onset atrial fibrillation: sex differences in presentation, treatment, and outcome. *Circulation*. 2001;103:2365–2370.
17. Sudlow M, Thomson R, Thwaites B, Rodgers H, Kenny RA. Prevalence of atrial fibrillation and eligibility for anticoagulants in the community. *Lancet*. 1998;352:1167–1171.
18. Hylek E, Singer D. Risk factors for intracranial hemorrhage in outpatients taking warfarin. *Ann Intern Med*. 1994;120:897–902.
19. 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.
20. Gustafsson C, Blomback M, Britton M, Hamsten A, Svensson J. Coagulation factors and the increased risk of stroke in nonvalvular atrial fibrillation. *Stroke*. 1990;21:47–51.
21. Minamino T, Kitakaze M, Asanuma H, Ueda Y, Koretsune Y, Kuzuya T, Hori M. Plasma adenosine levels and platelet activation in patients with atrial fibrillation. *Am J Cardiol*. 1999;83:194–198.
22. Conway DS, Heeringa J, Van Der Kuip DA, Chin BS, Hofman A, Witteman JC, Lip GY. Atrial fibrillation and the prothrombotic state in the elderly: the Rotterdam Study. *Stroke*. 2003;34:413–417.
23. Inoue H, Nozawa T, Okumura K, Jong-Dae L, Shimizu A, Yano K. Prothrombotic activity is increased in patients with nonvalvular atrial fibrillation and risk factors for embolism. *Chest*. 2004;126:687–692.
24. Feinberg WM, Pearce LA, Hart RG, Cushman M, Cornell ES, Lip GY, Bovill EG. Markers of thrombin and platelet activity in patients with atrial fibrillation: correlation with stroke among 1531 participants in the Stroke Prevention in Atrial Fibrillation III study. *Stroke*. 1999;30:2547–2553.
25. Wang TD, Chen WJ, Su SS, Su TC, Chen MF, Liau CS, Lee YT. Increased levels of tissue plasminogen activator antigen and factor VIII activity in nonvalvular atrial fibrillation: relation to predictors of thromboembolism. *J Cardiovasc Electrophysiol*. 2001;12:877–884.
26. Benjamin EJ, D'Agostino RB, Belanger AJ, Wolf PA, Levy D. Left atrial size and the risk of stroke and death: the Framingham Heart Study. *Circulation*. 1995;92:835–841.
27. Pritchett AM, Jacobsen SJ, Mahoney DW, Rodeheffer RJ, Bailey KR, Redfield MM. Left atrial volume as an index of left atrial size: a population-based study. *J Am Coll Cardiol*. 2003;41:1036–1043.
28. Stroke Prevention in Atrial Fibrillation Investigators Committee on Echocardiography. Transesophageal echocardiographic correlates of thromboembolism in high-risk patients with nonvalvular atrial fibrillation. *Ann Intern Med*. 1998;128:639–647.
29. Atrial Fibrillation Investigators. The efficacy of aspirin in patients with atrial fibrillation: analysis of pooled data from three randomized trials. *Arch Intern Med*. 1997;157:1237–1240.
30. Benesch C, Witter DM Jr, Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy of the *International Classification of Diseases* (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. *Neurology*. 1997;49:660–664.
31. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford M, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. *Med Care*. 2005;43:480–485.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biomedical Informatics 38 (2005) 220–228
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

Telephone-linked care for physical activity: A qualitative evaluation of the use patterns of an information technology program for patients

Ramesh Farzanfar*, Sophie Frishkopf, Jeffrey Migneault, Robert Friedman

Medical Information Systems Unit, Boston University Medical Center, 720 Harrison Avenue, Suite 1102, Boston, MA 02118, USA

Received 6 October 2004
Available online 8 December 2004

Abstract

Automated health behavior interventions that involve discretionary use by patients or consumers over extended periods of time are becoming more common and it is generally assumed that adherence to the recommended schedule is related to the impact of the system on users. Yet reasons for use or non-use of such systems have not been carefully explored. An understanding of factors that influence people to use, not use, or underutilize these automated behavioral change and self-care management systems can help in designing systems that are more effective and acceptable to users. Using qualitative research methods, this study explored the experiences of 45 users of a multiple-contact health promotion application with the goal of understanding the major factors that affect patterns of use (frequency of and duration of contact). The in-depth exploration of users' perceptions and views made possible by the qualitative research methods revealed a number of important themes. Reported reasons for underutilization or non-use were found to be both user-related and system-related. User-related reasons encompassed personal and individual events that prevented or impeded system utilization. System-related reasons included those that related to the medium itself as well as the content of the application. The qualitative methods employed in this study created a forum through which users' feedback could be fully explored and then synthesized to assist in the improvement of this and other automated health behavior interventions.

Keywords: Evaluation; System utilization; Qualitative research methods

1. Introduction

The use of information technologies in health care is beginning to change the health care industry in important and perhaps irreversible ways. By facilitating patient education, patient lifestyle change, and self-care as well as patient–provider communications, these technologies promote patients' involvement in their own care, assist in health care delivery, and improve patient outcome [1]. Evidence suggests, however, that a substantial number of programs that use these new technologies do not reach their potential, because of underutilization or non-use by providers, patients, and consumers (defined as users who are healthy) [2–4]. The adoption and diffusion of technological innovations, to a great extent, depend on a critical factor: *utilization* [5]. In fact, it is believed that the societal value of an innovation is ultimately determined by the repetition and range of "use" that the system receives [5].

Even among people who use a program, there are significant variations in patterns of use. These differences are an important issue to consider when evaluating health promotion and disease prevention applications, particularly those that involve discretionary use by patients and consumers over extended periods of time. Many automated behavioral change and self-care management interventions are designed to be utilized over

* Corresponding author. Fax: +1 617 638 8858.
E-mail address: rfarzanf@bu.edu (R. Farzanfar).

1532-0464/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2004.11.011

time and consequently, it is assumed that to achieve and maintain the targeted effects user adherence to the intervention schedule is necessary. However, reported research on the variations in use patterns of these systems is scant. Thus, an effort to identify and explore the factors that influence people to use, not use, or underutilize these systems would provide an important perspective for evaluation of these systems from the users' viewpoint. This in turn will help in designing systems that are more acceptable to users, and perhaps more effective.

Using qualitative research methods, we explored the experiences of users of a multiple-contact health promotion application with a focus on understanding factors that affected patterns of use (frequency and duration of contact). We recruited 82 healthy adults to use a physical activity promotion computer telephony program. Subsequently, 45 individuals were selected for in-depth interviews based on their use patterns. The results of the in-depth interviews provided insights into the factors that contribute to *use*, *non-use*, and *underutilization* with significant implications for design of the physical activity telephony system, in particular, and health technology systems used by patients, in general.

2. Methods

2.1. Telephone-linked care technology

Telephone-linked care (TLC) is a computer telephony technology with applications for behavioral change and chronic disease management. Most TLC applications are designed to be used repeatedly over time. Through totally automated telephone conversations, TLC uses digitized human speech to *talk* with patients; and either through touch tone or speech recognition technologies *understands what the patient communicates*. TLC asks questions, gives feedback based on the user's response and embedded logic, and provides education and counseling for a targeted health behavior. Either TLC or the user can initiate a conversation. In the TLC behavioral change programs, patients are asked to contact TLC, depending on the application, anywhere from daily to monthly for a period that varies from 1 to 12 months. TLC stores the user's feedback in a database, based on which current and future TLC conversations are carried out. These responses also provide the information for reports that are sent to users and/or to their providers.

TLC behavior change applications have been applied to changing dietary behavior [6], promoting physical activity [7], helping cigarette smokers quit [8], and promoting medication adherence in patients with hypertension [9] and depression [10,11] as well as promoting regular screening mammography. TLC chronic disease applications have been developed for chronic obstructive pulmonary disease (COPD) [12], coronary heart disease [13], and diabetes mellitus [13]. Two other TLC applications have been developed which help clinicians better manage cancer patients who are receiving chemotherapy [14,15]. Although pattern of use has varied across these applications, most TLC systems that have been fully evaluated have generally been effective and well-accepted.

2.2. Telephone-linked care for physical activity

This study evaluated an interactive health promotion technology application, TLC-physical activity (TLC-PA), which promotes moderate-intensity physical activity like brisk walking, to a healthy general adult population. The program's goal was the recommendation set by the Center for Disease Control (CDC) and the American College of Sports Medicine (ACSM) that all adults should engage in at least 30 min of moderate-intensity physical activity on most days of the week. To promote the regular attainment of physical activity, TLC-PA employed behavior change strategies derived from the transtheoretical model (TTM) of behavior change [16,17]. Stage of change (motivational readiness) is the central organizing construct of the TTM. The following five stages of change are integrated into the design of TLC-PA: (1) Precontemplation (not thinking of meeting the physical activity goals in the next 6 months); (2) contemplation (thinking about becoming physically active within 6 months); (3) preparation (intention to achieving activity recommendations in the next 30 days); (4) action (being sufficiently physically active for less than 6 months); and (5) maintenance (being physically active at or above recommendations for more than 6 months). During each TLC-PA conversation, the system assesses the user's current stage. The system then selects behavior change strategies for use during the conversation that are based on the user's stage. The theoretically based tailoring was expected to increase the relevance of the messages to each individual participant and thus contribute to the effectiveness of the system. It is generally believed that tailoring is an effective strategy when health messages are directed at a diverse population [18].

The duration of the study was 3 months and the study participants were asked to call the system two times per week. All participants met with the study staff prior to using the system. Based on the information users provided during this meeting, they were assigned to an appropriate stage of physical activity readiness. Users initiated all calls to TLC-PA.

At the beginning of each telephone conversation, the system begins with a salutation and information about the TTM stages and what it means to be in one stage versus another. The system describes the "*meaning of exercise*" (i.e., the definition of moderate or greater

intensity physical activity) at the beginning of each contact by saying: "just to make sure that we are talking about the same thing, when I talk about exercise I mean structured physical activity that makes you breath hard or break a sweat. This does not include things like housework, golfing using a cart, or walking around the office. This does include brisk walking, bicycling, or playing sports." The system then asks about the user's current level of physical activity, defined as the number of days and minutes/day during the previous week the user engaged in "exercise" and then stage of change.

The user's stage of readiness determined the content of the TLC-PA conversations. For example, those who were in the precontemplation stage were given information about the benefits of physical activity. This included such topics as "impact of physical activity on blood pressure," "prevention of breast cancer," "prevention of diabetes," "lowering stress," and more. Similarly, those in the contemplation and preparation stages were provided with information on the benefits of physical activity as well as suggestions for overcoming barriers to physical activity and were encouraged to set weekly exercise goals. Users in the *action* stage were given an option to hear the information on the barriers if there had been a decline in activity level, and were encouraged to increase their level.

3. Study design

We conducted in-depth interviews to evaluate reasons for the participants' use patterns of TLC-PA. The interviews explored: (1) how people felt about the system in general, and (2) why some people did not use or underutilized the system. The context of users' experiences, i.e., their lifestyle and cultural norms, were also queried.

3.1. Study participants

Eighty-four volunteers were recruited and found eligible by screening of whom 80 completed the study (two withdrew). Individuals were excluded if they had a serious medical condition or who were in the maintenance stage for physical activity. The study population included 48 women (58%), 29 blacks (35%), 7 Asians (8%), 3 Hispanics (3%), and 6 "Other" (7%). Over one-fourth (23%) of the participants were married while 37% were employed and nearly 63% had education beyond high school. The participants' age ranged from 21 to 74 with the mean age of 45.

3.2. Qualitative evaluation methods

After 4 weeks of using the system, the utilization pattern of each participant was classified into one of five categories. Four of these utilization patterns closely resembled those identified in other TLC studies. A new fifth category was defined based on the utilization pattern observed in this study of the TLC-PA system for a subset of participants who ended each conversation with the system by hanging up before the call was complete. The final five utilization categories, along with their size and the number interviewed, are as follows: (1) $\geqslant 80\%$ adherence to the call schedule ($N = 8$, 7 interviewed), (2) Intermittent but continuous use ($N = 18$, 11 interviewed), (3) Discontinued use (consecutive use of the system for two or more times after which the calling ceased completely) ($N = 36$, 16 interviewed), (4) Non-use or one-time-use ($N = 14$, 7 interviewed), and (5) Incomplete calls (one or more) ($N = 6$, 4 interviewed).

The participants were interviewed over time, with data collection (interviews) and analytic work (coding and interpretation of interviews, described below) occurring simultaneously in accordance with standard qualitative research methodology [19]. In qualitative research study samples are usually small and the selection method is purposive rather than random. Furthermore, the sample size is not predetermined, with recruitment for a particular cohort (in our study, each utilization group) ending when there are no longer any substantial new findings from the interviews. After this point there is little "incremental learning" as the researchers observe "phenomena seen before." Methodologically, this process is called "information saturation" or "redundancy" [19].

All members of the smaller utilization groups (groups 1, 4, and 5) were invited for in-depth interviews. As demonstrated above, the proportion who accepted the invitation varied considerably with 88% of the high adherence group (group 1) agreeing to be interviewed, whereas only 50% of group 4 agreed to be interviewed. Given the nature of these groups, this variation is to be expected. In each of these groups, however, saturation was fully achieved. In the larger groups (groups 2 and 3) saturation was judged to have occurred after 11 and 16 interviews, respectively.

In-depth interviews were conducted by the first and second authors together. Interviews took place at the date and time most convenient for the participants. The interviews followed a "general interview guide approach" in which a set of predefined issues were explored with the study participants. These issues were written in an interview guide that served as a question check list for the interview, to ensure that all relevant topics were covered. The interviews lasted between 20 and 45 min. The issues that were discussed during the in-depth interviews dealt with the following: (1) specific features and components of the system, such as the TLC-PA's voice, its tone, duration of the conversation, etc.; (2) participants' overall impressions such as their

likes/dislikes, satisfaction with the program, opinions about the program's helpfulness, their initial expectations, degree to which their expectations were met, possible behavior change effects, etc.; (3) reasons for participants' patterns of use including questions about why they used the system in a certain way. As we were interested in both negative and positive opinions about the TLC-PA program, addressed reasons for utilization and non-utilization with all participants, including those who were highly adherent to the planned twice a week calling schedule.

The interviews were tape-recorded, transcribed, coded, and stored both in a database and in hard copy. A systematic coding of transcripts by two independent coders identified 27 *constructs* or themes of interest. A secondary analysis of the coded transcripts condensed the constructs to 10. We also conducted a content analysis of the transcripts in which these *constructs* or *themes* were counted for frequency of occurrence (see Table 1). Constructs were defined as those that either shed light on the users' opinions and views about the system or helped reveal the reasons or provide explanations for certain behaviors, including the participants' TLC-PA use patterns and their physical activity behavior.

4. Results

4.1. ⩾80% Adherence group

Individuals in this category ($N = 8$, 7 interviewed) adhered to the call schedule ⩾80% of the time. Among the four individuals (57%) who reported behavior change, i.e., an increase in physical activity levels, only 2 (29%) reported benefits (i.e., description of outcome benefits such as a decrease in blood pressure, decrease in cholesterol, weight loss, general well-being, etc.). This group also had the highest ratio (86%) of individuals (6) who complained about too much repetition in the content of TLC-PA. Of the seven individuals interviewed four had positive views about the system (57%) while the other four felt neither positive nor negative. Two of the individuals in this group said that the program did not enhance their physical activity levels. These individuals, however, used the program to keep physically active and one said that using the system kept him aware of the amount of time that he was devoting to exercise. Of the seven individuals, two said that they used the system regularly because they had committed themselves to the study.

Except one person, these regular users were critical of different aspects of the system. One said that he used the system because he was curious and wanted to learn new information about physical activity and health, however, he was disappointed about the amount of repetition and the lack of new information. Others also complained about the repetition, the length of the conversations, and problems with being understood. A woman who maintained she benefited moderately from using the system said she was worried that we were going to ask her to use the program forever.

4.2. Intermittent user group

The intermittent users ($N = 18$, 11 interviewed) were those who used TLC-PA throughout the 2-month test period, but who called less then 80% of the time. Their average utilization was 52% (range 31–77%). The reasons provided by these individuals for not fully utilizing the system were identical to those expressed by users in other groups: life crises, being away, being too busy (including working too hard, being too tired), and health problems. Forgetting to use the system as a result of the reasons referred to above was mentioned frequently.

Surprisingly, these participants, as a group, were neither dissatisfied with TLC-PA nor perceived it to be without benefit to them. In fact, individuals in this group had the highest ratio (91%) of satisfied users (10 individuals) and better reported outcomes both in terms of physical activity levels (9 individuals—82%) (Table 1) and perceived benefits (8 individuals—73%). Even

Table 1
Interviewed participants: summary utilization and user response ($N = 82$)

	⩾80% $N = 8$ (7 Interviewed)	Intermittent $N = 18$ (11 Interviewed)	Block users $N = 36$ (16 Interviewed)	Non-use/one time use $N = 14$ (7 Interviewed)	Partial use $N = 6$ (4 Interviewed)
Positive Opinion	4 (57%)	10 (91%)	3 (19%)	1 (14%)	0
Behavior Change	4 (57%)	9 (82%)	4 (25%)	2 (29%)	2 (50%)
Reported Benefit	2 (29%)	8 (73%)	0	2 (29%)	2 (50%)
Failure and Avoidance	0	2 (18%)	6 (38%)	0	0
Helpful Information	5 (71%)	9 (82%)	5 (31%)	0	0
System as a Monitor	3 (43%)	7 (64%)	0	0	0
System as a Motivator	2 (29%)	4 (36%)	3 (19%)	0	0
Too Much Repetition	6 (86%)	6 (55%)	10 (63%)	0	0
Too Long	2 (29%)	0	10 (63%)	0	0
Voice Recognition Problems	1 (14%)	4 (36%)	6 (38%)	0	0

though utilization was irregular, some individuals in the intermittent group continued to use the system after we had conducted our in-depth interviews with them and their participation in the study was effectively over. Only one person expressed a negative opinion about TLC-PA in this group.

An important aspect of the physical activity behavior of the individuals in this group was the fact that their call pattern mirrored their exercise pattern. These individuals went for walks, to the gym, or performed other activities and subsequently reported the results to TLC-PA. As one of the participants described it: "I usually do it [calling the system] after I exercise, you know, take my walk... After I do it, it is kind of like gratification that I can call in and tell somebody that I did it." It is compelling that in some cases when a person had a "good week" in terms of physical activity behavior, then that person would make more calls to TLC-PA. One participant who exercised a lot during a particular week called the system six times for that week! When asked why he called so many times, he said because he "was doing so good." Similarly, another individual who made additional calls during a particular week said that the reason was because during that period he exercised more.

One important theme generated from the in-depth interviews of intermittent users revolved around the concept of *control*. They described their use patterns as being determined by themselves, not the designers of TLC-PA. For at least half of the intermittent users taking control of system utilization also reflected taking control of their exercise regimen. A man in this group in fact used the study's Users' Guide as a symbol to exercise such control. During the day when he planned to exercise and subsequently call the system, he would place his Users' Guide on his desk and at other times the Guide would "get stuck underneath somewhere." Control of both the exercise regimen and call schedule was thus instrumental in impeding or facilitating system utilization as in several cases the two went hand in hand.

Another theme that emerged was the central theme of being "monitored" as a motivator of behavior. The individuals in the intermittent group felt that the system was watching them and this perception motivated them to engage in physical activity. One woman described it this way: "You are more aware and more responsible. I think we need to be accountable. It's like answering to a higher... [authority]" And, a man who started going to the YMCA upon his participation said: "It kept me in check." In fact, seven of the individuals (64%) interviewed in this group considered the TLC-PA as an effective monitor. Monitoring in turn seems to have generated a certain degree of anxiety in a few individuals who said they exercised because they wanted to report that they had done well. For example, a woman who accomplished most of her exercise goals commented: "The next time I called, I wanted to be able to say I did this." This woman subsequently elaborated that the system changed her [physical activity] behavior "a little bit" because she wanted to report that she had accomplished her tasks. Another woman also commented, "I want to report accurately for myself as well. Umm, I felt re-encouraged that *I didn't get penalized.*" [Emphasis added.]

It is of great interest to us that the system's response might have seemed "penalizing" to this woman as the designers had done their utmost to ensure that the system's responses to the unaccomplished goals were polite, pleasant, and supportive with a positive tone. For example, "It is great that you are doing some exercise, but you did less than your goal. To receive the maximum benefits from an exercise program you need to gradually work yourself up to exercising at least 4 days per week for at least 30 minutes per day. Don't feel too badly. I will set another goal with you later in the call. Use the rest of this call to increase your commitment to regular exercise." The anxiety that these study participants felt about accomplishing their physical activity goals was in fact constructive as it reinforced their resolve and thus helped them achieve their goals. A young woman tried to describe her feelings this way: "I feel obligated. It's like something- I don't know what it is. Before, I didn't go to gym because I did too much and didn't have time, but now I make the time... I don't know how they [TLC-PA] make me like that, you know."

Finally, despite the overall positive opinions, six individuals (55%) in this group complained about too much repetition, while four (36%) had problems being understood by TLC-PA.

4.3. Non-users or one-time-users group

Of the 82 participants, 14 did not use the system at all or used it only once. We have put non-users and one-time users into a single utilization group because their reactions to the system were remarkably similar. Of note, it was difficult to arrange interviews for subjects in this group. We were eventually able to interview seven individuals among the 14, but two had disconnected their telephones and five did not return repeated calls. Of the seven individuals we interviewed, four claimed that they had actually used our system a few times although the system's log files did not show any contacts (three of them called TLC-PA for the first time on the day they were to meet with us for the in-depth interview). Two of the seven interviewees had called TLC-PA once; and two explained that they had lost the Users' Guide and thus did not have the information necessary (for example, the telephone number or the password) to use the system.

The results of the in-depth interviews with individuals in this group demonstrated that the reasons for non-use and one-time use mostly overlapped. They identified

personal events or situations in the users' lives such as the death of a loved one, getting robbed, financial distress (e.g., unemployment), illness (personal or family), being away (on a vacation or a business trip), working too hard, being too busy, and forgetting as the reasons for not calling TLC-PA. During our interviews, the majority of the individuals in this group presented their lives as too hectic, too disorganized, or too eventful for them to use a health promotion and disease prevention program. Most of those affected by "life events" were women, some of whom were parents (one a single parent) and all were experiencing financial problems. One woman said that she did not have water and heat in her apartment and that a close friend of hers was in the hospital dying of cancer. Another woman told us that she could not spend time to use the system because it took too long and she had to "deal with too much lately."

Devastating life events, however, were not the sole reasons for non-utilization. Among non-users were two participants who were not experiencing life crises. For these individuals using the system was not a priority. One woman commented: "this was not as important as other things in my life" and kept repeating "I forgot." Surprisingly, even though this individual had never used the system, she claimed that the idea of being in the study was a sufficient incentive to bring about behavior change. She said that she was now walking between 5 to 7 miles/day and had lost more than 70 lb. There was no way for us to verify the accuracy of this information. She said that she had pictures that could prove her claim. We gave this young woman a chance to use the system by explicitly asking her to use it and thus evaluate it for us. She enthusiastically accepted but she made no calls. We heard similar comments from another individual who had made only one call. Most individuals in this group had difficulty articulating the reasons why they could not reserve 15 minutes a few times a week to a health promotion program considering their perceived and reported need. One person chalked it up to "laziness."

4.4. Discontinued use group

A fourth group of TLC-PA users called the system for a period of time (calling from 2 to 14 times) but then stopped and never called back. This group had the highest number of participants (36). We interviewed 16 individuals from this group (nine women).

Two important themes emerged from the in-depth interviews with these participants: (1) most of the reasons for discontinuing TLC-PA were system-related, and (2) these individuals had negative opinions of the system as 10 (63%) complained of too much repetition, and 10 (63%) felt that the calls were too long, while six (38%) had problems being understood.

4.4.1. Failure and avoidance

Some of the interviews with participants in the discontinued group suggested an intriguing combination of complex emotional and psychological reactions to the system's content. We learned that six participants (38%), five of whom women, stopped using the system because they were reluctant to report that they had not exercised. These individuals stopped using the system once they had *failed* to accomplish the physical activity goals they had negotiated with TLC-PA during the previous conversation. If they had not exercised, they were reluctant to use the system to report that they had *not* accomplished their physical activity goal. This reluctance involved wanting to avoid reporting an unaccomplished goal to the system, and concern about the system' response to such as admission.

We were told by these individuals that having negotiated goals for physical activity and then having to report to the system that the goals were not met, felt like an admission of failure to an authority figure. One young woman, who attributed the problem to the tone of the TLC-PA's voice, said that it reminded her of her mother's admonishments. Another woman said that reporting unaccomplished physical activity goals to the system was like having to show a "bad report card" to her father. The following example is a remarkable testimony that speaks to this experience. "Cause, in the way you are talking to the system, it expects you to do better each day, you know. So, every day the system wants you to do a little bit better. It was an encouragement. But, when you didn't meet that goal, you are not happy with it. I wasn't too happy because I'd like to meet that goal... And, then I didn't-, I stopped. I don't know how to describe it. It's a feeling kind of like you failed; you failed a goal. Psychologically but then it's a study; you are just talking to the computer..."

There was one man in this group who said he felt uneasy about unaccomplished goals to TLC-PA. This individual called the system 10 times before stopping. It seems that for this individual the unease and anxiety initially worked in a positive manner, helping to increase his physical activity levels. However, in the long run, he could not keep up with the goals that he had negotiated with the system, and thus he stopped using the system altogether: "I tried to look at it objectively from the very beginning. I tried to walk more you know. Is it encouraging me or is it not? Is this lady on the recording going to embarrass me if I don't? ((laugh)). It's, uh, the recording – the lady on the recording said, 'do you intend to exercise four times a week?' And I said, 'yes.' So, now I had to live up to it. That's what encouraged me to do the exercise. Each time I came back it was like a building pattern. Do you remember what we did last time?' How can a recording be so smart? With the recording, there is no way of reversing what you promised the week before. And that is the part that kind of

frustrated me. The lady said, 'I am sorry to hear that.' ((laugh)) [He is referring to the system's response when he reported unmet goals.]

We have no self-evident explanation as to why there were more women among those who expressed anxiety about reporting unaccomplished tasks. Possible explanations may include gender differences in relation to exercise achievement and to negative judgment by others and/or the women felt more at ease with the two female interviewers than the men did. It is thus possible that the women more openly expressed their feelings, while the male participants were more reserved and reluctant to express a perceived *weakness* to the two female interviewers.

4.4.2. TLC-PA as a tailored intervention: "this is not for Me!"

Even though TLC-PA, based on the TTM, provided behavioral feedback tailored to stage of change for physical activity, an interesting and significant point brought up by several individuals across utilization categories, particularly those in the discontinued group was that TLC-PA was not responsive to their particular needs and personal lifestyles. Probing this issue further revealed that these participants believed the system was not tailored to their personal lifestyle and did not sufficiently address their perceived needs. The TTM cognitive and behavioral processes, as applied to physical activity, addressed such topics as confidence enhancement strategies, information on physical activity benefits, and ways to overcome barriers to physical activity. Some participants pointed out that receiving information about the benefits of physical activity, or overcoming barriers to physical activity, though acceptable and perhaps useful, was not exactly what they had in mind when they joined the study. Several of the participants contrasted TLC-PA with a personal trainer and used this analogy to describe their perceived needs. They maintained that TLC-PA's strategies did not help them engage in physical activity and that only a more personalized and tailored program, structured and planned specifically based on their personal exercise needs and requirement, would be helpful to them.

It seems that the duration and length of the conversation was particularly irritating to those who used cell phones (as the conversation used up their valuable minutes) and also those who called from their work, as the calls took 10–15 or sometimes 20 minutes during which they had to keep saying "yes," "no," etc.

4.5. Incomplete use group

These were users ($N = 6$) who made from 1 to 3 incomplete calls to TLC-PA (calls in which they hung up in the middle of the call); we interviewed four of them. By and large, these individuals disliked the system. Only one woman, who had used the system incompletely on the day of her interview, expressed some enthusiasm. The other three participants had negative views referring to system-related issues such as repetition of content and difficulty being understood and that the program did not address their particular concerns with regard to physical activity. One person who had made three incomplete calls said that she used the system out of sheer guilt but that she "could not stand it."

5. Discussion

Our study demonstrated that both user-related and system-related reasons accounted for non-use or underutilization. Among the themes that emerged from the in-depth interviews, one with the most important design implications was the concept of "failure and avoidance." "Failure and avoidance" was a negative response that was articulated by users, particularly in the discontinued group, towards the "monitoring" and "goal-setting" functions of TLC-PA. As was described, the anxiety associated with reporting unaccomplished goals did not always impede physical activity as our interviews with the individuals in the intermittent group demonstrated. In fact, in the intermittent use group, such anxiety was constructive and helped induce behavior change by motivating participants to increase their physical activity. However, it is the negative responses to "monitoring" and "goal-setting" that reveal the most about underutilization.

One possible explanation for this phenomenon may reside in the social psychological theory of *social facilitation*. Social facilitation occurs when an individual either enhances or diminishes a particular behavior in the presence of another [20]. Many experiments with both humans and animals demonstrate the consequences of such a presence [21]. Zajonc classifies research in social facilitation under two different paradigms: audience effects and co-action effects. Audience effects refer to the impact of the mere presence of others on behavior while co-action effects refer to the simultaneous involvement in action by all parties in full view of each other [22]. Experiments carried out by Zajonc demonstrated that in the audience effects paradigm, the response to the presence of others varied based on the difficulty or simplicity of the task. As a result, in the presence of a spectator if the tasks are easy, the response is enhanced. However, the response is diminished if the tasks are difficult. Similarly, through an experiment in which people completed a task alone, in front of two observers, or in the presence of two persons who were blindfolded, Cottrell demonstrated that the presence of others created heightened arousal. Tasks performed in front of the observers were negatively affected [diminished]; being alone or in the presence of

blindfolded individuals had no impact on performance [23].

Studies carried out with animated characters and other computer interfaces demonstrate that social facilitation does occur in the presence of automated systems [24,25]. Studies conducted on automated monitoring of work performance also confirm the impact of social facilitation [26]. For example, Rickenberg and Reeves [24] carried out an experiment in which they tested the impact of animated characters on user anxiety and task performance in a Web environment. The authors concluded that when a "social actor communicates an intention to monitor someone's work" there is an enhanced arousal and a diminished performance. In fact, "When the monitoring is obvious, thoughts and behavior change; there is more anxiety and less accurate performance of complex tasks." The finding that automated monitoring evokes anxiety in certain tasks and influences behavior has also been corroborated by other studies [27,28]. Our findings are consistent with these human and computer demonstrations of social facilitation. We observed both positive and negative responses to social facilitation in our study as different individuals responded differently to social facilitation. Positive responses were observed among individuals who accomplished their tasks while negative responses were expressed by those who found the tasks daunting and unachievable. Thus, it seems that social facilitation worked to both enhance and diminish performance as has been demonstrated in the cited studies. This explains both why some individuals in the intermittent group commented that they used the system as a "monitoring" agent and how system utilization and physical activity became interconnected among these individuals. For participants such as the man who called the system six times during the week that he had exercised more (see Section 4.2), being monitored by the system helped sustain his physical activity levels. Indeed, we followed up with this man 2 months after the study was completed and learned that he had stopped exercising once he stopped using the system.

Rickenberg and Reeves have also distinguished between the responses of the people who possess strong internal locus of control and those who have a strong external locus of control. Based on conclusions reached by Rickenberg and Reeves, "being monitored is less worrisome for people who believe that they control their own destiny than for those who think that their destiny is in the hands of others" [23]. We would have liked to explore this concept in our interviews. However, the theme did not emerge until late when most participants had completed the study. We can only speculate that the individuals who responded well to the "monitoring" functions of TLC-PA by increasing their physical activity levels had a strong "internal locus of control," and, conversely, those who stopped using the system because they felt uncomfortable to report unaccomplished goals, had a strong "external locus of control." This is a topic that we intend to explore in future evaluations.

6. Conclusions

The results of this study suggest two important design implications for developers of health information technology programs that use behavioral change strategies to interact with patients and consumers:

- Monitoring does have positive impact on some users' health behavior. However, not all users respond similarly to monitoring. Monitoring may also generate anxiety that is clearly an unpleasant experience and might impede utilization due to avoidance. Designers should be cognizant of this and thus should formulate precautionary measures. These precautionary measures may include the following: (1) providing an initial educational segment delivered by the system itself in which the reasons for the system's goal setting strategies are clearly stated and the likelihood that some goals may not be achieved are discussed. This can prepare the users for possible lapses and thus desensitize them to the resulting anxiety due to unachieved goals. (2) Ensuring that the system's tone and delivery are as supportive and empathetic as possible. The system should be able to provide insight [e.g., "This happens to many people in the process of behavior change and should not be considered as a failure."] and understanding [e.g., "It is not easy to change your lifestyle. It often takes many tries and extraordinary effort. Don't be disappointmented; just keep at it."].
- Users should be given significant *control* over their interaction with a health-promotion system. This may include control over the frequency and the duration of the interaction as well as over the content. For example, in the TLC-PA study, users were told to call the system two times a week with each call lasting between 15 and 20 min. As we noted, individuals in the intermittent group (group #2), appeared to take control of their utilization pattern and had the highest levels of satisfaction. Users also demonstrated a desire to exert more control over the system's content. They felt that they should have been able to select items from various lists of informational topics about exercise benefits and barriers. Even though these participants had some choice over which items they heard, they could not choose exactly what they wanted to listen to. For example, one user told us: "One thing about the call is that it asked you if you wanted to listen to ideas for indoor or outdoor [exercise]. I listened to indoor, but then I would have liked to listen to outdoor but there was no option to go

back. And, there was one other time when there was a choice between listening to this or that, and if you listened to one you couldn't listen to the other. It would be nice to have the option to do both." Similarly, there were occasions when users would have liked to have avoided the topics altogether. It irritated many users that they could not since they had to spend additional time interacting with the system. Thus, some individuals hung up in the middle of the conversation and others engaged in other activities as they used the system such as one woman who said sometimes she "would call" when she was "making dinner." Another user, who wanted to avoid hearing the "definition of exercise" that was repeated at the beginning of every contact, told us: "If there was a way to delete just certain parts at the beginning—It's, like, you know what's coming." These and other examples indicate that providing users with a choice about what they want to learn from a program [29] might positively affect their utilization, thus enhancing the system's impact on their health behavior.

Acknowledgment

This study was supported in part by the Robert Wood Johnson Foundation.

References

[1] Rice R. The Internet as a healthcare communications tool. Telemed Today 1995;3:6–7.
[2] Dambro M, Weiss BD, McCLure CL, Vuturo AF. An unsuccessful experience with computerized medical records in an academic medical center. J Med Educ 1988;63:617–23.
[3] Goddard B. Termination of a contract to implement an enterprise electronic medical record system. J Am Med Inform Assoc 2000;7:564–8.
[4] Lawler F, Cacy JR, Vicianni N, Hamm RM, Cobb SW. Implementation and termination of a computerized medical information system. J Fam Pract 1966;42:233–6.
[5] Bozeman B, Rogers JD. Use-and-transformation: a theory of knowledge value for research evaluation. Working Paper. Research Value Mapping Program. Georgia Institute of Technology; 1999.
[6] Delichatsios HK, Friedman RH, Glanz K, Tennstedt S, Smigelski C, Pinto BM, et al. Randomized trial of a "talking computer" to improve adults' eating habits. Am J Health Promot 2000;15:215–24.
[7] Pinto BM, Friedman R, Marcus BH, Kelley H, Tennstedt S, Gillman MW. Effects of a computer-based, telephone-counseling system on physical activity. Am J Prev Med 2002;23:113–20.
[8] Ramelson H, Friedman R, Ockene J. An automated telephone-based smoking cessation education and counseling system. Patient Educ Couns 1999;36:131–43.
[9] Friedman R, Kazis LE, Jette A, Smith BM, Stollerman J, Torgerson. A telecommunication.
[10] Farzanfar R, Locke SE, Vachon L, Charbonneau AM, Friedman RH. Computer telephony to improve adherence to antidepressants and clinical visits. Ann Behav Med Annual Meeting Supplement 2003;25:S161.
[11] Farzanfar R, Frishkopf S, Friedman R, Ludena K, Ward A. Humanizing the machine: refinement of an automated patient-management system. Ann Behav Med Annual Meeting Supplement 2004;27:S154.
[12] Young M, Sparrow D, Gottlieb D, Selim A, Friedman R. A telephone-linked computer system for COPD care. Chest 2001;119:1565–75.
[13] Friedman R. Automated telephone conversations to assess health behavior and deliver behavioral interventions. J Med Syst 1998;22:95–102.
[14] Mooney KH, Beck SL, Friedman R, Farzanfar R. Telephone-linked care for cancer symptom monitoring. Cancer Pract 2002;10:147–54.
[15] Mooney KH, Beck SL, Dudley WN, Farzanfar R, Friedman R. A computer-based telecommunication system to improve symptom care for women with breast cancer. Ann Behav Med Annual Meeting Supplement 2004;27:S152.
[16] Prochaska JW, DiClemente CC. Stages and processes of self-change in smoking: towards an integrative model of change. J Consult Clin Psychol 1983;51:390–5.
[17] Pinto BM, Friedman RH, Marcus BH, Kelley H, Tennstedt S, Gillman MW. Effects of a computer-based, telephone-counseling system on physical activity. Am J Prev Med 2002;23:113–20.
[18] Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27(Suppl. 3):S227–32.
[19] Glaser B, Strauss A. The discovery of grounded theory: strategies of qualitative research. London: Wiedenfiled & Nicholson; 1967.
[20] Guerin B. Social facilitation. Cambridge, UK: Cambridge University Press; 1993.
[21] Zajonc RB, Heingartner A, Herman EM. Social enhancement and impairment of performance in the cockroach. J Pers Soc Psychol 1969;13:83–92.
[22] Zajonc RB. Social facilitation. Science 1965;16:268–74.
[23] Cottrell NB, Wack DL, Skerak GJ, Rittle RH. Social facilitation of dominant responses by the presence of an audience and the mere presence of others. J Pers Soc Psychol 1968;9:245–50.
[24] Rickenberg R, Reeves B. The effects of animated characters on anxiety, task performance and evaluations of user interfaces. CHI Lett 2000;2:49–56.
[25] Jettmar E, Nass C. Adaptive testing: effects on user performance. CHI Lett 2002;4:129–34.
[26] Aiello JR, Svec CM. Computer monitoring of work performance: extending the social facilitation framework to electronic presence. J Appl Soc Psychol 1993;223:537–48.
[27] Stanton J, Barnes-Farrell J. Effects of electronic performance monitoring. J Appl Psychol 1996;6:738–45.

Editorial

Alcohol and Coronary Heart Disease
The Answer Awaits a Randomized Controlled Trial

Matthew S. Freiberg, MD, MSc; Jeffrey H. Samet, MD, MPH

In this issue of *Circulation*, Mukamal and colleagues report that frequent moderate consumption of alcohol is inversely associated with incident myocardial infarctions for both men and women.[1] This finding is consistent with numerous studies conducted in the course of 3 decades that have documented this inverse relationship. It has been described as a "J-" or "U-" shaped phenomenon, with moderate drinkers having a lower risk of myocardial infarctions as compared with abstainers and heavy drinkers.[2] A substantial contribution of this study is the results suggesting that the observed "benefits" of alcohol on myocardial infarction outcomes, particularly in men, are mediated in large part by only 3 factors: HDL-cholesterol, glucose intolerance (HbA_{1C}), and fibrinogen.

See p 1406

Previous studies have examined the potential causes for the observed benefits of moderate alcohol consumption and are consistent with these findings. Serum HDL-cholesterol concentrations, for example, increase in a dose-dependent response to alcohol consumption.[3] Prospective studies have reported a reduced incidence of type 2 diabetes mellitus[4] and 1 randomized controlled trial showed increased insulin sensitivity levels with moderate alcohol consumption.[5] Heavy drinking and binge drinking have been associated with an increased incidence of type 2 diabetes.[6] Similarly, moderate alcohol consumption is associated with lower levels of inflammatory and hemostatic markers as compared with never or occasional consumption.[7,8] Heavy alcohol drinkers, in contrast, have higher levels of the inflammatory marker C-reactive protein, as compared with moderate drinkers.[8] Even when risk factors are examined in aggregate, as in the case of the metabolic syndrome, current moderate consumers of alcohol have a lower prevalence of the syndrome than current nondrinkers.[9]

This abundance of evidence supporting the hypothesis that alcohol itself leads to lower incident coronary heart disease (CHD) events is not, however, definitive. Because the majority of the data describing the relationship between alcohol and CHD risk comes from observational studies, albeit well-established cohorts, the possibility that confounding may be partially or even entirely responsible for these observed effects remains controversial. Some contend that nondrinkers have a higher burden of CHD risk factors than do moderate drinkers and consequently have higher incidence rates of CHD events. In a recent study examining 30 cardiovascular-associated risk factors, 90% of the risk factors were significantly more prevalent in nondrinkers as compared with moderate drinkers.[10] It should be noted, however, that former drinkers and potentially some "sick quitters" were included in the abstainer group, thereby potentially increasing the risk burden in nondrinkers. Nevertheless, issues of confounding can be critical, as was recently appreciated in the highly publicized examples of hormone replacement, β-carotene, and vitamin E therapies. In each of these cases, the clinical adoption by some individuals of these potential therapies was proven premature when randomized controlled trials did not demonstrate the beneficial cardiovascular events that were found in observational studies.

Even beyond the important potential issue of uncontrolled or residual confounding in this case, the therapeutic cardioprotective use of alcohol raises serious concerns. Alcohol has considerable toxicities including abuse and dependence. In the United States, 17.6 million adults abuse alcohol or are dependent on alcohol.[11] Among current drinkers, 8% exceed weekly consumption criteria for risky drinking, defined by the National Institute on Alcohol Abuse and Alcoholism as >14 drinks for men or 7 drinks for women per week.[12] Binge drinking, which is associated with substantial morbidity, is common: 32% of current drinkers consumed ≥5 drinks on 1 occasion in the past year and 15% reported such behaviors on ≥12 days in the past year.[12] A recent study comparing the prevalence of alcohol abuse in the US population from 1990 to 1991 and 2000 to 2001 showed that alcohol abuse had significantly increased, from 3.03% to 4.65%.[11] Similarly, binge drinking increased from 1.2 billion episodes in 1993 to 1.5 billion episodes in 2001.[13] Although the prevalence of alcohol problems pales when compared with the prevalence of CHD, it is important to note that alcohol consumption was the third leading actual cause of death in the United States in 2000,[14] and in 1998 alone it was associated with economic costs totaling $184.6 billion.[15] Moreover, other side effects to be considered include the unfavorable associations between alcohol and violent crime, unintentional injuries including death and suicide, and myriad other health conditions includ-

From the Division of General Internal Medicine, Center for Research on Health Care, Departments of Medicine and Epidemiology, University of Pittsburgh Schools of Medicine and Public Health, Pittsburgh, Pa (M.S.F.), and the Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, Mass (J.H.S.).

Correspondence to Matthew S. Freiberg, MD, MSc, Division of General Internal Medicine, Center for Research on Health Care, University of Pittsburgh Medical Center, 230 McKee Pl, Suite 600, Pittsburgh, PA 15213. E-mail freibergms@upmc.edu

(*Circulation* 2005;112:1379-1381.)

Circulation is available at http://www.circulationaha.org
DOI: 10.1161/CIRCULATIONAHA.105.568030

ing certain forms of cancer.[16] Of particular importance with respect to the use of alcohol as a cardioprotective strategy, some investigators estimate that 5% to 7% of current abstainers and/or infrequent drinkers could develop diagnosable alcohol problems on beginning regular moderate alcohol-consumption regimens.[17] These estimates, however, are based on the general population and not on a particular subgroup of drinkers (eg, moderate drinkers consuming alcohol within "safe" limits).

Clinicians considering the use of alcohol as a treatment modality for CHD in clinical practice must also carefully consider the issues of dosing and duration of therapy and the needs of individual patients. Past research is not particularly informative about these issues. Observational data has consistently shown that 1 to 2 drinks per day is associated with lower cardiovascular risk, yet most studies did not ascertain and therefore did not consider risky patterns of alcohol consumption (eg, binge drinking) in the study design or subsequent analyses. For example, among people with infrequent episodes of heavy drinking, the risk of CHD is higher as compared with abstainers even when the overall volume of alcohol consumed is low.[18] None of the major prospective studies, including the present study, used measurement tools such as the Timeline Followback, the gold standard for ascertaining alcohol consumption in the past 30 days.[19] Other measurement tools only allow investigators to average daily patterns of consumption based on weekly, monthly, or annual reports. The Timeline Followback, however, specifically asks subjects to recall the amount of alcohol consumed during each of the past 30 days, thereby quantifying daily volumes and patterns of consumption.[19] Another issue of concern is the lack of adjustment for duration of alcohol consumption or the consideration of past quantities of consumption in the predictive models. Also, the alcohol data collected prospectively during the follow-up period in these studies are sparse. Assumptions of stable, unchanging alcohol consumption over time may simply not be accurate. Previous studies clearly show, for example, that younger adults tend to consume higher quantities of alcohol and demonstrate increased risky behavior associated with alcohol (eg, binge drinking) than do older adults.[13] Because the alcohol exposure for participants most likely predates their enrollment into established cohorts, not accounting for this past quantity or pattern of alcohol exposure may be problematic. Another concern, as in the case of the present study, is that participants in these established cohorts may not be necessarily generalizable to the overall population. They typically do not represent the minority of the population with risky alcohol consumption behavior patterns, nor are they designed to provide information pertaining to an individual's metabolic, dietary, or genetic make-up, which may affect alcohol metabolism and subsequent CHD risk. Consequently, the available observational data, although important, still do not answer important questions for physicians and patients such as: Is/are 1 to 2 drinks several days a week the appropriate amount of alcohol and pattern of consumption that is required to be cardioprotective? How long should this therapy continue? Does this therapy need to be lifelong? Is there an amount or lifetime quantity of alcohol consumption that needs to be consumed to achieve the "observed benefit?"

The important limitations of these observational studies do not diminish the significance of these studies for identifying important associations that make the compelling case for pursing clinical trials that can direct clinical care. The abundance of favorable data associated with alcohol consumption in relation to CHD juxtaposed with a side effect profile that includes risk of dependence, comorbid medical conditions, or even death compounded by the absence of specific dose or duration for the therapy to achieve a well-defined cardiovascular outcome presents the practicing physician with somewhat of a dilemma. Current recommendations are clear that if high-risk drinking behavior is present, patients should be asked to cut back or seek treatment.[20] If a patient is currently drinking within "safe" limits, then physicians can delineate and discuss the potential benefits of such behavior and the risks of drinking beyond such limits with patients.[21] For those who do not drink or only occasionally consume alcohol, recommendations for alcohol use as a protective agent should not be broadly endorsed. It is true that some proponents such as Dr R. Curtis Ellison of the Boston University School of Medicine argue that, "For appropriate patients without any contraindications (eg, history of abuse, medical conditions, religious or ethical inhibitions, etc) who do not drink or do so only occasionally, and who wish to do so, encouraging a glass of wine or other alcoholic beverage with dinner every evening may be the best advice you can give them."[22] Others, including the present editorialists, believe that observational studies are not enough to make a prescription.

Although no large prospective studies have examined the role of alcohol as a cardioprotective agent for CHD, some randomized controlled trials involving alcohol do show a beneficial effect on insulin and triglyceride concentrations among those who consume 30 g (2 drinks)/day,[5] and numerous trials have shown beneficial effects of moderate alcohol intake on lipids, clotting factors, and other intermediaries.[7] Primary prevention trials have been considered previously but judged doubtful for several important reasons: high costs, difficulty with blinding subjects, the need to find large numbers of people who have no contraindications to alcohol use (eg, medical reasons) and who are willing to forgo or continue alcohol use for an extended period of time, and the possibility of eventual alcohol abuse or dependence.[23] Others, however, have suggested that large secondary prevention trials among those at highest risk (eg, history of CHD) and on appropriate therapy may not only be possible but are also arguably justifiable.[24]

Consequently, given the importance and prevalence of CHD, the potential benefits of alcohol use, the risks of alcohol abuse, and the inherent limits of observational studies, it is time for serious consideration of the usual next step in the assessment of a prevention intervention: performance of a randomized controlled trial.

References

1. Mukamal KJ, Jensen MK, Grønbaek M, Stampfer MJ, Manson JE, Pischon T, Rimm EB. Drinking frequency, mediating biomarkers, and

risk of myocardial infarction in women and men. *Circulation.* 2005;112: 1406–1413.
2. Corrao G, Rubbiati L, Bagnardi V, Zambon A, Poikolainen K. Alcohol and coronary heart disease: a meta-analysis. *Addiction.* 2000;95: 1505–1523.
3. Gaziano JM, Buring JE, Breslow JL, Goldhaber SZ, Rosner B, VanDenburgh M, Willett W, Hennekens CH. Moderate alcohol intake, increased levels of high-density lipoprotein and its subfractions, and decreased risk of myocardial infarction. *N Engl J Med.* 1993;329: 1829–1834.
4. Rimm EB, Chan J, Stampfer MJ, Colditz GA, Willett WC. Prospective study of cigarette smoking, alcohol use, and the risk of diabetes in men. *BMJ.* 1995;310:555–559.
5. Davies MJ, Baer DJ, Judd JT, Brown ED, Campbell WS, Taylor PR. Effects of moderate alcohol intake on fasting insulin and glucose concentrations and insulin sensitivity in postmenopausal women: a randomized controlled trial. *JAMA.* 2002;287:2559–2562.
6. Carlsson S, Hammar N, Grill V, Kaprio J. Alcohol consumption and the incidence of type 2 diabetes: a 20-year follow-up of the Finnish Twin Cohort Study. *Diabetes Care.* 2003;26:2785–2790.
7. Rimm EB, Williams P, Fosher K, Criqui M, Stampfer MJ. Moderate alcohol intake and lower risk of coronary heart disease: meta-analysis of effects on lipids and haemostatic factors. *BMJ.* 1999;319:1523–1528.
8. Imhof A, Froehlich M, Brenner H, Boeing H, Pepys MB, Koenig W. Effect of alcohol consumption on systemic markers of inflammation. *Lancet.* 2001;357:763–767.
9. Freiberg MS, Cabral HJ, Heeren TC, Vasan RS, Curtis Ellison R. Alcohol consumption and the prevalence of the metabolic syndrome in the US: a cross-sectional analysis of data from the Third National Health and Nutrition Examination Survey. *Diabetes Care.* 2004;27:2954–2959.
10. Naimi TS, Brown DW, Brewer RD, Giles WH, Mensah G, Serdula MK, Mokdad AH, Hungerford DW, Lando J, Naimi S, Stroup DF. Cardiovascular risk factors and confounders among nondrinking and moderate-drinking U.S. adults. *Am J Prev Med.* 2005;28:369–373.
11. Grant BF, Dawson DA, Stinson FS, Chou SP, Dufour MC, Pickering RP. The 12-month prevalence and trends in *DSM-IV* alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. *Drug Alcohol Depend.* 2004;74:223–234.
12. *Alcohol Consumption by Persons 18 Years of Age and Over, According to Selected Characteristics: United States, Selected Years 1997–2001.* Hyattsville, Md: Centers for Disease Control and Prevention, National Center for Health Statistics; 2003: 223–226.
13. Naimi TS, Brewer RD, Mokdad A, Denny C, Serdula MK, Marks JS. Binge drinking among US adults. *JAMA.* 2003;289:70–5.
14. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA.* 2004;291:1238–1245.
15. Harwood HJ. *Updating Estimates of the Economic Costs of Alcohol Abuse in the United States: Estimates, Update Methods, and Data.* Bethesda, Md: National Institute on Alcohol Abuse and Alcoholism; 2000.
16. Rehm J, Room R, Graham K, Monteiro M, Gmel G, Sempos CT. The relationship of average volume of alcohol consumption and patterns of drinking to burden of disease: an overview. *Addiction.* 2003;98: 1209–1228.
17. Gunzerath L, Faden V, Zakhari S, Warren K. National Institute on Alcohol Abuse and Alcoholism report on moderate drinking. *Alcohol Clin Exp Res.* 2004;28:829–847.
18. McElduff P, Dobson AJ. How much alcohol and how often? Population based case-control study of alcohol consumption and risk of a major coronary event. *BMJ.* 1997;314:1159–1164.
19. Sobell LC, Sobell MB. *Alcohol Timeline Followback (TLFB) Users' Manual.* Toronto: Addiction Research Foundation; 1995.
20. U.S. Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. *Ann Intern Med.* 2004;140:554–556.
21. Goldberg IJ, Mosca L, Piano MR, Fisher EA. AHA Science Advisory. Wine and your heart: a science advisory for healthcare professionals from the Nutrition Committee, Council on Epidemiology and Prevention, and Council on Cardiovascular Nursing of the American Heart Association. *Circulation.* 2001;103:472–475.
22. Wilson JF. Should doctors prescribe alcohol to adults? *Ann Intern Med.* 2003;139:711–714.
23. Mukamal KJ, Rimm EB. Alcohol's effects on the risk for coronary heart disease. *Alcohol Res Health.* 2001;25:255–261.
24. Goldberg IJ. To drink or not to drink? *N Engl J Med.* 2003;348:163–164.

KEY WORDS: Editorials ■ alcohol ■ myocardial infarction ■ coronary disease ■ therapy

Warfarin Maintenance Dosing Patterns in Clinical Practice*

Implications for Safer Anticoagulation in the Elderly Population

David Garcia, MD; Susan Regan, PhD; Mark Crowther, MD, MSc; Robert A. Hughes, MD; and Elaine M. Hylek, MD, MPH†

Background: **The use of anticoagulant therapy is expanding among the elderly population, in part because of the increasing prevalence of atrial fibrillation. Published data describing the warfarin maintenance dose requirements for this age group are limited. Because warfarin therapy is often initiated in the outpatient setting where significant barriers to daily monitoring exist for this patient population, a better understanding of the factors that predict lower dose requirements may reduce the risk of unanticipated overanticoagulation and hemorrhage.**
Objective: **To define the effects of age and gender on the warfarin maintenance dose among ambulatory adult patients with an international normalized ratio target between 2.0 and 3.0.**
Design: **Prospective cohort study and retrospective cohort secondary data source.**
Setting: **One hundred one community-based physician practices with dedicated warfarin management systems and an academic medical center anticoagulation clinic.**
Patients: **A total of 4,616 patients comprised the prospective cohort, and 7,586 patients comprised the retrospective cohort. Of the 12,202 patients, 2,359 were ≥ 80 years of age.**
Measurements: **Median weekly and daily maintenance warfarin dose.**
Results: **The warfarin dose was inversely related to age and was strongly associated with gender. The median weekly dose ranged from 45 mg (6.4 mg/d) for men who were < 50 years of age to 22 mg (3.1 mg/d) for women ≥ 80 years of age. The weekly dose declined by 0.4 mg/yr (95% confidence interval [CI], 0.37 to 0.44; $p < 0.001$) and women required 4.5 mg less per week than men (95% CI, 3.8 to 5.3; $p < 0.001$). Among patients who were > 70 years of age, the often-suggested initiation dose of 5 mg/d will be excessive for 82% of women and 65% of men.**
Conclusions: **Warfarin dose requirements decrease greatly with age. Older women require the lowest warfarin doses. These observations suggest that, when warfarin is being initiated, the commonly employed empiric starting dose of 5 mg/d will lead to overanticoagulation for the majority of patients in the geriatric age group; lower initiation and maintenance doses should be considered for the elderly.** ***(CHEST 2005; 127:2049–2056)***

Key words: anticoagulation; warfarin

Abbreviations: CI = confidence interval; INR = international normalized ratio

The number of elderly patients who are eligible to receive warfarin is steadily expanding, in part because of the increasing prevalence of atrial fibrillation.[1] Atrial fibrillation affects approximately 10% of individuals who are ≥ 80 years of age, and it is projected that 4 million individuals in the United States will have atrial fibrillation by the year 2030.[2] Both the risk of stroke in patients with atrial fibrillation and the risk of mortality in patients who

*From the Department of Medicine (Dr. Garcia), Division of General Internal Medicine, University of New Mexico School of Medicine, Albuquerque, NM; the Department of Medicine (Drs. Regan and Hughes), General Medicine Division, Massachusetts General Hospital, Harvard Medical School, Boston, MA; the Department of Medicine (Dr. Crowther), McMaster University, Hamilton, ON, Canada; and the Department of Medicine (Dr. Hylek), General Internal Medicine Research Unit, Boston University School of Medicine, Boston, MA.
This research was supported by a grant from Bristol-Myers Squibb, Princeton, NJ. Dr. Crowther holds a Canadian Institutes for Health Research Scholarship.
Manuscript received October 19, 2004; revision accepted December 3, 2004.
†A list of practices and directors that participated in the study is located in the Appendix.

Correspondence to: David A. Garcia, MD, Department of Internal Medicine, MSC10 5550, Albuquerque, NM 87131; e-mail: davgarcia@salud.unm.edu

experience a stroke increase with age.[3] Warfarin therapy reduces the risk of stroke in patients with atrial fibrillation by 68%.[4] However, despite evidence that the benefit of anticoagulation therapy is greatest in patients who are > 75 years of age, advanced age independently predicts warfarin nonuse.[5–8] Warfarin is likely to remain the only widely available oral anticoagulant agent for the foreseeable future.

The fear of hemorrhage and the anticipated difficulty with the control of anticoagulation therapy have been cited as reasons for not prescribing warfarin to elderly patients.[9] Compared to younger patients, older individuals exhibit an enhanced dose response to warfarin, are more likely to become overanticoagulated, and take longer to return to safe international normalized ratio (INR) levels when they have a supratherapeutic INR.[10–16] Furthermore, increasing age and elevated INR have been shown to be risk factors for bleeding among patients receiving anticoagulant therapy.[17–20]

Currently, few data are available to guide clinicians in the dosing of warfarin therapy for the very elderly. The largest published data set included patients who were treated to achieve different target INR intensities and included only 143 individuals who were ≥ 80 years of age.[16] An article from The Seventh American College of Chest Physicians Consensus Conference on Antithrombotic and Thrombolytic Therapy[21] stated that "a starting dose of < 5 mg might be appropriate in the elderly," but did not make specific recommendations for the geriatric population. Although prediction models[22] of the maintenance warfarin dose have been published, it is not clear that octogenarians were well-represented in the derivation cohorts, and these models are not easily applied in all clinical settings. Similarly, warfarin initiation nomograms are largely based on younger patients.[23] Two other published nomograms[24,25] incorporated adjustments for age, but were derived from study populations that included only 11 patients who were ≥ 80 years of age and 30 patients who were > 75 years of age. Warfarin nomograms are limited not only by the paucity of older patients included in the cohorts from which they were derived, but also by their reliance on daily or alternate-day monitoring of the INR. Such frequent laboratory testing represents a significant burden and may be unrealistic for many elderly patients who are initiating warfarin therapy. Compared to younger patients, older patients are more often dependent on others for transportation and are challenged by physical limitations to mobility. These impediments to frequent testing are especially relevant for those patients who live in less densely populated areas or in less developed countries. An awareness of these common, practical access barriers has prompted the authors of at least one guideline[26] to recommend INR testing "at least weekly during the initiation of oral anticoagulant therapy."

Because warfarin dose requirements are known to decrease with age, the often-suggested empiric starting dose of 5 mg/d will likely result in supratherapeutic INR values for many elderly patients. Systematic overdosing of patients who are ≥ 80 years of age may partly explain the observation that the risk for major hemorrhage is highest in older patients and in the early phase of therapy.[27,28]

Given the practical barriers that may limit the use of initiation nomograms, and the uncertainty about whether prediction models and nomograms can be generalized to patients who are > 80 years of age, we sought to describe warfarin maintenance dosing patterns that were experienced in the clinical care of two large, independent, ambulatory patient populations. We have provided the distribution of the warfarin maintenance dose by age, gender, and indication. Additional knowledge of warfarin dose requirements for very elderly individuals will help physicians with empiric dosing decisions. A more informed decision about dosing will likely decrease the risk of excessive anticoagulation, particularly among those patients for whom daily testing is not possible.

Materials and Methods

Study Populations

A prospective cohort was assembled as part of a larger study assessing anticoagulation care in the United States. Sites were invited to participate by letter and through a study web site. All enrolled sites had at least one dedicated provider managing warfarin therapy, usually within the setting of a community-based physician group office practice. Patients were invited to participate at the time of their routine appointment. Study enrollment began in August 2000 and ended in February 2002. All study participants provided written informed consent. During the study period, 6,761 patients were enrolled into the study from 101 sites in 38 states. We included those patients who were ≥ 18 years of age, had a physician-specified INR target range of 2.0 to 3.0, and had achieved a stable warfarin dose during the observation period. Health-care personnel (typically, a nurse or pharmacist dedicated to anticoagulation therapy management) from the participating clinic made these decisions.

Encrypted patient-derived data from each site were transmitted weekly to an independent data-coordinating center (McKessonHBOC BioServices; Rockville, MD). Missing data fields and data-entry errors were flagged and resolved prior to data transfer. Study investigators were blinded to the identification and location of the participating practices and patients. Nearly all sites had experience using the designated data-entry system (CoumaCare; Bristol-Myers Squibb; Princeton, NJ) before the study began. The designated software program has been used by many anticoagulation therapy management services for clinical pur-

poses (*eg*, patient tracking, data entry, and record keeping). The software does not make dosing or follow-up recommendations.

Variables of interest included age, gender, ethnicity, indication for oral anticoagulant therapy, physician-specified INR target range, INR values and dates of testing, warfarin dose at each visit, and the presence of other medical diagnoses, specifically, COPD, coronary artery disease, diabetes mellitus, heart failure, hypertension, hyperlipidemia, active malignancy, and prior stroke. The effect of amiodarone was specifically sought out because of the magnitude of its potentiating effect on warfarin.[29–31] Patient weight was routinely recorded in seven of the participating sites. The relationship between weight and gender was explored within this subset of patients.

Retrospective Cohort Secondary Data Source

We additionally assessed the effects of age and gender on warfarin dose in an independent cohort of ambulatory patients who were managed at the Anticoagulation Therapy Unit of Massachusetts General Hospital. We included all patients who were ≥ 18 years of age, had a specified INR target range of 2.0 to 3.0, were followed up during the period from 1993 to 2003, and had achieved a stable warfarin dose. The available data for this cohort included age, gender, indication for oral anticoagulant therapy, INR values, dates of testing, and warfarin dose prescribed. The retrospective cohort decisions about warfarin dosing and follow-up were made by nurses employed in the Anticoagulation Therapy Unit at the Massachusetts General Hospital.

Statistical Analysis

We defined a stable warfarin dose as a dose that was prescribed twice consecutively after two consecutive in-range INR measurements (*ie*, 2.0 to 3.0). For each patient, we used the first stable dose observed. The median warfarin dose was calculated across gender and decade of age. Univariate comparisons were performed using *t* tests and rank sum tests. Nonparametric tests for trend were used to assess the change in median dose with age.[32]

We assessed the independent effects of age and gender for the prospective cohort using multiple linear regression. Variables with the potential to influence the maintenance warfarin dose were included in the model if the p value was < 0.1 in the univariate analysis. A two-sided p value of < 0.05 was considered to be statistically significant. The final model included terms for age, gender, coronary artery disease, diabetes mellitus, heart failure, hypertension, amiodarone use, and venous thromboembolic disease. Analyses were performed using a statistical software package STATA software; (STATA Corp; College Station, TX). For the subset of patients with weight available, we present the median dose and age for men and women stratified by weight to better illustrate the relationships among these variables.

Human Studies Approval

The Western Institutional Review Board (Olympia, WA) provided approval of the registry protocol and the informed consent form. Each participating site director was responsible for obtaining internal human studies approval, if locally required, and to provide evidence that these requirements had been met before the site was enrolled. The study protocol was also approved by the institutional review board at Massachusetts General Hospital.

Role of the Funding Source

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

Prospective Cohort

A total of 4,616 patients met the study inclusion criteria. The mean age was 72 years, and 58% of patients were men (Table 1). There were 1,127 patients who were ≥ 80 years of age. Sixty-two percent of the cohort was receiving warfarin therapy for stroke prevention in atrial fibrillation, and 15% for venous thromboembolic disease. The maintenance warfarin dose was significantly greater for men compared to women (30 vs 25 mg/wk, respectively; $p < 0.0001$). Trends of decreasing dose with increasing age were significant overall and within each gender (all $p < 0.01$) [Fig 1]. The decrements in the median weekly maintenance warfarin dose between the youngest age group (*ie*, < 50 years) and the oldest age group (*ie*, ≥ 90 years) were 47% for women and 39% for men.

Similar relationships between both warfarin dose and age, and warfarin dose and gender were seen in patients with atrial fibrillation and in those with venous thromboembolic disease (Table 2). Within each indication, younger men required the highest doses (median daily dose: for atrial fibrillation, 5.4 mg; for venous thromboembolic disease, 6.4 mg), and women who were ≥ 80 years of age required the lowest doses (median daily dose: for atrial fibrillation, 3.1 mg; for venous thromboembolic disease, 3.6 mg). Based on our study population, the often-recommended initiation dose of 5 mg/d would exceed the maintenance warfarin dose requirement for approximately 65% of women 60 to 69 years of age,

Table 1—*Clinical Features of Patients with Target INR of 2.0 to 3.0 in the Prospective Cohort (n = 4,616)

Variables	Values
Age, yr	72 (22–100)
Male gender	58 (2,655)
Non-white race	5 (234)
Indication	
Atrial fibrillation	62 (2,849)
Venous thromboembolism	15 (688)
Stroke	11 (499)
Cardiomyopathy	3 (160)
Coronary artery disease	3 (147)
Prosthetic heart valve	3 (151)
Other	3 (117)
Medical illness	
Hypertension	45 (2,065)
Coronary artery disease	30 (1,407)
Heart failure	22 (996)
Diabetes mellitus	17 (769)
Malignancy	7 (334)
COPD	6 (298)
Amiodarone use	3 (144)

*Values given as mean (range) or No. (%)

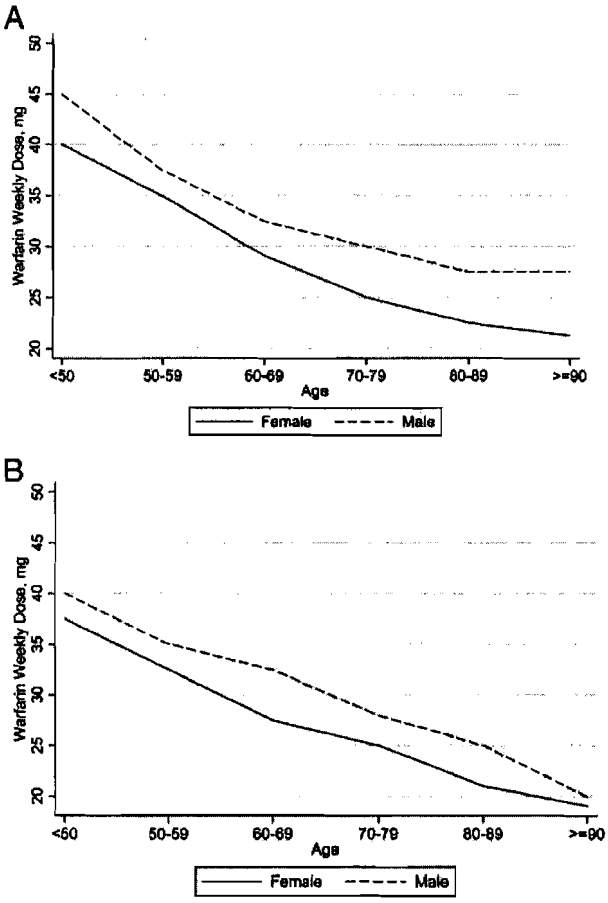


FIGURE 1. Median weekly maintenance warfarin dose by age and gender for patients with target INR between 2.0 and 3.0 in the prospective cohort (*top, A*) and the retrospective cohort (*bottom, B*). Patients receiving amiodarone have been excluded from the prospective sample.

79% of women 70 to 79 years of age, and 84% of women ≥ 80 years of age (Table 3).

The effects of age and gender were independent and did not change in magnitude after controlling for comorbid illness and amiodarone use. For each year of age, the weekly warfarin dose declined by 0.4 mg (95% confidence interval [CI], 0.37 to 0.44 mg; $p < 0.001$). At any given age, the weekly warfarin dose for women was 4.5 mg lower (95% CI, 3.8 to 5.3 mg; $p < 0.001$) than that of men. Amiodarone use was associated with a reduction of 7.3 mg/wk (95% CI, 5.2 to 9.5 mg/wk; $p < 0.001$) in warfarin maintenance dose. Other less potent factors that were independently associated with warfarin dose included heart failure, coronary artery disease, and diabetes mellitus (Table 4).

Retrospective Cohort

The retrospective cohort consisted of 7,586 patients, 1,232 of whom were ≥ 80 years of age. The median daily doses by age and gender for patients with atrial fibrillation or venous thromboembolic disease are presented alongside those of the prospective cohort in Table 2. Similar patterns of change in the median dose by age for women and men were evident in the two cohorts (Fig 1). The proportion of patients requiring a dose of < 35 mg/wk approximates that of the prospective cohort (Table 3).

Relationship of Gender and Weight to Maintenance Warfarin Dose

We examined the median warfarin dose by gender-stratified deciles of weight in the subset of 369 patients from the seven sites where weight was routinely recorded. The sample size was too small for significant dose comparisons within each stratum; however, in seven of the nine weight classes women required lower doses. There was no significant difference in mean age between genders for these strata (Table 5). The exceptions to this trend were the weight class of 150 to 159 lbs, in which the higher dose among women may be explained by the fact that the women were significantly younger than the men ($p = 0.017$), and the heaviest weight class (*ie*, those weighing > 234 lbs).

DISCUSSION

A limited amount of data on the dosing of warfarin in an elderly population has been published. The largest such study included only 143 individuals who were ≥ 80 years of age and included patients who had been treated to achieve various target intensities. Existing data that form the basis of the current guidelines[21,33] were largely obtained from younger cohorts. In this study, we report on the maintenance warfarin dose of 12,202 ambulatory outpatients with an INR target range of 2.0 to 3.0. We provide the warfarin dosing experience of 2,359 individuals who were ≥ 80 years of age. We were able to assess the independent effects of age and gender through multiple linear regression models that controlled for medical conditions that are prevalent in this population and the known powerful effect of amiodarone.

We found highly significant associations between warfarin dose and age, and warfarin dose and gender in two large cohorts. For each additional year of age, the weekly warfarin dose declined by 0.4 mg. At any given age, the mean weekly warfarin dose for women was 4.5 mg lower than that for men. Among patients with atrial fibrillation as the indication for warfarin therapy, the median daily dose ranged from 5.4 mg, for men aged 50 to 59 years of age, to 3.1 mg, for women ≥ 80 years of age. These findings have important implications for patient safety and quality

Table 2—*Median Daily Warfarin Dose by Age and Gender for Patients With Atrial Fibrillation or Venous Thromboembolic Disease*

	Prospective Cohort (n = 2,849)						Retrospective Cohort (n = 3,014)					
	Male			Female			Male			Female		
Age, yr	Median	Warfarin Dose,* mg	No.	Median	Warfarin Dose,* mg	No.	Median	Warfarin Dose,* mg	No.	Median	Warfarin Dose,* mg	No.
Atrial fibrillation												
50–59	5.4	(4.0, 6.4)	117	5.0	(3.9, 6.0)	42	5.0	(3.9, 6.8)	233	4.3	(2.6, 5.9)	72
60–69	4.6	(3.6, 5.7)	349	4.0	(2.9, 5.4)	199	4.6	(3.6, 5.7)	466	3.9	(2.9, 5.4)	263
70–79	4.3	(3.2, 5.4)	736	3.5	(2.5, 4.6)	526	4.0	(2.9, 5.0)	679	3.4	(2.5, 4.6)	519
80–89	3.9	(2.5, 5.0)	393	3.2	(2.5, 4.3)	364	3.6	(2.5, 4.6)	270	2.9	(2.1, 3.9)	353
≥ 90	3.6	(2.6, 4.0)	27	3.0	(2.0, 3.6)	41	3.4	(2.1, 4.6)	12	3.1	(2.3, 3.9)	35
Venous thromboembolic disease												
< 50	6.4	(5.0, 8.2)	55	6.1	(5.0, 7.7)	48	5.7	(4.3, 8.0)	186	5.0	(3.6, 7.1)	192
50–59	5.7	(4.3, 7.1)	65	5.2	(3.6, 6.8)	50	5.7	(4.3, 7.5)	107	5.0	(3.6, 6.4)	117
60–69	5.0	(3.6, 6.1)	45	4.2	(2.9, 5.7)	84	5.0	(3.2, 6.1)	129	4.5	(3.0, 6.1)	152
70–79	4.5	(3.4, 6.0)	114	3.6	(2.7, 5.0)	135	5.0	(3.2, 6.1)	135	3.6	(2.5, 5.0)	150
80–89	3.9	(3.2, 5.0)	47	3.6	(2.6, 4.3)	39	4.0	(2.6, 5.4)	50	3.2	(2.4, 4.0)	77
≥ 90			(< 5)			(< 5)	4.0	(3.0, 5.0)	6	2.0	(1.0, 3.0)	18

*Values given as 25th percentile, 75th percentile of daily warfarin dose.

of anticoagulation control in the outpatient setting where daily or alternate-day monitoring of the INR is not always possible. Based on the actual warfarin maintenance doses observed in our study populations, the often-recommended 5 mg daily warfarin starting dose would be too high for up to 82% of women (and 65% of men) who were > 70 years of age. Thus, our data suggest that the higher rates of bleeding and erratic INR values that have been reported[17–20] in the early phase of warfarin therapy may in part be explained by overly aggressive dosing in the elderly.

The mechanism for increased sensitivity to warfarin with aging is not well-understood. Factors such as hypoalbuminemia (leading to a reduced volume of distribution), decreased dietary vitamin K intake (resulting in a decreased capacity to synthesize functional clotting factors), reduced absorption of vitamin K, and polypharmacy (producing drug-drug interactions that potentiate warfarin) may lead to increased warfarin sensitivity for an individual patient. The pharmacokinetics of warfarin do not explain the lower dose requirements, because the distribution and elimination half-lives of warfarin change very little with advancing age.[34,35]

Pharmacodynamic changes with age have been studied much less. A negative correlation between age and hepatic drug-metabolizing enzymes has been observed in rat models *in vitro*,[36,37] and a study[38] of hepatic microsomal enzymes in humans showed a 30% decline in hepatic drug metabolism and cytochrome P450 content with age. Others have suggested an age-related decrease in the diffusion of oxygen within the hepatocyte that compromises the oxygen-dependent mixed-function oxidase enzymes.[39] Changes in hepatic glutathione levels and decreased activity of the nicotinamide adenine dinucleotide phosphate-dependent reductase of vitamin K are other touted hypotheses.[40]

The observation that women, independent of age,

Table 3—*Percentage of Patients With Target INR of 2.0 to 3.0 Whose Maintenance Warfarin Dose Was < 35 mg/wk (5 mg/d)**

Age, yr	Prospective Cohort		Retrospective Cohort	
	Male	Female	Male	Female
50–59	36 (270)	45 (145)	40 (702)	52 (373)
60–69	53 (556)	65 (376)	53 (1,103)	64 (707)
70–79	61 (1,107)	79 (798)	66 (1,341)	76 (1,112)
80–89	71 (562)	84 (481)	76 (495)	88 (650)
≥ 90	82 (33)	90 (51)	86 (21)	91 (66)

*Values given as the % (total No. of patients).

Table 4—*Independent Effect of Age and Gender on Warfarin Dose in the Prospective Cohort**

Variables	Adjusted Difference in Weekly Warfarin Dose,† mg	p Value
Age, per year	− 0.4 (− 0.44 to − 0.37)	< 0.001
Female gender	− 4.5 (− 5.30 to − 3.75)	< 0.001
Amiodarone	− 7.3 (− 9.48 to − 5.17)	< 0.001
Coronary artery disease	− 1.6 (− 2.43 to −0.74)	< 0.001
Heart failure	− 1.0 (− 1.97 to − 0.10)	0.03
Diabetes mellitus	1.4 (0.43 to 2.47)	0.005

*The full multiple linear regression model included age, gender, amiodarone therapy, coronary artery disease, diabetes mellitus, heart failure, hypertension, and venous thromboembolic disease.
†Values in parentheses are 95% CI.

Table 5—*Median Daily Maintenance Warfarin Dose and Median Age for Men and Women by Deciles of Weight (n = 369)*

Weight, lbs	Daily Warfarin Dose				Age, yr	
	Male, mg	No.	Female, mg	No.	Male	Female
< 135		< 5	2.9	37	85	83
135–149	4.0	10	3.7	33	74	78
150–159	3.6	15	4.3	17	80	76
160–169	4.3	17	3.7	12	79	72
170–179	4.4	28	2.9	16	75	75
180–189	4.3	26	4.0	11	76	74
190–199	4.3	24	3.4	6	73	66
200–214	4.7	39	3.3	5	74	78
215–234	5.4	22	4.7	9	71	70
> 234	6.0	28	6.4	11	67	62

require less warfarin than men may be explained partly by differences in mean body size or hepatic fat content, or by intrinsic differences in warfarin metabolism. Reports[41] of sex-related differences in hepatic clearance by cytochrome P450 enzymes warrant further investigation. One set of experiments[42] using an animal model has suggested that the expression of at least some cytochrome P450 enzymes may be regulated by sex steroids.

Our study was based on two large patient populations. Therefore, we were unable to account for individual genetic variation (*eg*, polymorphism in P450 CYP2C9). However, differences in the CYP2C9 genotype would not be expected to confound our central findings of increased warfarin sensitivity with advancing age and gender. Furthermore, pharmacogenetic testing will not be widely available to assist physicians making time-sensitive treatment decisions in the office setting, and there is currently no evidence describing the specific influence of CYP2C9 genotype on warfarin dosing in individual patients.

The possible effect of weight on warfarin dose could not be completely accounted for in our data set. It is likely that female gender largely subsumes the effect of weight on dose requirement. In our analysis of a small subset of patients, the effect of weight was most evident in the uppermost deciles (*ie*, > 234 lbs). Overall, our data suggest that gender-related differences in dose are independent of weight.

The data used in our analysis were derived entirely from outpatients who were receiving long-term warfarin therapy. Although our study did not directly address the initiation doses of warfarin for the immediate treatment of an acute arterial or venous thrombosis, our results support a cautious approach to empiric dose selection, particularly in older women. Care must be taken in extrapolating the results from studies of younger cohorts with male predominance in whom initiation doses of 10 mg/d were shown to be safe.[43–47] The median doses listed in Table 2 may be a reasonable empiric choice for some elderly outpatients in whom warfarin therapy is being initiated. It must be remembered, however, that optimal initiation and maintenance doses of warfarin will likely be lower among hospitalized patients, because they are more likely to experience concomitant decreased oral intake, recent surgery, decompensated heart failure, liver impairment, polypharmacy, or other factors associated with reduced warfarin dose requirement.[10,48] Individuals who are sensitive to warfarin would be expected to manifest an early, exuberant anticoagulant response, usually after the second dose. Finally, for patients whose weight deviates significantly from the norm for their age and gender, body mass should also be considered in any empiric dosing decisions.

Conclusions

This study found that mean warfarin doses declined predictably with age and were lower in women than in men. The average maintenance dose for young men was 6.4 mg/d; for women > 80 years of age, the average maintenance dose was only 3.1 mg/d. These observations suggest that the current initiation and maintenance doses should be modified to reduce the risk of inadequate therapy in young patients, and excessive anticoagulation in elderly patients.

Appendix

The following practices and directors participated in the study, with the sites listed in decreasing order of the number of patients enrolled: Lutheran General Hospital, Niles, IL: W. Fried, M. Pubentz; Physicians, Inc., Lima, OH: D. Parker; Idaho Cardiology Associates, Boise, ID: F. Badke; North Clinic, Robbinsdale, MN: V. Krug; Rockwood Clinic P.S.–Main, Spokane, WA: J.S. Pennock; Wenatchee Valley Clinic, Wenatchee, WA: R. Kirby Primm, L. Vaughn; Framingham Heart Center, Framingham, MA: J. Dangel, S.R. Hewett; Clinic Pharmacy Consultants-Brainerd Medical Center, Brainerd, MN: B. Twamley, R. Sorenson; Woodland Healthcare, Woodland, CA: L. Smith, T. Fajerson; Cardiology, PC, Syracuse, NY: S. O'Donnell; Health Care American Corp, Bradenton, FL: C. Hoffman; DuPage Medical Group-Department of Cardiology, Winfield, IL: N. Kinsley; Camino Medical Group, Sunnyvale, CA: S. Edwards; Ohio Valley Heartcare, Evansville, IN: L. Janeira, J. Robb; Desert Medical Group/Oasis IPA, Palm Springs, CA: H.F. Bellaci, J. Bellaci; Anchor Health Center, Naples, FL: M. Means; Sutter Gould Medical Foundation, Modesto, CA: J.E. Baker; Hannibal Clinic Inc., Hannibal, MO: L. Chalton; Saratoga Cardiology, Saratoga Springs, NY: R. Sheldon, D. Kandath; Lima Memorial Hospital, Lima, OH: C.L. Thompson, J. Recker; Staten Island University Hospital, Staten Island, NY: M. Howard; Jacksonville Cardiovas-

cular Clinic, Jacksonville, FL: R.A. Benson; River Valley Healthcare, Silvis, IL: K.Carroll; Family Physician Incorporated, N. Canton, OH: H. Marshall; Internal Medicine of Northern Michigan, Petoskey, MI: P.D. Blanchard; Redmond Internal Medicine, Redmond, OR: D. Palmer, C. Gangan; Grove Hill Medical Center, New Britain, CT: M.S. Werner; Olean Medical Group, Olean, NY: H.D. Storch, T.L. Buzzard; Internal Medicine Associates of Greenville, Greenville, SC: J.S. Moore; Magan Medical Clinic, Covina, CA: R. Sakamoto; Owatonna Clinic–Mayo Health System, Owatonna, MN: T. Price; Dearborn Cardiology, Dearborn, MI: S. Dabbous; Westchester Medical Group, White Plains, NY: B. Newman; Central Cardiology Medical Clinic, Bakersfield, CA: W. Nyitray; Salem Clinic, Salem, OR: M. Smith; East Carolina University, Greenville, NC: C. Estrada; Northwest Primary Care Group, Milwaukie, OR: D. McAnulty, P. Devisser; The William W Backus Hospital, Norwich, CT: S. Johnson; Jefferson City Medical Group, Jefferson City, MO: C. Balcer; Saint Louis University Department of Neurology, St. Louis, MO: S. Cruz-Flores, E. Holzemer; Wellspan Health–Yorktowne, York, PA: J.D. Horton; Mercy Medical Center, Canton, OH: M. Cudnik; Cardiovascular Group, Lawrenceville, GA: B. Craig-Allen; Asheville Cardiology Assoc, Asheville, NC: W. Wharton, A. Moser; Cardiac Consultants Chartered, Bethesda, MD: L. Chappell; Valley Care Health System, Pleasanton, CA: N. Huynh; Bloomington Hospital, Bloomington, IN: K. Kalotta; Samaritan Anticoagulation Service, Corvallis, OR: R. Stockberger; Covenant Clinic, Waterloo, IA: D. Kohls; Dartmouth-Hitchcock Nashua, Nashua, NH: L. Cook; Cardiology Consultants, PC, Hamden, CT: A.M. Radoff; Seventh Avenue Family Health Center, Fort Lauderdale, FL: J. Berges; Diagnostic Cardiology, PA, Jacksonville, FL: P.D. Kuhlman; Norlanco Medical Associates, Elizabethtown, PA: J. Rittenhouse; University of Texas Medical Branch, Galveston, TX: H. von Marensdorff; Bend Memorial Clinic, Bend, OR: M. Hegewald; Memorial Primary Care Center, Hollywood, FL: J. Beck; Batey Cardiovascular Center, Bradenton, FL: D. Calabrila, E.J.Sanchez; Western Montana Clinic, Missoula, MT: W.B. Bekemeyer, D. Ramsey; Winona Clinic, Winona, MN: L. Tschumper; Cardiac Consultants, Lancaster, PA: M. Lesko; Hattiesburg Clinic, Hattiesburg, MS: A.J. Jackson; Bryn Mawr Medical Specialist Association, Bryn Mawr, PA: H. Mayer; River Valley Healthcare, Moline, IL: B. Cady; Cardiovascular Group, Snellville, GA: L. Lesser; Medicor, Bridgewater, NJ: P. Saulino, C. Hartpence; Bond Clinic, PA, Winter Haven, FL: P. Lundsford. K. Bhatia; University of Cincinnati-Pharmacy Anticoagulation Services, Cincinnati, OH: J. McQueen; Senior Healthcare Center, Gainesville, FL: M.L. Breeser; North Canton Medical Foundation, North Canton, OH: H.M. Schenker; Manor Family Health Center, Millersville, PA: J. Ichter; Cardiology Associates of Central Florida, Ocala, FL: L. McDaniel; Cardiovascular Associates Ltd, Chesapeake, Virgina: S.R. Jones; Woodburn Medical Clinic, Woodburn, OR: F. Golden; Rockwood Clinic North, Spokane, WA: C. Laudenbach, J.S. Pennock; Wachspress, Shatkin & Rainear, Vineland, NJ: L. Assink; Chambersburg Hospital, Chambersburg, PA: D. Grant; Wellspan Pharmacy-Dallastown, Dallastown, PA: T.G. Williams; Pulmonary & Critical Care Associates, Ypsilanti, MI: W.F. Patton; Island Cardiac Specialist, Mineola, NY: P. Ragno; Portland Cardiovascular Institute 2, Portland, OR: R. Chelfky; River Valley Healthcare ACS, Bettendorf, IA: W. Langley; Consultants in Cardiology, Farmington Hills, MI: G.M. McKendrick; Portland Cardiovascular Institute 1, Portland, OR: R. Chelfky; Cleveland Clinic Florida, Weston, FL: B. Fernandez; BiState Medical Consultants, St. Louis, MO: P.M. Stein, C.B. Lomnel; Medical Consultants, PC, Muncie, IN: J. Bow; Cardiovascular Associates of South Florida, Coral Gables, FL: J.S. Palmer; Parkway Cardiology Associates, Oak Ridge, TN: S. Cooke; Northwest Georgia Diagnostic Clinic, Gainesville, GA: J. Jackson; Cardiovascular Associates, Kingsport, TN: L.H. Cox; Heart Place, Dallas, TX: C.N. Bowers; Rockwood Clinic, Spokane, WA: C. Laudenbach; J.S. Pennock; Delaware Heart Group, Newark, DE: C. Bowens; and Abilene Diagnostic Clinic, Abilene, TX: P. Howard.

References

1 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

2 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

3 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

4 Investigators AF. 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

5 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

6 Brass LM, Krumholz HM, Scinto JM, et al. Warfarin use among patients with atrial fibrillation. Stroke 1997; 28:2382–2389

7 Munschauer FE, Priore RL, Hens M, et al. Thromboembolism prophylaxis in chronic atrial fibrillation: practice patterns in community and tertiary-care hospitals. Stroke 1997; 28: 72–76

8 Smith NL, Psaty BM, Furberg CD, et al. Temporal trends in the use of anticoagulants among older adults with atrial fibrillation. Arch Intern Med 1999; 159:1574–1578

9 Kutner M, Nixon G, Silverstone F. Physicians' attitudes toward oral anticoagulants and antiplatelet agents for stroke prevention in elderly patients with atrial fibrillation. Arch Intern Med 1991; 151:1950–1953

10 Hylek EM, Regan S, Go AS, et al. 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 Fihn SD, Callahan CM, Martin DC, et al. The risk for and severity of bleeding complications in elderly patients treated with warfarin: the National Consortium of Anticoagulation Clinics. Ann Intern Med 1996; 124:970–979

12 Gurwitz JH, Avorn J, Ross-Degnan D, et al. Aging and the anticoagulant response to warfarin therapy. Ann Intern Med 1992; 116:901–904

13 Wynne H, Cope L, Kelly P, et al. The influence of age, liver size and enantiomer concentrations on warfarin requirements. Br J Clin Pharmacol 1995; 40:203–207

14 Wynne HA, Kamali F, Edwards C, et al. Effect of ageing upon warfarin dose requirements: a longitudinal study. Age Ageing 1996; 25:429–431

15 Redwood M, Taylor C, Bain BJ, et al. The association of age with dosage requirement for warfarin. Age Ageing 1991; 20:217–220

16 James AH, Britt RP, Raskino CL, et al. Factors affecting the maintenance dose of warfarin. J Clin Pathol 1992; 45:704–706

17 Landefeld CS, Goldman L. Major bleeding in outpatients treated with warfarin: incidence and prediction by factors known at the start of outpatient therapy. Am J Med 1989; 87:144–152

18 van der Meer FJ, Rosendaal FR, Vandenbroucke JP, et al. Bleeding complications in oral anticoagulant therapy: an analysis of risk factors. Arch Intern Med 1993; 153:1557–1562
19 Palareti G, Leali N, Coccheri S, et al. Bleeding complications of oral anticoagulant treatment: an inception-cohort, prospective collaborative study (ISCOAT); Italian Study on Complications of Oral Anticoagulant Therapy. Lancet 1996; 348: 423–428
20 Hylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. Ann Intern Med 1994; 120:897–902
21 Ansell J, Hirsh J, Poller L, et al. The pharmacology and management of the vitamin K antagonists: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 2004; 126(suppl):204S–233S
22 Gage BF, Eby C, Milligan PE, et al. Use of pharmacogenetics and clinical factors to predict the maintenance dose of warfarin. Thromb Haemost 2004; 91:87–94
23 Fennerty A, Dolben J, Thomas P, et al. Flexible induction dose regimen for warfarin and prediction of maintenance dose. Br Med J (Clin Res Ed) 1984; 288:1268–1270
24 Gedge J, Orme S, Hampton KK, et al. A comparison of a low-dose warfarin induction regimen with the modified Fennerty regimen in elderly inpatients. Age Ageing 2000; 29: 31–34
25 Roberts GW, Druskeit T, Jorgensen LE, et al. Comparison of an age adjusted warfarin loading protocol with empirical dosing and Fennerty's protocol. Aust N Z J Med 1999; 29:731–736
26 Fuster V, Ryden LE, Asinger RW, et al. ACC/AHA/ESC Guidelines for the Management of Patients With Atrial Fibrillation: Executive Summary A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines and Policy Conferences (Committee to Develop Guidelines for the Management of Patients With Atrial Fibrillation) Developed in Collaboration With the North American Society of Pacing and Electrophysiology. Circulation 2001; 104:2118–2150
27 Fihn SD, McDonell M, Martin D, et al. Risk factors for complications of chronic anticoagulation: a multicenter study; Warfarin Optimized Outpatient Follow-up Study Group. Ann Intern Med 1993; 118:511–520
28 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
29 O'Reilly RA. The stereoselective interaction of warfarin and metronidazole in man. N Engl J Med 1976; 295:354–357
30 O'Reilly RA. Stereoselective interaction of trimethoprim-sulfamethoxazole with the separated enantiomorphs of racemic warfarin in man. N Engl J Med 1980; 302:33–35
31 Wells PS, Holbrook AM, Crowther NR, et al. Interactions of warfarin with drugs and food. Ann Intern Med 1994; 121: 676–683
32 Cuzick J. A Wilcoxon-type test for trend. Stat Med 1985; 4:87–90
33 Hirsh J, Poller L. Practical dosing considerations with warfarin and optimal therapeutic range. In: Poller L, Hirsh J, eds. Oral anticoagulants. London, UK: Arnold, 1996; 167–179
34 Chan E, McLachlan AJ, Pegg M, et al. Disposition of warfarin enantiomers and metabolites in patients during multiple dosing with rac-warfarin. Br J Clin Pharmacol 1994; 37:563–569
35 Routledge PA, Chapman PH, Davies DM, et al. Pharmacokinetics and pharmacodynamics of warfarin at steady state. Br J Clin Pharmacol 1979; 8:243–247
36 Woodhouse KW, James OF. Hepatic drug metabolism and ageing. Br Med Bull 1990; 46:22–35
37 Schmucker DL. Aging and the liver: an update. J Gerontol A Biol Sci Med Sci 1998; 53:B315–B320
38 Sotaniemi EA, Arranto AJ, Pelkonen O, et al. Age and cytochrome P450-linked drug metabolism in humans: an analysis of 226 subjects with equal histopathologic conditions. Clin Pharmacol Ther 1997; 61:331–339
39 Le Couteur DG, McLean AJ. The aging liver: drug clearance and an oxygen diffusion barrier hypothesis. Clin Pharmacokinet 1998; 34:359–373
40 Loguercio C, Taranto D, Vitale LM, et al. Effect of liver cirrhosis and age on the glutathione concentration in the plasma, erythrocytes, and gastric mucosa of man. Free Radic Biol Med 1996; 20:483–488
41 Schwartz JB. The influence of sex on pharmacokinetics. Clin Pharmacokinet 2003; 42:107–121
42 Niemela O, Parkkila S, Pasanen M, et al. Induction of cytochrome P450 enzymes and generation of protein-aldehyde adducts are associated with sex-dependent sensitivity to alcohol-induced liver disease in micropigs. Hepatology 1999; 30:1011–1017
43 Kovacs MJ, Rodger M, Anderson DR, et al. Comparison of 10-mg and 5-mg warfarin initiation nomograms together with low-molecular-weight heparin for outpatient treatment of acute venous thromboembolism: a randomized, double-blind, controlled trial. Ann Intern Med 2003; 138:714–719
44 Gage BF. Randomized trial of warfarin nomograms [letter]. Ann Intern Med 2004; 140:489
45 Crowther MA, Harrison L, Hirsh J. Randomized trial of warfarin nomograms [letter]. Ann Intern Med 2004; 140:490
46 Peterson GM, Jackson SL. Randomized trial of warfarin nomograms [letter]. Ann Intern Med 2004; 140:488
47 White R, Dager W. Randomized trial of warfarin nomograms [letter]. Ann Intern Med 2004; 140:488–489
48 Ageno W, Turpie AG, Steidl L, et al. Comparison of a daily fixed 2.5-mg warfarin dose with a 5-mg, international normalized ratio adjusted, warfarin dose initially following heart valve replacement. Am J Cardiol 2001; 88:40–44

Pap Smear Rates Among Haitian Immigrant Women in Eastern Massachusetts

ERIC H. GREEN, MD, MSc[a]
KAREN M. FREUND, MD, MPH[a]
MICHAEL A. POSNER, MS[a]
MICHELE M. DAVID, MD, MPH, MBA[a]

SYNOPSIS

Objective. Given limited prior evidence of high rates of cervical cancer in Haitian immigrant women in the U.S., this study was designed to examine self-reported Pap smear screening rates for Haitian immigrant women and compare them to rates for women of other ethnicities.

Methods. Multi-ethnic women at least 40 years of age living in neighborhoods with large Haitian immigrant populations in eastern Massachusetts were surveyed in 2000–2002. Multivariate logistic regression analyses were used to examine the effect of demographic and health care characteristics on Pap smear rates.

Results. Overall, 81% (95% confidence interval 79%, 84%) of women in the study sample reported having had a Pap smear within three years. In unadjusted analyses, Pap smear rates differed by ethnicity ($p=0.003$), with women identified as Haitian having a lower crude Pap smear rate (78%) than women identified as African American (87%), English-speaking Caribbean (88%), or Latina (92%). Women identified as Haitian had a higher rate than women identified as non-Hispanic white (74%). Adjustment for differences in demographic factors known to predict Pap smear acquisition (age, marital status, education level, and household income) only partially accounted for the observed difference in Pap smear rates. However, adjustment for these variables as well as those related to health care access (single site for primary care, health insurance status, and physician gender) eliminated the ethnic difference in Pap smear rates.

Conclusions. The lower crude Pap smear rate for Haitian immigrants relative to other women of color was in part due to differences in (1) utilization of a single source for primary care, (2) health insurance, and (3) care provided by female physicians. Public health programs, such as the cancer prevention programs currently utilized in eastern Massachusetts, may influence these factors. Thus, the relatively high Pap rate among women in this study may reflect the success of these programs. Public health and elected officials will need to consider closely how implementing or withdrawing these programs may impact immigrant and minority communities.

[a]Women's Health Unit, Section of General Internal Medicine, Evans Department of Medicine, Boston University School of Medicine, Boston, MA

Address correspondence to: Eric H. Green, MD, MSc, Dept. of Medicine, Montefiore Medical Center, 111 E. 210 St., Bronx, NY 10467; tel. 718-920-7332; e-mail <ergreen@montefiore.org>.

The incidence of and mortality from invasive cervical cancer in the United States have decreased since the 1970s.[1] However, disparities in mortality from this preventable disease are found when African American or Latina women are compared to non-Hispanic white women.[1–4] These findings may be due to later detection of cervical cancer in minority women. A retrospective review of data from a tumor registry[5] and data from a case series[6] suggest that the burden of cervical cancer is higher in the Haitian immigrant community than in other communities of color. There is some evidence of lower Pap smear rates among women who self-identified as Haitian than among other women of color in New York City;[7] however, little direct research has been done on cervical cancer in the Haitian immigrant community. This reflects the difficulty in studying this community. Haitian immigrants are poorly sampled by large national surveys because of their small numbers and geographical concentration in a few metropolitan areas,[8,9] as well as their inclusion in groups labeled "black" or "African American" in some large databases.[10] Further, many Haitians in the U.S. speak primarily Haitian Creole, requiring additional efforts on the part of survey organizers to make their surveys accessible to this population.

For the present study, we examined Pap smear rates for women living in neighborhoods with substantial Haitian immigrant populations in four cities in eastern Massachusetts. We determined whether Haitian immigrant women had lower screening rates than other women residing in the same neighborhoods and identified the demographic and health care characteristics associated with receipt of screening.

METHODS

We conducted a cross-sectional, community-based survey of women living in four cities in eastern Massachusetts who were 40 years of age or older and spoke English, Haitian Creole, or Spanish. The survey used an area probability sample to identify subjects. We first generated canvass maps of potential survey areas in four eastern Massachusetts cities (Boston, Cambridge, Brockton, and Somerville) using information from community-based agencies, businesses, the city of Boston Assessing Department's Property Parcel Data for fiscal year 1997, and Property Parcel Data for fiscal year 1999 from the Assessing Departments of the cities of Brockton, Cambridge, and Somerville. After consultation with Haitian community leaders, we used these maps to generate, for each of the four cities, an initial list of city blocks where the community leaders believed at least 20% of households included Haitian immigrants. We supplemented this list with city blocks where at least two study participants believed that 20% of households included Haitian immigrants.

Trained multilingual interviewers visited randomly selected households from 2000–2002. In households with multiple eligible women, the youngest woman meeting the eligibility criteria completed the survey. Interviewers visited each household up to four times before it was eliminated from consideration for inclusion in the study and a new household was selected. No more than 12 women were surveyed from any individual city block.

Each participant completed a 30- to 45-minute interviewer-administered survey using a standardized instrument in English, Haitian Creole, or Spanish, according to her preference. The survey included questions on demographics, health care access, and preventive health practices. We adapted questions from the Cancer Control Needs in Multi-Ethnic Communities study,[11] the National Health Interview Survey 1992 epidemiological and cancer control supplements,[12] and a prior survey administered to 332 similar women.[13] We translated all questions from English into Haitian Creole and Spanish and independently back-translated them to English to ensure the accuracy of the initial translation. We piloted the survey with women from diverse cultural backgrounds and revised the questionnaire based on these sessions. The Boston University School of Medicine Institutional Review Board approved all aspects of this study.

Women were asked to self-identify their ethnicity and race. The interviewer first asked, "How would you describe your ethnic group?" Respondents were offered 19 specific response choices as well as "other," "don't know," and "refused." Respondents who indicated "other" were asked for a further description of their ethnicity. A separate question asked about race: "How do you define your race? Would you say you're black; American Indian, Eskimo, and Aleut; Asian or Pacific Islander; white; other race [specify]; don't know; refused."

Women were also asked to specify their country of birth, their mother's and father's countries of birth, their first language, and the country that they "most identif[ied] with."

We grouped women into five major ethnic groups: "Haitian," "African American," "Latina," "English-speaking Caribbean," and "white, not Hispanic," based on self-report. We attempted to classify the 114 (16%) women who reported their ethnicity as "other" or who listed multiple ethnicities into these categories and were able to do so for 76 of the 114 women. We based this categorization on the women's self-reported race and country of origin, with secondary reliance on their parents' countries of origin. For example, one women who self-identified as "black" reported that she had been born in Haiti to Haitian parents and had grown up there speaking Haitian Creole; we classified her as Haitian for the present study.

We grouped annual household income into four categories (<$20,000, $20,000–$40,000, >$40,000, and unknown) and analyzed age groups by decades (40–49 years, 50–59 years, . . . , ≥80 years.). We considered women who reported having had a Pap smear within three years prior to the interview to have had a recent Pap smear.

We performed bivariate analyses using chi-square or Fisher's exact tests, as required, for categorical variables. We used analysis of variance to analyze the relationship of race with age to determine the statistical significance of any age differences between racial/ethnic groups. We constructed two multiple logistic regression models to assess Pap smear rates, controlling for other factors. Our first model adjusted for age and demographic factors known to affect Pap smear rates in other populations (age, marital status, education level, and household income).[14–20] Our second model included all variables in the first study, as well as elements of health care delivery known to influence Pap smear rates (insurance status, physician gender, and single site for primary care).[7,14,15,21] We performed extensive sensitivity analyses to confirm that our findings were not overly dependent

on assumptions in variable definition. All statistical analyses were performed using SAS,[22] and all p-values are two-sided. We defined statistical significance as $p<0.05$.

RESULTS

We identified 2,304 potentially eligible addresses. We excluded 161 (7%) because of vacancy, safety concerns, or because no housing unit existed at that address. At an additional 369 (16%) addresses, no one was available on four separate visits. At 792 (34%) of the households visited, there were no eligible women. A total of 982 eligible women were identified (43% of visited households); of these women, 753 (77%) participated.

Among the 753 participants, 38 (5.0%) could not be classified into one of the five pre-defined ethnic groups, and 15 (2.0%) said they did not know when they had last received a Pap smear or did not answer this question. The remaining 700 women formed our study sample.

The characteristics of the sample are shown in Table 1. Of the 700 women, 278 (40%) were categorized as Haitian, 156 (22%) as African American, 52 (7%) as English-speaking Caribbean, 72 (10%) as Latina, and 142 (20%) as white, not Hispanic.

The Haitian group included 275 women who reported having been born in Haiti and who self-identified as "Haitian" or "Haitian American"; two women who self-identified as Haitian or Haitian American but reported having been born in other Caribbean countries (Dominican Republic, Cuba); and one women (as described above) who did not identify an ethnicity but reported having been born in Haiti to Haitian parents.

Seventy-five percent of the women in the study sample lived in Boston, and 87% reported having some form of health insurance. Most surveyed women had reportedly engaged in preventive health behaviors: 90% reported having had a "check-up visit" within two years, and 68% reported having had a mammogram within one year. Also, 87% reported that someone at their doctor's office, i.e., their health care provider, a member of the support staff, or an interpreter, was available to communicate with them in their first or preferred language.

Table 1. Characteristics of women in study sample (N=700 women ≥40 years of age)

Characteristic	n	Percent
Ethnic category		
Haitian[a]	278	40
African American	156	22
English-speaking Caribbean	52	7
Latina	72	10
White, not Hispanic	142	20
High school graduate	373	54
Married or living with a domestic partner	337	48
Annual household income		
>$40,000	86	12
$20,000–40,000	155	22
<$20,000	204	29
Unknown	255	36
Health insurance	591	87
Single site for primary care	667	95
Female physician[b]	303	45
Language access[c]	607	87
Lived in Boston	473	75
"Check-up" visit within 2 years	566	91
Mammogram within 1 year	435	69
Age (mean ± standard deviation)	55 ± 12	

[a]Included 275 women who reported having been born in Haiti who self-identified as "Haitian" or "Haitian American," two women who reported having been born in other Caribbean countries (Dominican Republic and Cuba) who self-identified as "Haitian," and one women who did not identify an ethnicity who reported having been born in Haiti to Haitian parents.

[b]n=671 who identified a usual provider.

[c]Physician, nurse, support staff, or translator available at primary care site who spoke first or preferred language.

We found differences in demographic and health care access variables across ethnicities (see Table 2). The Haitian immigrant women were less likely to have graduated from high school than African American, Latina, non-Hispanic white, or English-speaking Caribbean women ($p<0.01$ for each comparison) and were least likely to have health insurance or a female physician. They were also least likely to report that they had access to someone who spoke their language when they saw a physician. Haitian, Latina, and English-speaking Caribbean women were less likely to report their incomes than African American or non-Hispanic white women, thus justifying use of a four-level variable for income. White women were oldest and least likely to be married.

We found that 570 women reported having had a Pap smear within three years (81%; 95% confidence interval [CI] 79%, 84%). Pap smear rates differed by ethnicity ($p=0.003$). Women who identified as Latina, African American, and English-speaking Caribbean women had the highest rates, while women identified as Haitian had a somewhat lower Pap smear rate (78%; 95% CI 74%, 83%), and non-Hispanic white women had the lowest rate (74%; 95% CI 67%, 81%). Characteristics that made Pap smear acquisition more likely included being married or living with a domestic partner, having some form of health insurance, utilizing a single site of health care, and having a female physician (Table 3). Pap smear rates also differed by income and age.

Logistic regression models that adjusted for confounding accounted for much of the difference in Pap smear rates between women of different ethnicities (Table 4). In our models, which sequentially adjusted for demographics and health care delivery variables, the adjusted odds ratios (ORs) for having a recent Pap smear for Haitian women compared to African American women approached 1.0 and had progressively wider CIs. As we adjusted for more confounders, the overall statistical significance associated with ethnicity was progressively reduced ($p=0.003$ unadjusted; $p=0.03$ adjusted for demographics; and $p=0.19$ adjusted for

Table 2. Pap smear rates and selected demographic and health care access characteristics, by ethnicity (N=700 women ≥40 years of age)

Characteristic	Haitian (n=278)[a] Percent (95% CI)	African American (n=156) Percent (95% CI)	English-speaking Caribbean (n=52) Percent (95% CI)	Latina (n=72) Percent (95% CI)	White, not Hispanic (n=142) Percent (95% CI)	p-value
Pap smear within past 3 years	78 (74, 83)	87 (81, 92)	88 (80, 97)	92 (85, 98)	74 (67,81)	0.003
High school graduate	24 (20, 30)	80 (75, 87)	67 (54, 80)	46 (34, 57)	80 (73, 86)	<0.001
Married or living with a domestic partner	56 (50, 61)	43 (35, 50)	58 (44, 71)	47 (36, 54)	37 (29, 44)	0.001
Annual family income <$20,000	34 (28, 39)	22 (16, 29)	8 (1, 15)	25 (15, 35)	38 (30, 46)	<0.001
Health insurance	79 (74, 84)	94 (90, 98)	100	84 (76,93)	92 (87, 95)	<0.001
Single site for primary care	94 (92,97)	98 (96,100)	100	97 (93,100)	92 (87, 96)	0.03
Female physician[b]	38 (32, 43)	51 (45, 59)	46 (40, 67)	65 (54, 77)	42 (34, 51)	<0.001
Language access[c]	83 (78, 87)	99 (98, 100)	100	96 (92, 100)	96 (93, 98)	<0.001
Age (mean ± standard deviation)	54 ± 11	55 ± 9	56 ± 11	51 ± 10	61 ± 14	<0.001

[a]Includes 275 women who reported having been born in Haiti who self-identified as "Haitian" or "Haitian American," two women who reported having been born in other Caribbean countries (Dominican Republic and Cuba) who self-identified as "Haitian," and one women who did not identify an ethnicity who reported having been born in Haiti to Haitian parents.

[b]n=671 who identified a usual provider.

[c]Physician, nurse, support staff, or translator available at primary care site who spoke first or preferred language.

CI = confidence interval

demographics and health care delivery). Our second model produced a C-statistic (a measure of predictive validity) of 0.74, indicating that it discriminated well between women with and without Pap smears.

Age, physician gender, and insurance were statistically significant predictors of Pap smear acquisition in our final model. In sensitivity analyses, we found no statistically significant interactions between any of these factors and ethnicity. During further sensitivity analyses in which we varied variable definitions, we found that the adjusted ORs varied by less than 10% from those in the base model and the *p*-values for ethnicity were similar to those calculated in the base model.

DISCUSSION

In our population-based sample of multi-ethnic urban women at least 40 years old, self-reported Pap smear rates among all women were relatively high. In the study sample, women identified as Haitian had a lower Pap smear rate than women identified as African American, English-speaking Caribbean, or Latina. Of note, non-Hispanic white women who lived in the same neighborhoods in eastern Massachusetts as Haitian immigrants had a lower crude Pap smear rate than women of color. However, adjustment for differences in demographics and health delivery eliminated the association between ethnicity and Pap smears.

Ethnicity has been reported to be an important predictor of Pap smear acquisition. Multiple data sources suggest that African American women have slightly higher Pap smear rates than non-Hispanic white women.[15,19,23,24] Data from the National Health Interview Survey[16] and the Behavioral Risk Factor Surveillance System (BRFSS) Survey[15] suggest that Latina, Native American, and women classified as "other" have lower Pap smear screening rates than either African American or white women. However, there is little evidence directly comparing Pap smear rates of Haitian immigrant women to those of other black women. Fruchter et al. found the rate of *in situ* cervical carcinoma, which is almost always detected via Pap smears, to be lower in Haitian immigrant women than in other minority women.[25] In a telephone survey of minority and immigrant women completed in 1992 in New York, Mandelblatt and colleagues found that women identified as Haitian had the lowest Pap smear rate (69%) of any group sampled.[11] These data, along with the results of the present study, highlight differences among black populations, and reinforce the need for detailed reporting of race and ethnicity in monitoring the health of minority communities.[10]

Although the 81% Pap smear rate for a three-year period in the present study does not meet the Healthy People 2010 goal for 90% of women at least 18 years old to receive triennial Pap smears,[26] the Pap smear rate is higher than that seen in a recent study of similar low-income minority women.[11] Our data, along with state and local BRFSS data,[27,28] reveal higher Pap rates in Massachusetts and Boston than nationally. These higher rates may be attributed to expanded health care access in this region.[29] The services available in eastern Massachusetts during the study period included both government and non-government programs[30] such as broad Medicaid eligibility requirements, a general insurance "safety net" (the Massachusetts Uncompensated Care Pool),[31] a special program designed to increase cancer screening access (the Massachusetts Department of Public Health's Women's Health Network),[32] and a statutory requirement for translation services at hospital-based medical practices.[33] In addition, the relatively large number of teaching hospitals and community health centers in Boston and Cambridge may increase low-income women of color's access to medical care. The health care environment in eastern Massachusetts accounts for the many women in this study with insurance

Table 3. Self-reported Pap smear rates by selected demographic and health care access variables (N=700 women ≥40 years of age)

Characteristic	Pap smear within past 3 years	
	Percent	p-value
Overall	81	
Education		0.35
High school graduate	83	
Not high school graduate	80	
Marital status		0.005
Married or living with a domestic partner	86	
Not married	78	
Annual household income		<0.001
>$40,000	94	
$20,000–$40,000	85	
<$20,000	80	
Unknown	76	
Health insurance status		<0.001
Insured	84	
Not insured	65	
Single place of health care		<0.001
Yes	82	
No	55	
Provider gender[a]		<0.001
Female	89	
Male	78	
Age		<0.001
40–49	89	
50–59	84	
60–69	78	
70–79	62	
≥80	53	

[a]n=671 who identified a usual provider.

coverage and access to primary care offices where someone spoke their language.

Despite the relatively high overall rates of Pap smear acquisition, disparities still exist. Many studies have found that women are more likely to have had a Pap smear if they are younger, have more education, have higher incomes, or are married.[14–20] Others have found that higher Pap smear rates are associated with health insurance,[14,15] a usual site of care,[7] and a female physician.[21] In our data, the ethnic disparities observed in unadjusted analyses can be explained by ethnic differences in demographic and health delivery variables. Demographic variables such as age, income, education, and marital status are either unmodifiable or unlikely to be affected by public health programs. However, differences in Pap smear acquisition between women of different ethnicities were not eliminated by adjustment for age, income, and education. It was only after adjustment for differential insurance rates, care by a female physician, and use of a single site for primary care that the differences in Pap smear rates were fully explained. This implies that if these factors could be equalized across ethnicities, women of all ethnicities would have Pap rates similar to those seen in the African American, English-speaking, and Latina women in this study. Thus, if disparities in health care delivery could be minimized, Pap smear acquisition can approach the Healthy People 2010 goal of 90%.[26]

This study had several strengths. We achieved a high response rate in sampling a community-based population of multi-ethnic women. The interviewer-administered format in three languages allowed us to collect detailed data regarding ethnicity and cancer screening practices. Our use of an area-probability sampling strategy increased the likelihood that women of different ethnicities would be comparable in terms of unmeasured confounders such as access to public transit and exposure to public service announcements in the mass media.

This study also had several limitations. Ascertainment of Pap smear status was based on self-report. Although there is some evidence that self-report of Pap smear status has a low specificity,[34] most studies of Pap smear acquisition have used unverified self-report.[11,14,16,18,24,35–38] Our study was limited to

Table 4. Multivariate logistic regression models predicting a recent Pap smear (N=700 women ≥40 years of age)

Ethnic category	Crude and adjusted odds ratios (95% confidence intervals)		
	Unadjusted	Model 1[a]	Model 2[b]
Haitian[c]	0.57 (0.33, 0.97)	0.66 (0.34, 1.26)	0.82 (0.41, 1.67)
English-speaking Caribbean	1.19 (0.45, 3.14)	2.29 (0.71, 7.40)	2.07 (0.62, 6.94)
Latina	1.71 (0.66, 4.44)	2.09 (0.76, 5.74)	1.69 (0.60, 4.80)
White, not Hispanic	0.44 (0.24, 0.78)	0.69 (0.36, 1.35)	0.68 (0.33, 1.39)
African American	1.0 (Reference)	1.0 (Reference)	1.0 (Reference)
p-value[d]	0.003	0.03	0.19

[a]Adjusted for age, marital status, education level, and household income.

[b]Adjusted for age, marital status, education level, household income, insurance status, physician gender, and single place of care.

[c]Included 275 women who reported having been born in Haiti who self-identified as "Haitian," or "Haitian American," two women who reported having been born in other Caribbean countries (Dominican Republic and Cuba) who self-identified as "Haitian," and one women who did not identify an ethnicity who reported having been born in Haiti to Haitian parents.

[d]Overall difference in Pap smear rates by ethnicity

women at least 40 years of age and may not generalize to younger women. We also included women older than 65. Cervical cancer screening in this age group is controversial;[39] however, older minority women have a relatively high rate of abnormal Pap smears[40,41] and are at higher risk for invasive cervical cancer than younger or white women.[42] We also stratified by age, creating separate models for women older or younger than 65 (not shown). Although our power for this analysis was limited, it suggests that ethnic differences in Pap smear rates are more pronounced among older women. Further research will be needed to better characterize any possible interaction of age and ethnicity with respect to Pap smear acquisition.

The higher incidence of invasive cervical cancer in Haitian immigrant women[5,6,25,43] has been attributed to lower rates of Pap smear acquisition in this community. However, in the present study, the initially observed relationship of ethnicity to receipt of Pap smears was explained by health care delivery variables and specific demographic characteristics. Many of the health care delivery variables are potentially modifiable by public health programs, and such programs may account for the relatively high Pap smear rates seen in this study population. Thus, our data suggest that a health care system with government mandate and funding to serve the needs of limited-English-speaking immigrants can succeed. Other regions with lower Pap smear rates may be able to use this model to improve preventive health delivery to immigrant women. At the same time, however, these programs will be challenged to continue operation in a more fiscally restricted era. Public health and elected officials will need to consider closely how implementing or withdrawing these programs may impact immigrant and minority communities.

The authors thank Linda Ko, MPH, and Arlene Ash, PhD, for contributions to the study design and analysis as well as for reviewing drafts of the manuscript.

This study was presented in a poster session at the Annual Meeting of the Society of General Internal Medicine, May 2–4, 2002, Atlanta, Georgia.

REFERENCES

1. American Cancer Society. Cancer facts and figures: 2003. Atlanta: American Cancer Society; 2003.
2. U.S. mortality age-adjusted rates, total U.S., 1990–1999 [cited 2003 Feb 2]. Available from: URL: http://canques.seer.cancer.gov/cgi-bin/cq_submit?dir=seer1999&db=8&rpt=TAB&sel=^0^43^1^^2^0&x=Race/ethnicity^0,1,2,3,4,5&y=Statistic%20type^1,2,3&dec=3,1,1
3. SEER incidence age-adjusted rates, 11 registries, 1992–1999 [cited 2003 Feb 2]. Available from: URL: http://canques.seer.cancer.gov/cgi-bin/cq_submit?dir=seer1999&db=2&rpt=TAB&sel=^0^^53^0^^2^0&x= SEER%20registry^0,1,2,3,4,5,6,7,8,9,10,11,12^Race/ethnicity^0,1,2,3,4,5&y=Statistic%20type^1,2,3&dec=3,1,1
4. Jemal A, Thomas A, Murray T, Thun M. Cancer statistics, 2002. CA Cancer J Clin 2002;52:23-47.
5. Pitchenik AE, Becker DM, Hilsenbeck SG, Trapido EJ. Cancer among Haitians in Florida. Cancer 1987;59:184-7.
6. Fruchter RG, Remy JC, Burnett WS, Boyce JG. Cervical cancer in immigrant Caribbean women. Am J Public Health 1986;76:797-9.
7. O'Malley AS, Mandelblatt J, Gold K, Cagney KA, Kerner J. Continuity of care and the use of breast and cervical cancer screening services in a multiethnic community. Arch Intern Med 1997;157:1462-70.
8. Census Bureau (US). American FactFinder. PCT16. Ancestry (first ancestry reported) [cited 2004 May 19]. Available from: URL: http://factfinder.census.gov/servlet/DTTable?_bm=y&-geo_id=01000US&-ds_name=DEC_2000_SF3_U&-_lang=en&-state=dt&-format=&-mt_name=DEC_2000_SF3_U_PCT016
9. Trinity College Haiti Program. The emerging presence in the U.S. of the Haitian diaspora and its impact on Haiti: a symposium report [cited 2004 May 19]. Available from: URL: http://www.trinitydc.edu/academics/depts/interdisc/international/pdf%20files/report_%20final_%20version.pdf
10. Williams DR, Jackson JS. Race/ethnicity and the 2000 Census: recommendations for African American and other black populations in the United States. Am J Public Health 2000;90:1728-30.
11. Mandelblatt JS, Gold K, O'Malley AS, Taylor K, Cagney K, Hopkins JS, Kerner J. Breast and cervical cancer screening among multiethnic women: the role of age, health, and source of care. Prev Med 1999;28:418-25.
12. Current estimates from the National Health Interview Survey, 1992. Vital Health Stat 10 1994;189.
13. David MM, Rachelle J-B, Jeanty G, Laws MB, Prudent N. Screening mammography rates in Boston Haitian neighborhoods. J Gen Intern Med 1999;14(Suppl 2):23.
14. Segnan N. Socioeconomic status and cancer screening. In: Kogevinas M, Pearce N, Susser M, Boffetta P, editors. Social inequalities and cancer. Lyon (France): International Agency for Research on Cancer; 1997. p. 369-76.
15. Blackman DK, Bennett EM, Miller DS. Trends in self-reported use of mammograms (1989–1997) and Papanicolaou tests (1991–1997)—Behavioral Risk Factor Surveillance System. MMWR Morb Mortal Wkly Rep 1999;48(SS-06):1-22.
16. Calle EE, Flanders D, Thun MJ, Martin LM. Demographic predictors of mammography and Pap smear screening in U.S. women. Am J Public Health 1993;83:53-60.
17. Katz SJ, Hofer TP. Socioeconomic disparities in preventive care persists despite universal coverage: breast and cervical cancer screening in Ontario and the United States. JAMA 1994;272:530-4.
18. Sung JF, Blumenthal DS, Coates RJ, Alema-Mensah E. Knowledge, beliefs, attitudes, and cancer screening among inner-city African-American women. J Natl Med Assoc 1997;89:405-11.
19. Wilcox LS, Mosher WD. Factors associated with obtaining health screening among women of reproductive age. Public Health Rep 1993;108:76-86.
20. Goel V. Factors associated with cervical cancer screening: results from the Ontario Health Survey. Can J Public Health 1994;85:125-7.
21. Lurie N, Slater J, McGovern P, Ekstrum J, Quam L, Margolis K. Preventive care for women: does the sex of the physician matter? N Engl J Med 1993;329:478-82.
22. SAS Institute, Inc. SAS: Version 8.2. Cary (NC): SAS Institute, Inc.; 2001.
23. American Cancer Society. Cancer prevention and early detection: facts and figures 2003. Atlanta: American Cancer Society; 2003.
24. American Cancer Society. Cancer facts and figures for African-Americans 2000–2001. Atlanta: American Cancer Society; 2001.
25. Fruchter RG, Nayeri K, Remy JC, Wright C, Feldman JG, Boyce JG, Burnett WS. Cervix and breast cancer incidence in immigrant Caribbean women. Am J Public Health 1990;80:722-4.
26. Cancer. In: Department of Health and Human Services (US). Healthy People 2010. 2nd ed. Washington: DHHS; 2000. p. 3-1–3-32.
27. Massachusetts Department of Public Health, Bureau of Health Statistics, Research and Evaluation, Health Survey Program. Women's health in Massachusetts: highlights from the Massachusetts Behavioral Risk Factor Surveillance System (BRFSS): 2000–2001 [cited 2003 Mar 8]. Available from: URL: http://www.mass.gov/dph/bhsre/cdsp/brfss/wmnhealth.ppt
28. Massachusetts Department of Public Health. Massachusetts Community Health Information Profile (MassCHIP). Custom reports: Behavioral Risk Factor Surveillance System (BRFSS): Boston (v3.0 r303.0) [cited 2003 Mar 8]. Available from: URL: http://masschip.state.ma.us
29. Freund KM. Women's health and Medicaid reform. Womens Health Issues 2003;13:45-6.

30. Ruthardt L. Access to health care in Massachusetts. Boston: Massachusetts Division of Health Care Finance and Policy; 2002.
31. Ruthardt L. Uncompensated Care Pool: PFY 2001 annual report. Boston: Massachusetts Division of Health Care Finance and Policy; 2002 Aug.
32. Massachusetts Department of Public Health, Bureau of Family and Community Health. Women's Health Network [cited 2003 Jul 28]. Available from: URL: http://www.state.ma.us/dph/fch/whn/index.htm
33. Interpreter Service Requirement, 105 C.M.R. Sect. 130.1100 (2001).
34. Newell S, Girgis A, Sanson-Fisher R, Ireland M. Accuracy of patients' recall of Pap and cholesterol screening. Am J Public Health 2000;90:1431-5.
35. Fruchter RG, Wright C, Habenstreit B, Remy JC, Boyce JG, Imperato PJ. Screening for cervical and breast cancer among Caribbean immigrants. J Community Health 1985;10:121-35.
36. Wismer BA, Moskowitz JM, Chen AM, Kang SH, Novotny TE, Min K, et al. Rates and independent correlates of Pap smear testing among Korean-American women. Am J Public Health 1998;88:656-60.
37. Maxwell AE, Bastani R, Warda US. Demographic predictors of cancer screening among Filipino and Korean immigrants in the United States. Am J Prev Med 2000;18:62-8.
38. Zambrana RE, Breen N, Fox Sa, Gutierrez-Mohamed ML. Use of cancer screening practices by Hispanic women: analysis by subgroup. Prev Med 1999;29:466-77.
39. Hartmann KE, Hall SA, Nanda K, Boggess JF, Zolnoun D. Screening for cervical cancer. Systematic Evidence Review No. 25. Rockville (MD): Agency for Healthcare Research and Quality; 2002.
40. Mandelblatt J, Gopaul I, Wistreich M. Gynecological care of elderly women: another look at Papanicolaou smear testing. JAMA 1986; 256:367-71.
41. Mandelblatt J, Traxler M, Lakin P, Kanetsky P, Thomas L, Chauhan P, et al. Harlem Study Team. Breast and cervical cancer screening of poor, elderly, black women: clinical results and implications. Am J Prev Med 1993;9:133-8.
42. Mandelblatt J, Andrews H, Kerner J, Zauber A, Burnett W. Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type. Am J Public Health 1991;81:646-9.
43. Mitacek EJ, St Vallieres D, Polednak AP. Cancer in Haiti 1979–84: distribution of various forms of cancer according to geographical area and sex. Int J Cancer 1986;38:9-16.

ORIGINAL ARTICLE

Reforming Internal Medicine Residency Training

A Report from the Society of General Internal Medicine's Task Force for Residency Reform

Eric S. Holmboe, MD, Judith L. Bowen, MD, Michael Green, MD, MSc, Jessica Gregg, PhD, MD, Lorenzo DiFrancesco, MD, Eileen Reynolds, MD, Patrick Alguire, MD, David Battinelli, MD, Catherine Lucey, MD, Daniel Duffy, MD

DOI: 10.1111/j.1525-1497.2005.0249.x
J GEN INTERN MED 2005; 20:000–000.

The structure, process, and outcomes of internal medicine residency training have concerned the profession for over 20 years.[1–9] Over the last decade the initiative to move to outcomes-based education redefined the competencies physicians should obtain during training.[10,11] The core principle of outcomes-based education is the objective demonstration that a graduating trainee, whether from medical school or a residency, possesses the knowledge, skills, and attitudes necessary to progress to the next stage of his or her professional career.[12,13] The Accreditation Council for Graduate Medical Education (ACGME) and the Institute of Medicine (IOM) have defined core competencies for physicians shown in Table 1.[10,14] While both the ACGME and IOM provide a framework for the desired outcomes, medical educators bear the burden of designing the structures and processes to achieve them.[15]

Educators face several key challenges in redesigning residency programs. First, residency programs must prepare trainees for a variety of general internal medicine and subspecialty careers. Second, the settings and resources for residency training are highly heterogeneous. Third, an aging and increasingly diverse population, combined with rapidly expanding medical information and procedural technology, challenges all internists to acquire and maintain the knowledge, skills, attitudes, and performance necessary to provide high-quality care within their chosen discipline.[16,17] Finally, growing public dissatisfaction, substantial health care disparities, increased acuity but shorter lengths of stay for hospitalized patients, new work hour requirements, increasing medical student debt, and changing student demographics and lifestyle concerns further complicate residency reform.[18–25]

To provide recommendations for residency reform. The Society of General Internal Medicine (SGIM) convened a task force consisting of physicians representing a broad range of views within general medicine, expertise and experience in clinical education, and who represented internal medicine organizations outside of SGIM (Appendix 1). The task force focused on reform in 5 specific areas: ambulatory education, inpatient education, residency curriculum, health disparities, and life-long learning skills. To prepare this report, 4 subcommittees performed literature reviews that guided a prospective, systematic process to develop the final recommendations. The guiding principles, task force timeline, and the specific findings of the 4 subcommittees can be viewed at www.sgim.org. We acknowledge this report cannot cover all important aspects of residency training. The task force enthusiastically welcomes comments from other educators and internal medicine specialty organizations. Only through active collaboration and serious dialogue can we improve residency training.

TRAINING IN THE AMBULATORY SETTING

Recognizing the need for greater emphasis in outpatient training is not new.[1,5,26–29] The outpatient setting is where patients now receive most of their acute, chronic disease, and preventive care. Hospitalized patients are often discharged before many conditions have been fully evaluated or treated. Ambulatory settings, particularly continuity clinic settings, provide the ideal location for training in several key IOM competencies: learning to provide care based on continuous healing relationships, patient-centered care based on patient needs and values with the patient as the source of control, learning and designing systems of care that anticipate patients' needs, and learning to work in teams that model cooperation among clinicians (and nonclinician team members), including collaboration, coordination, and exemplary communication.[14] Although they are the essence of General Internal Medicine, these competencies apply equally for all internists.[14,17]

Two different types of ambulatory training occur in internal medicine residency programs: continuity clinic, where residents care for a panel of patients over time, and concentrated ambulatory block rotations. The learning in continuity clinic is experiential and is often augmented by case-based teaching conferences before or after clinics.[30,31] The ambulatory block rotations provide a venue for developing competence in managing the transitions and coordination of care within internal medicine, and experience with nonmedicine specialty care such as office orthopedics, gynecology, urology, ophthalmology, dermatology, and otolaryngology. Teaching strategies used in ambulatory blocks include didactic seminars teaching the principles of ambulatory medicine, case-based teaching sessions, and experiential learning with patients in other specialty clinics such as rheumatology, adolescent medicine, geriatrics, and women's health.[32–34]

Research has shown that graduating residents feel uncomfortable in managing common chronic conditions such as diabetes in the ambulatory setting, suggesting that the quality

The authors have no conflicts of interest to report.

Author's note: For author affiliations, please see Appendix 1.

Address correspondence and requests for reprints to Holmboe: American Board of Internal Medicine, 510 Walnut Street, Suite 1700, Philadelphia, PA 19106 (e-mail: eholmboe@abim.org).

Received for publication May 23, 2005
and in revised form July 26, 2005
Accepted for publication July 26, 2005

Table 1. Comparison of the IOM and ACGME Competencies

IOM Competencies	ACGME Competencies
Provide patient-centered care	Patient care, professionalism, interpersonal skills, and communication
Work in interdisciplinary teams	Professionalism, interpersonal skills and communication, systems-based practice
Utilize informatics	Patient care, practice-based learning and improvement, systems-based practice
Employ evidence-based medicine	Patient care, medical knowledge, practice-based learning and improvement
Apply quality improvement	Practice-based learning and improvement, systems-based practice

ACGME, Accreditation Council for Graduate Medical Education; IOM, Institute of Medicine.

of ambulatory education should be improved.[5,29,35,36] We identified several challenges to teaching and learning in ambulatory settings. First, there is often inadequate infrastructure to allow residents to provide high-quality, longitudinal care for complicated patients.[37–39] Residency clinics frequently lack the team structures needed to provide patient care when the primary resident is unavailable.[40–42] Also lacking are systems to monitor the quality of care provided to the residents' patients. The current approach to training in continuity care attenuates residents' growth toward independent practice and their ability to work in teams.[43–46]

Second, the residency clinics of academic medical centers typically attract a patient population with a disproportionate share of complex medical and psychosocial issues. The patient with multiple serious medical problems complicated by poverty, illiteracy, and substance abuse may overwhelm the clinical abilities of an internist in training, particularly in the absence of multidisciplinary resources. This is particularly true for interns. Too often they are assigned to care for a graduating senior's patient panel, containing patients whose conditions may be too complex for a novice to manage properly while they are learning the fundamentals of outpatient medicine.

Third, it is difficult for residents to develop expertise in continuity of care because they spend an insufficient amount of time in the outpatient arena. In most residency programs, residents attend their continuity clinic for only one half-day per week. Continuity clinic sessions are frequently cancelled when the resident is assigned to intensive care and night float rotations, limiting total continuity clinic exposure during training. To highlight the point, graduating residents starting a new outpatient practice will spend more time in the ambulatory setting in their first 3 months of practice than they do during an entire 3-year residency.

Finally, the quality and quantity of learning in ambulatory block rotations can be variable. Residents are often relegated to the role of observers during brief assignments in nonmedicine clinics, where programs rely on noninternal medicine specialists to donate teaching time. Residency programs also tend to use residents assigned to ambulatory block rotations as a workforce reservoir when unexpected vacancies on hospital services (e.g., the assigned resident is sick) require coverage. Vacations are frequently assigned during these rotations. Both factors contribute to a lack of consolidation of important skills. We found little information about ambulatory training during specialty consult rotations, that may be an important untapped resource for ambulatory skill development.

TRAINING IN THE INPATIENT SETTING

The inpatient setting is essential for educating residents about the care of acutely and critically ill patients. The hospital inpatient service has been the predominant setting of internal medicine education for over 100 years.[47] It is not surprising, therefore, that graduates of internal medicine residency programs report feeling most prepared to care for the acutely ill hospitalized patient.[29] Multiple changes in the health care system are significantly affecting the inpatient training experience.

First, hospitalized patients are sicker yet spend increasingly shorter periods of time in the hospital.[19,48] Residents have little time to establish a healing relationship with their patients and only infrequently learn about patient outcomes such as final diagnoses, symptom resolution, functional status, and satisfaction with care after discharge.[49–52] Even more problematic is the observation that much of residents' time is spent in nonclinical or noneducational tasks.[53,54] Second, although the traditional physician-centric teaching model (attending, resident, intern, and student teams) predominates, the care model of hospitalized patients is now interdisciplinary with patients receiving care from teams consisting of physicians, nurses, dieticians, case managers, and others.[46,55,56]

Third, new work hour rules, while appropriate given the large body of evidence of the pernicious effects of fatigue, have greatly challenged the ability of residency programs to meet expected service needs.[57–64] Programs have coped with the need to provide continuous patient coverage (in most teaching hospitals the residents are the only internists in house overnight) in the setting of work hours restrictions by instituting night and/or day float services. The handoffs necessitated by these float services may be associated with higher rates of preventable adverse events.[65,66] Other solutions include the hiring of hospitalists and physician extenders.[67–70] We know little about the effects of these changes on patient care and resident education,[70–72] although early reports regarding teaching by hospitalists are encouraging.[68–71,73,74] Finally, recent research on medical errors suggests that there is a need for new approaches to supervision, evaluation, and teaching in the inpatient setting. Studies have documented substantial rates of clinical errors, many of which have immediate implications for patient care, committed by trainees in teaching hospitals for over 30 years.[75–84] Data have shown that resident findings and presentations are often at odds with the findings of more experienced attendings.[85,86] A recent systematic review suggested that better supervision was associated with better quality of care.[86] Despite this knowledge, the type and quality of supervision and evaluation by faculty has changed little. Research has shown that learners value direct observation, bedside teaching, and role modeling of clinical skills by faculty.[87–93] These activities are also well received by patients.

THE RESIDENCY CURRICULUM

The Federated Council on Internal Medicine curriculum helped to define the breadth of internal medicine,[94] but the

curriculum has not achieved widespread adoption and programs may not be able to provide enough experiences to cover it.[95] Furthermore, as the body of biomedical knowledge expands, curricula will need to adapt. Internal Medicine needs to more clearly define the *core* content of the knowledge, skills, and attitudes that is required for all internists, regardless of their eventual career tracks. Programs must then identify how best to teach this content in the context of their institutional setting. Surprisingly little study has been performed on what aspects of the inpatient experience are most important for implementation of a successful residency curriculum and optimal patient care.[72] Finally, Internal Medicine must define the level of competency expected for each content area. It may no longer be feasible for residents to acquire "mastery" in all aspects of internal medicine, if indeed that ever was achieved.[96-99]

TRAINING TO REDUCE HEALTH DISPARITIES

The 2000 census confirms the increasing ethnic diversity of inhabitants living in the United States, and ethnic minorities lag behind white Americans on nearly every health care indicator.[18] Residency programs care for a large proportion of patients from socioeconomically disadvantaged populations. Academic Health Centers (AHC) have assumed increasing responsibility for care of the underserved—between 1991 and 1996, AHC expenditures on uninsured patients rose 40%, and at a cost of nearly 4 billion dollars in 1999.[7] Given that one of the core reasons for public funding of graduate medical education (GME) is the social contract between residency programs and the care of underserved patients, a core tenet of patient-centered training should be to educate residents on how to address health disparities in order to reduce or eliminate these gaps.[47,100] All internists, regardless of their ultimate specialty, will encounter health disparities throughout their careers. In addition, lessons learned in addressing health disparities, such as skills from the social sciences, apply across all patient groups.[100-104]

Unfortunately, little data exist regarding the evaluation of curriculum in cultural competency and health disparities.[103,104] A recent systematic review concluded that educational interventions in cultural competence do produce changes in learner knowledge and perhaps some skills, but no study has examined important patient outcomes.[104] Some critics argue that training solely focused on cultural competency training may actually have a deleterious impact on patient care.[105] However, training in health disparities and the specific social science domains of physician roles and behavior, social and cultural issues in health care, and health policy and economics should be incorporated in residency training and involve actual patient care.[100] Compounding the problem is the insufficient number of minority faculty and medical students at academic medical centers and residency programs.[106]

TRAINING FOR LIFE-LONG LEARNING

With the rapid production of new medical information, few would challenge Osler's assertion that medical education is "a life course, for which the work of a few years under teachers is but a preparation." Yet many physicians fail to meet their emerging information needs,[107] witness their up-to-date medical knowledge deteriorate over the years after their training,[108] and, ultimately, demonstrate wide practice variations for procedures with established efficacy.[109]

Self-directed learning (SDL) represents any study form in which individuals have primary responsibility for planning, implementing, and evaluating the effort.[110] Self-directed learners perform a sequence of tasks, including recognizing intrinsic information needs, seeking appropriate information, appraising the information, and applying the information to the triggering scenario. However, they will not engage in the process without sufficient motivation and will not construct personal meaning or sustain the process without reflecting on the learning process itself (metacognition).[111,112] In medicine, empiric studies demonstrate that physicians engage in SDL episodes in response to problems, which may be specific (an uncertainty arising from an encounter with a particular patient) or general (a need for an update in a skill or body of knowledge).[113]

Residency programs currently dedicate relatively little explicit curricular time to SDL, often in the forms of journal clubs, evidence-based medicine (EBM) curricula, and problem-based learning (PBL) conferences, all of which include training in some or all of SDL skills.[114,115] Studies involving journal clubs suffer from weak designs and outcome measures and have shown limited effectiveness.[116] Taking a lesson from undergraduate medical curricula, some programs replaced some of their didactic conferences with PBL sessions. In a controlled trial, pediatrics residents participating in PBL conferences exhibited more frequent SDL behaviors than controls, but the differences did not persist after the intervention ended.[117] In a surgery program, attendance at a basic science PBL conference correlated positively with in-training-examination scores.[118] Notably, this stands in contrast with an internal medicine program study that found no relationship between attendance at a traditional "noon conference" and ABIM certifying examination scores.[119]

From 1998[120] to 2003,[121] the number of programs offering EBM curricula increased from 37% to 71%. Several pre-post controlled studies with objective outcomes have demonstrated the effectiveness of these curricula in improving EBM knowledge and skills.[122-124] Studies of the impact on behaviors suffer from outcome measures lacking validation, such as retrospective self-reports or the frequency of their EBM "utterances" in audiotaped teaching interactions.

In addition to specific courses, we should consider residents' day-to-day experience with SDL on their clinical rotations. Too often programs and faculty fail to utilize the experiential learning through the integration and application of clinical skills, judgment, and EBM. Residents fail to take advantage of SDL opportunities due in part to barriers such as insufficient time, underdeveloped skills, limited access to resources, dysfunctional team dynamics, and an unsupportive institutional culture.[121-125] More recently, educators are exemplifying EBM in "real time" as part of the flow of clinical work, which confronts the logistical constraints faced by busy practitioners and leverages the imperative of immediate learning needs.[126-129] Reform measures should help residents to capitalize on the clinical questions that arise in the care of their patients.

GME FINANCING

The biggest contributor to GME financing is the Center for Medicare and Medicaid Services (CMS). States, the Veterans

Administration, and the Department of Defense provide the bulk of the rest of GME funding that totals nearly 15 billion dollars. Center for Medicare and Medicaid Services uses a complicated formula initially developed in 1982 to determine support for teaching hospitals based on the hospital's Medicare burden. However, payments go to the hospital and not directly to the residency program. Private insurances do not contribute direct financial support to GME and no longer support reimbursement at a higher rate for teaching hospitals.

There is consensus that the current system is flawed and that reform is needed. Multiple organizations or groups have proposed financing reform.[101,130–135] However, there is no consensus on what reforms should be enacted and the result is policy inertia. Policy experts argue that financing GME is a collateral duty of CMS, and if CMS is to continue to fund at least some portion of GME, then residency programs need to demonstrate their public good function more effectively.[101,136–138] The majority of recommendations include changing to an all payer system, distributing GME funds directly to the residency programs, and reducing the variation in GME payments among geographic regions.

Future reform efforts will require the involvement of a broader cross-section of all specialties. We concur with the recommendations to move to an all payer system and to distribute GME funds more directly to programs, but hold the programs accountable for how the funds are utilized. Dedicated funds for educational research are also desperately needed.

SUMMARY OF TASK FORCE RECOMMENDATIONS

The task force offers the following recommendations for reforming internal medicine residency training. In sum, the task force believes we can no longer make changes at the margins of current program structures and that we should embrace bold and innovative reforms for the good of our patients and all trainees. This will require a combination of courage and innovation from all stakeholders.

1. *Residency programs must teach patient-centered care by providing the highest quality of care possible during the training process.*

Broadly defined, patient-centered residency education means providing developmentally appropriate training experiences integrated with interdisciplinary teams to provide coordinated, comprehensive, safe, and high-quality care that simultaneously meets the needs of residents and patients.

2. *There must be better balance between educational experiences in the ambulatory and hospital settings.*

Regardless of their ultimate career choice, all residents need a minimal level of competence in continuing care that focuses on relationships with patients and the community. Residents pursuing careers in specialties that are predominantly outpatient based will require greater time spent in ambulatory education settings.

3. *To honor the social contract and embrace our professional obligations, programs should explicitly address health disparities and incorporate teaching in the social sciences.*

To solve the problems of disadvantaged patients, institutions that sponsor residency programs will need to develop programs to address health disparities and residency programs will need to develop curricula so that residents can effectively participate in these programs. Attention to health disparities will strengthen the bond between the public and residency education.

4. *All stakeholders should work together to better define the "core" knowledge, skills, and attitudes of internal medicine training.*

This core curriculum should prepare all internal medicine residents for any career pathway they choose and include more attention to clinical skills. This work should be a collaborative effort among the programs, medical specialty societies, and the certification boards.

5. *SGIM and other stake-holders should work with regulatory organizations to permit greater flexibility to promote innovative approaches to training.*

Residents now have a broad array of career choices, including urban and rural primary care practice, hospitalist practice, subspecialty training for practice or academics, public health, academic general internal medicine, and others. A "one size fits all" approach to internal medicine training is no longer logical in this new environment.[139–144] The RRC's educational innovations project provides one opportunity to create more flexible, state-of-the-art programs.[139] SGIM should partner with other internal medicine organizations to discuss different approaches to certification, pathways to subspecialization, and new pathways to certification in hospitalism, geriatrics, and other areas of special concentration.

6. *Medical educators must improve the substantial and widespread inadequacies in the current evaluation practices by faculty and programs.*

Trainees must not be advanced to the next level of training without clear evidence they are ready.[145–149] A substantial number of tools and methods currently exist for effective evaluation. This is not an unfunded mandate; the public contributes billions of dollars a year to GME and educators have a moral and ethical responsibility to ensure the competence of graduating trainees.

7. *Clinical work and educational processes in teaching hospitals and clinics need substantial redesign.*

Residents should become learning members of high-functioning interdisciplinary teams without absolute reliance on the resident workforce to compensate for failures in the institution's infrastructure. Furthermore, evidence is accumulating quickly that team-based care leads to better patient outcomes. A growing number of programs have successfully incorporated residents in both inpatient and outpatient interdisciplinary teams.

8. *Internal medicine resident education must develop a more robust faculty supervision system.*

Patients have the right to expect safe and effective care in a training setting. Graduated responsibility and autonomy for clinical decision making will remain central to the educational experience. Residents must learn clinical reasoning skills through the actual care of patients and appropriate faculty supervision must be assured to avoid clinical errors too common in the current educational process. Faculty supervising residents in the inpatient arena need to have protected time to supervise to ensure quality of care.

In the outpatient setting, patients should have a "primary" ambulatory attending and health care team to improve

continuity and create a long-term healing relationship. New faculty-resident co-management strategies will be needed to ensure quality of care, patient safety, and resident learning. To accomplish these goals, faculty development will be needed in new supervision and observation skills, principles of microsystems and safety, chronic disease management, and quality improvement.

9. *SGIM should partner with others to approach CMS and policy makers to change how GME funds are allocated*

Without reform of the financial support for GME any large-scale reform in residency training will be more difficult. The push by CMS for quality and accountability among practitioners is an opportunity for residency programs to step up to the plate. Programs can leverage residency reform as part of a genuine effort to improve patient care to support a change in how funds are allocated. This will require a cohesive effort among organizations to work with CMS and policy makers.

10. *Residency programs must explicitly prepare residents for life-long learning.*

Training in life-long SDL deserves much more explicit emphasis in internal medicine residency training. Residents should be actively involved in answering clinical questions in "real time" and should work with clinical performance data to improve the systems of care in which they work. These skills will lay the groundwork for life-long SDL and improvement. Regardless of the length of training, no resident will attain mastery in all areas of internal medicine, and substantial learning will and must occur throughout a career. In terms of learning infrastructure, residents should have rapid, reliable, and continuously available access to electronic medical information resources at the point of care in every clinical setting. Internet-based portfolios, validated instruments for SDL knowledge and skills, and a SDL readiness scale all show promise as effective tools to improve SDL behaviors and evaluation.[150–152]

11. *Residency reform must also occur in the context of reforms in undergraduate and continuing medical education.*

Residency educators should work with medical student and fellowship educators, and continuing medical education organizations to define benchmarks of competence and coordinate training from undergraduate through graduate and post-GME. The current state of fragmentation among internal medicine educational organizations is counterproductive to effective reform.

12. *Redesign of internal medicine training must promote collaboration among residency programs for better education research and sharing of best educational practices.*

Many important questions about the interface of education and patient care need urgent answers. Questions for research include: how should we address health disparities in the context of a residency program? What is the ideal ambulatory training system? How do new models of continuity affect education and patient care? What are the optimal models of team learning in both the inpatient and outpatient settings? What are the actual costs of training a resident? How will new approaches to supervision affect learning? Finally, better research methodology is needed to address past limitations in residency education research, including cluster designs and combined qualitative–quantitative approaches.[153–156]

The opinions reflected within this manuscript are solely those of the authors and do not represent the views of the Department of Defense, Department of the Army, the American Board of Internal Medicine, or the American College of Physicians. Portions of this work were supported by the Society of General Internal Medicine.

REFERENCES

1. **Schroeder SA, Showstack JA, Gerbert B.** Residency training in internal medicine: time for change? Ann Intern Med. 1986;104:554–61.
2. **Greganti MA, Fletcher SW.** Residency training in the inpatient setting: a new dilemma for internal medicine. J Gen Intern Med. 1989;4:136–8.
3. **Fallon HJ.** Residency reform: a perspective from the Association of Professors in Medicine. Ann Intern Med. 1992;116:1041.
4. **Nolan JP, Inui TS.** Tinkering or real reform? The choice is ours. Ann Intern Med. 1992;116:1042–5.
5. **Cantor JC, Baker LC, Hughes RG.** Preparedness for practice: young physicians' views of their professional education. JAMA. 1993;270: 1035–40.
6. **Cohen JJ.** Honoring the "E" in GME. Acad Med. 1999;74:108–13.
7. **Task Force on Academic Health Centers.** Training tomorrow's doctors: the medical education mission of academic health centers. Commonwealth Fund. April 2002.
8. **Whitcomb ME.** Putting patients first: the need to reform graduate medical education. Acad Med. 2003;78:851–2.
9. **Donabedian A.** The Definition of Quality and to its Assessment. Ann Arbor, Mich: Health Administration Press; 1980.
10. **Accreditation Council for Graduate Medical Education.** ACGME Outcome Project: The General Competencies. Accessed at www.acgme.org on April 24, 2004.
11. **American Board of Internal Medicine.** Portfolio for Internal Medicine Residency Programs. Philadelphia: American Board of Internal Medicine; 2001.
12. **Ben-david MF.** AMEE guide no. 14: outcome-based education: part 3—assessment in outcome-based education. Med Teach. 1999;21:23–5.
13. **Harden RM.** AMEE guide no. 14: outcome-based education: part 1: an introduction to outcome-based education. Med Teach. 1999;21:7–14.
14. **Institute of Medicine.** Health Professions Education: A Bridge to Quality. Washington: National Academy Press; 2003.
15. **Ende J, Atkins E.** Conceptualizing curriculum for graduate medical education. Acad Med. 1992;67:528–34.
16. **Choudhry NK, Fletcher RH, Soumerai SB.** Systematic review: the relationship between clinical experience and quality of health care. Ann Intern Med. 2005;142:260–73.
17. **Institute of Medicine.** Crossing the Quality Chasm: A New Health System for the 21st Century. Washington: National Academy Press; 2001.
18. **Institute of Medicine.** Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington: National Academy Press; 2002.
19. **Institute for the Future.** Health and Healthcare 2010: The Forecast, the Challenge. San Francisco: Jossey-Bass; 2003.
20. **Stillman PL, Regan MB, Swanson DB, et al.** An assessment of the clinical skills of fourth year students at four New England medical schools. Acad Med. 1990;65:320–6.
21. **Pfeiffer C, Madray H, Ardolino A, Willms J.** The rise and fall of students' skill in obtaining a medical history. Med Educ. 1998;32: 283–8.
22. **Mangione S, Nieman LZ.** Cardiac auscultatory skills of internal Medicine and Family Practice trainees: a comparison of diagnostic proficiency. JAMA. 1997;278:717–22.
23. **Newton DA, Grayson MS.** Trends in career choice by US medical school graduates. JAMA. 2003;290:1179–82.
24. **Dorsey ER, Jarjoura D, Rutecki GW.** Influence of controllable lifestyle on recent trends in specialty choice by US medical students. JAMA. 2003;290:1173–8.
25. **Jolly P.** Medical school tuition and young physician indebtedness. American Association of Medical Colleges, 2004. Accessed at www.aamc.org August 10, 2004.
26. **Petersdorf RG, Goltein L.** The future of internal medicine. Ann Intern Med. 1993;119:1130–7.
27. **Perkoff GT.** Teaching medicine in the ambulatory setting: an idea whose time may finally have come. N Engl J Med. 1986;314:27–31.

28. **Walter D, Whitcomb ME.** Venues for clinical education in Internal Medicine residency programs and their implications for future training. Am J Med. 1998;105:262–5.
29. **Wiest FC, Ferris TG, Gokhale M, Campbell EG, Weissman JS, Blumenthal D.** Preparedness of internal medicine and family practice residents for treating common conditions. JAMA. 2002;288: 2609–14.
30. **Spickard A III, Ryan SP, Muldowney JA III, Farnham L.** Outpatient morning report: a new conference for internal medicine residency programs. J Gen Intern Med. 2000;15:822–4.
31. **Wenderoth S, Pelzman F, Demopoulos B.** Ambulatory morning report: can it prepare residents for the American Board of Internal Medicine examination? J Gen Intern Med. 2002;17:207–9.
32. **Randall DC, Strong J, Gibbons R.** A longitudinal subspecialty experience for internal medicine residents. Mil Med. 2001;166:40–3.
33. **Hellmann DB, Flynn JA.** Development and evaluation of a coordinated, ambulatory rheumatology experience for internal medicine residents. Arthr Care Res. 1999;12:325–30.
34. **Bharel M, Jain S, Hollander H.** Comprehensive ambulatory medicine training for categorical internal medicine residents. J Gen Intern Med. 2003;18:288–93.
35. **Bowen JL, Irby DM.** Assessing quality and costs of education in the ambulatory setting: a review of the literature. Acad Med. 2002;77: 621–80.
36. **Blumenthal D, Gokhale M, Campbell EG, Weissman JS.** Preparedness for clinical practice: reports of graduating residents at academic health centers. JAMA. 2001;286:1027–34.
37. **Wickstrom GC, Kelley DK, Keyserling TC, et al.** Confidence of graduating internal medicine residents to perform ambulatory procedures. J Gen Intern Med. 2000;15:353–60.
38. **Bernard AM, Anderson L, Cook CB, Phillips LS.** What do internal medicine residents need to enhance their diabetes care? Diabetes Care. 1999;22:661–6.
39. **Boutin-Foster C, Charlson ME.** Problematic resident-patient relationships: the patient's perspective. J Gen Intern Med. 2001;16: 750–4.
40. **Nelson EC, Batalden PB, Huber TP, et al.** Microsystems in health care: part 1. Learning from high performing front line units. Jt Comm J Qual Safety. 2002;28:472–93.
41. **Godfrey MM, Nelson EC, Wasson JH, Mohr JJ, Batalden PB.** Microsystems in health care: part 3. Planning patient centered services. Jt Comm J Qual Safety. 2003;29:159–70.
42. **Wasson JH, Godfrey MM, Nelson EC, Mohr JJ, Batalden PB.** Microsystems in health care: part 4. Planning patient-centered care. Jt Comm J Qual Safety. 2003;29:227–37.
43. **Von Korff M, Gruman J, Schaefer J, Curry SJ, Wagner EH.** Collaborative management of chronic illness. Ann Intern Med. 1997;127:1097–102.
44. **Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A.** Improving chronic illness care: translating evidence into action. Health Aff. 2001;20:64–78.
45. **Wagner EH.** The role of patient care teams in chronic disease management. BMJ. 2000;320:569–72.
46. **Hall P, Weaver L.** Interdisciplinary education and teamwork: a long and winding road. Med Educ. 2001;35:867–75.
47. **Ludmerer KM.** Time to Heal: American Education from the Turn of the Century to the Era of Managed Care. Oxford: Oxford University Press; 1999.
48. **Morelock JA, Stern DT.** Shifting patients: how residency programs respond to Residency Review Committee requirements. Acad Med. 2003;115:48–54.
49. **Wenger NS, Shpiner RB.** An analysis of morning report: implications for internal medicine education. Ann Intern Med. 1993;119:395–9.
50. **Matter CA, Speice JA, McCann R, et al.** Hospital to home: improving internal medicine residents' understanding of the needs of older persons after a hospital stay. Acad Med. 2003;78:793–7.
51. **Wright SM, Durbin P, Barker LR.** When should learning about hospitalized patients end? Providing housestaff with post-discharge follow-up information. Acad Med. 2000;75:380–3.
52. **Diem SJ, Prochazka AV, Meyer TJ, Fryer GE.** Effects of a postdischarge clinic on housestaff satisfaction and utilization of hospital services. J Gen Intern Med. 1996;11:179–81.
53. **Dresselhaus TR, Luck J, Wright BC, Spragg RG, Lee ML, Bozzette SA.** Analyzing the time and value of housestaff inpatient work. J Gen Inten Med. 1998;13:534–40.
54. **Nerenz D, Rosman H, Newcomb C, et al.** The on-call experience of interns in internal medicine. Medical education task force of the Henry Ford Hospital. Arch Intern Med. 1990;150:2294–7.
55. **Cooper H, Carlisle C, Gibbs T, Watkins C.** Developing an evidence base for interdisciplinary learning: a systematic review. J Adv Nurs. 2001;35: 228–37.
56. **Moore SM, Alemi F, Headrick LA, et al.** Interdisciplinary learning in continuous improvement of healthcare: four perspectives. Jt Comm J Qual Improvement. 1996;22:165–87.
57. **Accreditation Council for Graduate Medical Education.** General program requirements; Residency Review Committee. Accessed at www.acgme.org on April 24, 2004.
58. **Schroeder SA.** How many hours is enough? An old profession meets a new generation. Ann Intern Med. 2004;140:838–9.
59. **Skeff KM, Ezeji-Okoye S, Pompei P, Rockson S.** Benefits of resident work hours regulation. Ann Intern Med. 2004;140:816–7.
60. **Charap M.** Reducing resident work hours: unproven assumptions and unforeseen outcomes. Ann Intern Med. 2004;140:814–5.
61. **Shanafelt TD, Bradley KA, Wipf JE, Back AL.** Burnout and self-reported patient care in an internal medicine residency program. Ann Intern Med. 2002;136:358–67.
62. **Lockley SW, Cronin JW, Evans EE, et al.** Effects of reducing interns' weekly work hours on sleep and attentional failures. N Engl J Med. 2004;351:1829–37.
63. **Landrigan CP, Rothschild JM, Cronin JW, et al.** Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med. 2004;351:1838–48.
64. **Barger LK, Cade BE, Ayas NT, et al.** N Engl J Med. 2005;352: 125–34.
65. **Petersen LA, Brennan TA, O'Neil AC, Cook EF, Lee TH.** Does housestaff discontinuity of care increase the risk for preventable adverse events? Ann Intern Med. 1994;121:866–72.
66. **Patterson ES, Roth EM, Woods DD, Chow R, Gomes JO.** Handoff strategies in settings with high consequences for failure: lessons for healthcare operations. Int J Qual Healthcare. 2004;16:125–32.
67. **Wong J, Holmboe ES, Huot S.** A novel dayfloat rotation to address the 80 hour work restriction. J Gen Intern Med. 2004;19(part 2):519–23.
68. **Kripalani S, pope AC, Rask K, et al.** Hospitalists as teachers: how do they compare to subspecialty and general medicine faculty? J Gen Intern Med. 2004;19:8–15.
69. **Kulaga ME, Charney P, O'Mahony SP, et al.** The positive impact of initiation of hospitalist clinician educators. J Gen Intern Med. 2004;19: 293–301.
70. **Landrigan CP, Muret-Wagstaff S, Chiang VW, Nigrin DJ, Goldmann DA, Finkelstein JA.** Effect of a pediatric hospitalist system on housestaff education and experience. Archiv Pediatr Adolescent Med. 2002;156:877–83.
71. **Nishimura RA, Linderbaum JA, Naessens JM, Spurrier B, Koch MB, Gaines KA.** A nonresident cardiovascular inpatient service improves residents' experience in an academic medical center: a new model to meet the challenges of the new millennium. Acad Med. 2004;79: 426–31.
72. **Griffith CH III, Rich EC, Hillson SD, Wilson JF.** Internal medicine residency training and outcomes. J Gen Intern Med. 1997;12:390–6.
73. **Tenner PA, Dibrell H, Taylor RP.** Improved survival with hospitalists in a pediatric intensive care unit. Crit Care Med. 2003;31:847–52.
74. **Chung P, Morrison J, Jin L, Levinson W, Humphrey H, Meltzer D.** Resident satisfaction on an academic hospitalist service: time to teach. Am J Med. 2002;112:597–601.
75. **Chaudhry SI, Olofinboda KA, Krumholz HM.** Detection of errors by attending physicians on a general medicine service. J Gen Intern Med. 2003;18:595–600.
76. **Fox RA, Clark CLI, Scotland AD, Dacre JE.** A study of pre-registration house officers' clinical skills. Med Educ. 2000;34:1007–12.
77. **Weiner S, Nathanson M.** Physical examination. Frequently observed errors. JAMA. 1976;236:852–5.
78. **Wray NP, Friedland JA.** Detection and correction of house staff error in physical diagnosis. JAMA. 1983;249:1035–7.
79. **Mangione S, Burdick WP, Peitzman SJ.** Physical diagnosis skills of physicians in training: a focused assessment. Acad Emerg Med. 1995; 2:622–9.
80. **Li JTC.** Assessment of basic examination skills of internal medicine residents. Acad Med. 1994;69:296–9.
81. **Johnson JE, Carpenter JL.** Medical house staff performance in physical examination. Arch Intern Med. 1986;146:937–41.

82. **Battles JB, Shea CE.** A system of analyzing medical errors to improve GME curricula and programs. Acad Med. 2001;76:125–33.
83. **Reilly BM.** Physical examination in the care of medical inpatients: an observational study. Lancet. 2003;362:1100–5.
84. **Holliman CJ, Wuerz RC, Kimak MJ, et al.** Attending supervision of nonemergency medicine residents in a university hospital ED. Am J Emerg Med. 1995;13:259–61.
85. **Gennis VM, Gennis MA.** Supervision in the outpatient clinic: effects on teaching and patient care. J Gen Intern Med. 1993;9:116–20.
86. **Kilminster SM, Jolly BC.** Effective supervision in clinical practice settings: a literature review. Med Educ. 2000;34:827–40.
87. **Phy MP, Offord KP, Manning DM, Bundrick JB, Huddleston JM.** Increased faculty presence on inpatient teaching services. Mayo Clin Proceed. 2004;79:332–6.
88. **Norcini JJ, Blank LL, Arnold GK, Kimball HR.** The mini-CEX (clinical evaluation exercise): a preliminary investigation. Ann Intern Med. 1995;123:795–9.
89. **Norcini JJ, Blank LL, Duffy FD, Fortna GS.** The mini-CEX: a method for assessing clinical skills. Ann Intern Med. 2003;138:476–81.
90. **Irby DM.** Three exemplary models of case-based teaching. Acad Med. 1994;69:947–52.
91. **Grant J, Kilminster S, Jolly B, Cottrell D.** Clinical supervision of SpRs: where does it happen, when does it happen and is it effective? Specialist registrars. Med Educ. 2003;37:140–8.
92. **Lehmann LS, Brancati FL, Chen M-C, Roter D, Dobs AS.** The effect of bedside case presentations on patients' perceptions of their medical care. N Engl J Med. 1997;336:1150–6.
93. **Wright SM, Kern DE, Kolodner K, Howard DM, Brancati FL.** Attributes of excellent attending-physician role models. N Engl J Med. 1998;339:
1986–93.
94. **Federated Council of Internal Medicine.** Curriculum for Internal Medicine. Accessed at acponline.org on April 24, 2004.
95. **Hripcsak G, Stetson PD, Gordon PG.** Using the Federated Council of Internal Medicine curricular guide and administrative codes to assess IM residents' breadth of experience. Acad Med. 2004;79:557–63.
96. **Dreyfus HL.** On the Internet: Thinking in Action. New York: Routledge; 2001.
97. **Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S.** General competencies and accreditation in graduate medical education. Health Aff. 2002;21:103–11.
98. **Ogrinc G, Headrick LA, Mutha S, Coleman MT, O'Donnell J, Miles PV.** A framework for teaching medical students and residents about practice-based learning and improvement, synthesized from a literature review. Acad Med. 2003;78:748–56.
99. **Mallett S, Clarke M.** How many reviews are needed to cover existing evidence on the effects of health care interventions? ACP J Club. 2003;139:A11.
100. **Institute of Medicine.** Improving Medical Education: Enhancing the Behavioral and Social Science Content of Medical School Curricula. Washington: National Academy Press; 2004.
101. **Council on Graduate Medical Education (COGME).** Proceedings of the GME Financing Stakeholders Meeting. Public Response to COGME's Fifteenth Report, Department of Health and Human Services, Rockville, Md, September, 2001.
102. **Institute of Medicine.** Academic Health Centers. Leading Change in the 21st Century. Washington: National Academy Press; 2004.
103. **Price EG, Beach MC, Gary T, et al.** Quality of educational literature on cultural competence training of health professionals. J Gen Intern Med Vol. 2004;19(suppl 1):200.
104. **Beach MC, Price EG, Gary TL, et al.** Cultural competence. A systematic review of health care provider educational interventions. Med Care. 2005;43:356–73.
105. **Wardlow H.** Giving birth to Gonolia: "culture" and sexually transmitted disease among the Huli of Papua New Guinea. Medl Anthropol Quart. 2002;16:151–75.
106. **Lypson ML, Gruppen L, Stern DT.** Warning signs of declining faculty diversity. Acad Med. 2002;77:S10–2.
107. **Covell DG, Uman GC, Manning PR.** Information needs in office practice: are they being met? Ann Intern Med. 1985;103:596–9.
108. **Ramsey PG, Carline JD, Inui TS, et al.** Changes over time in the knowledge base of practicing internists. JAMA. 1991;266:1103–7.
109. **McGlynn EA, Asch SM, Adams J, et al.** The quality of health care delivered to adults in the United States. N Engl J Med. 2003;348:2635–45.
110. **Brockett RG, Hiemstra R.** Self-Direction in Learning: Perspectives in Theory, Research, and Practice. London: Routledge; 1991.
111. **Garrison DR.** Self-directed learning: toward a comprehensive model. Adult Educ Quart. 1997;48:18–33.
112. **Candy PC.** Self-Directed Lifelong Learning: A Comprehensive Guide to Theory and Practice. San Francisco: Jossey & Bass; 1991.
113. **Slotnick HB.** How doctors learn: physicians' self-directed learning episodes. Acad Med. 1999;74:1106–17.
114. **Coomarasamy A, Taylor R, Khan KS.** A systematic review of postgraduate teaching in evidence-based medicine and critical appraisal. Med Teach. 2003;25:77–81.
115. **Alguire PC.** A review of journal clubs in postgraduate medical education. J Gen Intern Med. 1998;13:347–53.
116. **Green ML.** Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. Acad Med. 1999;74:686–94.
117. **Ozuah PO, Stein RE.** More on problem-based learning and self-directed learning [comment]. Arch Pediatr Adolesc Med. 2001;155:1278.
118. **Itani KM, Miller CC, Church HM, McCollum CH.** Impact of a problem-based learning conference on surgery residents' in training exam (ABSITE) scores. American Board of Surgery in Training Exam. J Surg Res. 1997;70:66–8.
119. **FitzGerald JD, Wenger NS.** Didactic teaching conferences for IM resident: who attends, and is attendance related to medical certifying examination scores? Acad Med. 2003;78:84–9.
120. **Green ML.** Evidence-based medicine training in internal medicine residency programs a national survey. J Gen Intern Med. 2000;15:
129–33.
121. **Dellavalle RP, Stegner DL, Deas AM, et al.** Assessing evidence-based dermatology and evidence-based internal medicine curricula in US residency training programs: a national survey. Arch Dermatol. 2003;
139:369–72; discussion 372.
122. **Green ML, Ellis PJ.** Impact of an evidence-based medicine curriculum based on adult learning theory. J Gen Intern Med. 1997;12:742–50.
123. **Smith CA, Ganschow PS, Reilly BM, et al.** Teaching residents evidence-based medicine skills: a controlled trial of effectiveness and assessment of durability. J Gen Intern Med. 2000;15:710–5.
124. **Ross R, Verdieck A.** Introducing an evidence-based medicine curriculum into a family practice residency—is it effective? Acad Med. 2003;78:412–7.
125. **Green ML.** Evidence-based medicine training in graduate medical education: past, present and future. J Eval Clin Pract. 2000;6:121–38.
126. **Richardson WS.** One slice or the whole pie? Evidence-Based Health Care Newslett. 2001;21:17–18.
127. **Bradley DR, Rana GK, Martin PW, Schumacher RE.** Real-time, evidence-based medicine instruction: a randomized controlled trial in a neonatal intensive care unit. J Med Library Assoc. 2002;90:
194–201.
128. **Green ML, Ciampi MA, Ellis PJ.** Residents' medical information needs in clinic: are they being met? Am J Med. 2000;109:218–23.
129. **Dinkevich E, Ozuah PO.** Self-directed learning activities of paediatric residents. Med Educ. 2003;37:388–9.
130. **Council on Graduate Medical Education (COGME).** Fifteenth Report: Financing Graduate Medical Education in a Changing Health Care Environment, Department of Health and Human Services, Rockville, Md, December 2000.
131. **American College of Physicians.** The physician workforce and financing graduate medical education. Ann Intern Med. 1998;128:142–8.
132. **Sundwall DN.** Another alternative for financing graduate medical education. A proposal from the Council on Graduate Medical Education. Health Aff. 2001;20:156–8.
133. **Fryer GE, Green LA, Dovey S, Phillips RL.** Direct graduate medical education payments to teaching hospitals by Medicare: unexplained variation and public policy contradictions. Acad Med. 2001;76:439–45.
134. **Covey AS, Freidlaender GE.** Financing graduate medical education: sorting out the confusion. J Bone Jt Surg. 2003;85:1594–1604.
135. **Rich EC, Liebow M, Srinivasan M, et al.** Medicare financing of graduate medical education. Intractable problems, elusive solutions. J Gen Intern Med. 2002;17:283–92.
136. **Guamano M, Schlesinger M.** The social roles of Medicare: assessing Medicare collateral benefits. J Health Politics, Policy, Law. 2001;26:
37–79.
137. **Council on Graduate Medical Education (COGME).** Collaborative Education to Ensure Patient Safety, Department of Health and Human Services, Rockville, Md, 2000.

138. **Liaison Committee on Medical Education.** Functions and Structure of a Medical School: Standards for Accreditation of Medical Education Programs Leading to the MD Degree, Washington, DC, September 2003.
139. **Gorroll AH, Sirio C, Duffy FD, et al.** A new model for accreditation of residency programs in internal medicine. Ann Intern Med. 2004;140: 902–9.
140. **Duffy FD, Zipes DP.** The future of certification and recertification. Am J Med. 2004;117:140–4.
141. **Bowen JL.** Adapting residency training. Training adaptable residents. West J Med. 1998;168:371–7.
142. **Cleghorn CD, Headrick LA.** The PDSA cycle at the core of learning in health professions education. Jt Comm J Qual Improv. 1996;22: 206–12.
143. **Humphries HJ.** Customizing residency education. Ann Intern Med. 2004;140:663–4.
144. **Goldman L.** Modernizing the paths to certification in Internal Medicine and its subspecialties. Am J Med. 2004;117:133–6.
145. **Miller G.** The assessment of clinical skills/competence/performance. Acad Med. 1990;65:S63–7.
146. **Holmboe ES.** The importance of faculty observation of trainees' clinical skills. Acad Med. 2004;79:16–22.
147. **Long DM.** Competency-based residency training: the next advance in graduate medical education. Acad Med. 2000;75:1178–83.
148. **Lypson ML, Frohna JG, Gruppen LD, Woolliscroft JO.** Assessing residents' competencies at baseline: identifying the gaps. Acad Med. 2004;79:564–70.
149. **Silber CG, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski JJ.** Do global rating forms enable program directors to assess the ACGME competencies? Acad Med. 2004;79:549–56.
150. **Ramos KD, Schafer S, Tracz SM.** Validation of the Fresno test of competence in evidence based medicine. BMJ. 2003;326:319–21.
151. **Fung MF, Walker M, Fung KF, et al.** An internet-based learning portfolio in resident education: the KOALA multicentre programme. Med Educ. 2000;34:474–9.
152. **Guglielmino LM.** Development of the self-directed learning readiness scale (doctoral dissertation, University of Georgia). Dissertation Abstracts International. 1997;38:6467A.
153. **Shea JA, Arnold L, Mann KV.** A RIME perspective on the quality and relevance of current and future medical education research. Acad Med. 2004;79:931–8.
154. **Dauphinee WD, Wood-Dauphinee S.** The need for evidence in medical education: the development of best evidence medical education as an opportunity to inform, guide, and sustain medical education research. Acad Med. 2004;79:925–30.
155. **Chen FM, Baucher H, Burstin H.** A call for outcomes research in medical education. Acad Med. 2004;79:955–60.
156. **Wartman SA.** Revisiting the idea of a national center for health professions education research. Acad Med. 2004;79:910–7.

Appendix 1

SGIM Reforming Residency Task Force

Core Task Force Members:
Eric S. Holmboe, MD (Yale University)
Catherine Lucey, MD (Ohio State University)
Judith L. Bowen, MD (Oregon Health & Science University)
Daniel Duffy, MD (represented ABIM)
Patrick Alguire, MD (represented ACP)
Jessica Gregg, PhD, MD (Tulane University)
Lorenzo DiFrancesco, MD (Emory University, represented SHM)
David Battinelli, MD (Boston University, represented APDIM)
Eileen Reynolds, MD (Harvard University, represented SGIM council)
Michael Green, MD, MSc (Yale University)
Subcommittees:
Ambulatory Medicine
Chair: Judith L. Bowen, MD (Oregon Health & Science University)
Suzanne Brandenburg, MD (University of Colorado)
John Chamberlain, MD (Practicing internist, Rochester, NY)
Helen Chen, MD (University of California at San Francisco)
Elizabeth Eckstrom, MD, MPH (Legacy Health System, Portland, Ore)
Steve Salerno, MD, MPH (Tripler Army Hospital, Hawaii)
Inpatient Medicine
Chair: Lorenzo Di Francesco, MD (Emory University)
Andrew Auerbach, MD (University of California at San Francisco)
Robert Nardino, MD (Hospital of Saint Raphaels, Connecticut)
Michael Pistoria, DO (Lehigh Valley Hospital)
Cultural Competence and Health Disparities
Chair: Jessica Gregg, PhD, MD (Tulane)
Jada Bussey-Jones, MD (Emory University)
Leonor Fernandez, MD (Harvard University)
Maurice Lemon, MD (Rush Medical College)
Mukta Panda, MD (University of Tennessee at Chattenooga)
Joe Ravenell, MD (Cornell Medical College)
William Salazar MD (Medical College of Georgia)
Life-Long Learning
Chair: Michael Green, MD, MSc (Yale University)
Michael Zaroukian, MD, PhD (Michigan State College of Human Medicine)
Sharon Strauss, MD (University of Toronto)
Mark Wilson, MD (University of Iowa)
Hank Slotnick, PhD (University of Wisconsin, physician learning specialist)

Supplementary Material

The following supplementary material is available for this article online at www.blackwell-synergy.com:

Appendix 2. Guiding Principles for Task Force Report.

Prescription drug use and abuse

Risk factors, red flags, and prevention strategies

J. Harry Isaacson, MD John A. Hopper, MD Daniel P. Alford, MD, MPH Ted Parran, MD

PREVIEW

When a patient is in physical or emotional pain, prescribing controlled substances often appears to be the simplest and most efficient way to relieve suffering and distress. However, in a minority of cases, this approach leads to prescription drug abuse and patient harm. In this article, the authors review the epidemiologic factors of prescription drug abuse, legal policies designed to safeguard against it, risk factors and red flags, and practical ways to minimize the chances of misuse.

Identification of prescription drug abuse is admittedly imprecise, in part because of variable physician and societal beliefs about what constitutes appropriate use of medication. Table 1 outlines a practical approach to defining prescription drug abuse and expands on terminology originally described by Finch.[1]

Certain patterns of drug use identify greater risk of harm than benefit. These patterns include escalating use of a substance without consultation with a physician, use of a substance for effects independent of a defined medical condition (ie, using medication to "get high"), continued use of a substance despite negative consequences, and preoccupation with obtaining the substance.

Physical dependence does not equal abuse

Physical dependence refers to the pharmacologic principle that abrupt cessation of intake of some substances can lead to characteristic withdrawal symptoms. *Tolerance* refers to the state in which escalating doses of a substance must be ingested to attain the same effect. A patient treated on a long-term basis with a controlled substance may exhibit physical dependence or tolerance (biologic phenomena) to the medication without any misuse or abuse (behavioral phenomena). For example, a patient with metastatic cancer taking high doses of narcotics for pain relief becomes physically dependent and experiences acute opioid withdrawal if the medication is abruptly stopped. In these situations, the patient must be educated about the drug's benefits and about the difference between physical dependence and drug abuse.

In contrast, some patients may meet criteria for prescription drug abuse without having physical dependence on a substance. This situation may occur when the use of a substance is intermittent yet still significantly interferes with function. A substance can also be ingested for reasons other than to treat a defined medical condition without being used

continued

The authors disclose no financial interests in this article.

Table 1. Types of use and misuse of prescription drugs

Term	Definition	Clinical example	Intervention strategy
Appropriate use	Use of controlled substance as prescribed for defined condition with no signs of misuse or abuse	10-day course of postoperative narcotics taken as prescribed	Explain in advance to patient that narcotics will be used for limited time only
Misuse/ inappropriate use	Use of controlled substance for reason other than that for which it was prescribed or in dosage different than that prescribed No pattern of misuse leading to disability or dysfunction	Single episode of narcotic used twice as often as prescribed Use of old prescription for new clinical problem without consulting physician	Educate patient on proper use of medication
Abuse	Use of controlled substance outside normally accepted standards of use, resulting in disability and/or dysfunction	Continued misuse despite interventions Use of narcotic for recreational purposes unrelated to medical condition	Express concern in empathetic manner Discontinue medication of abuse Consult with expert (chemical dependency, pain management)
Catastrophic use	Use of controlled substance that involves illegal activity or places patient in immediate harm	Altering prescription or selling controlled substance Overdose	Immediately stop prescribing controlled substances Consult chemical dependency expert Notify legal authorities if indicated

continuously or at a sufficient dose for physical dependence. Finally, it is important to realize that patients may be denied appropriate use of a controlled substance for a variety of reasons and be mislabeled as having a drug problem when they do not.

Epidemiologic factors of drug abuse

The lack of a universally agreed-upon definition of prescription drug abuse has led to limitations in epidemiologic research. Although many classes of drugs are subject to misuse, physicians are most often concerned with intentional abuse of prescription controlled substances. Prescription stimulants, sedatives, tranquilizers, and analgesics can all be misused.

Prescription analgesics are the most widely abused of the prescription psychoactive medications, followed by tranquilizers, stimulants, and sedatives. The National Household Survey on Drug Abuse (NHSDA) provides insight into the scope of psychoactive

medication misuse in the United States. According to data from 2003, an estimated 6.3 million persons aged 12 or older (2.7% of the population) had used prescription psychotherapeutic medications nonmedically in the month before being surveyed. Pain relievers were most commonly used (4.7 million), followed by tranquilizers (1.8 million), stimulants (1.2 million), and sedatives (0.3 million). More than 9% of high school seniors reported using OxyContin, Vicodin, or Percocet in the past year.

In the United States, women are nearly 50% more likely than men to be prescribed a controlled drug (primarily opiates and anxiolytics).[3] Despite the greater exposure of women to medications of abuse, rates of prescription drug misuse are similar for women and men.[2] Patients with psychiatric disorders are more likely than the general population to have comorbid substance use disorders.[4]

From 1990 to 1998, the number of new nonmedical users of opioid analgesics increased by 181%, or about 1.5 million people, according to the NHSDA.[5] Emergency department visits related to the use of oxycodone hydrochloride increased 452%, from 4,069 reports in 1994 to 18,409 in 2001.[6] Nonmedical use increased 132% for tranquilizers, 90% for sedatives, and 165% for stimulants during this same time.[5] This rise has led the National Institute on Drug Abuse to launch a campaign against prescription drug abuse.[7]

A recent study of patients using opioid medications for noncancer pain in primary care settings found prescription drug abuse in up to 31% of patients. A lifetime history of substance use and younger age were significantly associated with prescription drug abuse.[8]

Legal issues

Physicians are legally responsible when prescribing controlled drugs and need to be familiar with federal and state prescribing laws and regulations. In all cases, the most stringent law—whether it is federal or state—takes precedence. Controlled substances are considered drugs with potential for physical dependence, psychologic dependence, or both. Under the Food, Drug, and Cosmetics Act, the US Food and Drug Administration (FDA) approves drugs established to be safe and effective. Physicians may choose to prescribe drugs for indications other than their approved labeling if it is done on the basis of acceptable medical practice. The FDA's role does not include regulating medical practice.[9]

The Controlled Substances Act of 1970 is the legal foundation by which the federal government regulates the use and availability of controlled drugs. It places all controlled drugs into one of five scheduled classes (table 2). The schedules are based on the degree of medicinal value, potential for abuse, and relative safety of the drug.[9] Each schedule carries different penalties for unlawful use. Schedule I contains drugs that have no acceptable medical use and are illegal except for approved research use. Schedules II, III, IV, and V have accepted medical uses and decreasing potential for abuse.

Schedule II drugs must be prescribed with a written prescription for no more than 30 days, and refills are prohibited unless the patient contacts his or her physician and obtains a new prescription. However, in cases of medical emergencies, written prescriptions for limited amounts of Schedule II medications may be telephoned or faxed to a pharmacy, with the written prescriptions delivered within 72 hours. As many as five refills are allowed with Schedule III, IV, and V medications.

The US Department of Justice's Drug Enforcement Agency was established in 1973 to enforce controlled-substance laws and regulations, including the investigation of cases of diversion. Physicians need to register with the Department of Justice before they can prescribe controlled substances.

The Federation of State Medical Boards develops guidelines for state boards. One such guideline is the Model Guidelines for the Use of Controlled Substances for the Treatment of Pain,[10] which was developed in

continued

1998 and serves as a framework for the use of opioids in the management of chronic pain. These guidelines improve physician practice and protect legitimate medical use of controlled drugs.

Influence of type of drug

Because prescription medications are approved by the FDA, patients often consider abuse of these substances safer than abuse of illicit, or "street," drugs. The most commonly abused medications, which include opioids, benzodiazepines, barbiturates, and stimulants, have mood-altering effects. The drugs within these classes that are the most abused and have the greatest value for diversion (eg, hydrocodone bitartrate, oxycodone, alprazolam [Xanax]) have rapid onset of action and an intense effect.

Long-acting or sustained-release preparations have historically been less often abused because of their slow onset of action. However, when a sustained-release preparation is crushed, dissolved, and injected, its onset of action becomes rapid and intense. Brand-name medications are more desirable among prescription drug abusers because they are more easily recognized as the "real thing."

Influence of the patient

A number of factors may put a patient at risk or be clues to an increased chance of prescription drug abuse. Patients with a history of alcohol or drug problems are clearly at risk when they are exposed to controlled substances.[8]

As patients become desperate to obtain medica-

Table 2. Schedules of controlled drugs

Schedule		Examples
I	High abuse potential and no accepted medical indications	Heroin, marijuana, lysergic acid diethylamide, mescaline
II	High abuse potential and accepted medical indications with severe restrictions. May lead to severe psychologic or physical dependence	Morphine sulfate, codeine, methadone HCl, oxycodone HCl, hydromorphone HCl (Dilaudid), meperidine HCl (Demerol), fentanyl, cocaine, amphetamines, dronabinol (Marinol), secobarbital sodium (Seconal Sodium Pulvules)
III	Accepted medical uses Abuse potential less than with Schedule II drugs	Compounds with limited quantities of controlled substances (eg, acetaminophen with codeine or hydrocodone bitartrate), buprenorphine HCl (Buprenex, Subutex)
IV	Accepted medical uses Abuse potential less than with Schedule III drugs	Propoxyphene (Darvon), pentazocine (Talwin), phentermine HCl, phenobarbital, benzodiazepines
V	Accepted medical uses Abuse potential less than with Schedule IV drugs	Preparations with limited quantities of certain opioids and stimulants, including many antitussive and antidiarrheal drugs, which are often nonprescription (eg, diphenoxylate HCl with atropine sulfate)

tions, they may engage in a number of scams designed to dupe physicians.[11] Examples include complaining of pain syndromes such as toothache, migraine, or renal colic (in some cases pricking their finger and dropping blood in their urine sample to mimic hematuria). They often present with such complaints to new physicians or emergency departments, or they call their clinic after hours, when they know their primary physician is unavailable. Patients may seek out multiple physicians, claiming to be new to the area. Additional red flags are noted in table 3.

Table 3. Red flags for drug seeking by patients

Are more concerned about the drug than the problem
Report multiple medication sensitivities
Say that they cannot take generic drugs
Refuse diagnostic workup or consultation
Have sophisticated knowledge of drugs
Say "You are the only one who can help me"
Say they have lost prescriptions

Influence of the physician

We live in a culture of "a pill for every ill," which includes the expectation by both patients and physicians that a medication will be prescribed at each visit. As a result, physicians may experience a conflict between wanting to provide symptom relief and needing to set limits.

Smith and Seymore[12] call the physician factors associated with prescription drug abuse the "Four Ds." *Dated* refers to physicians who have not kept up with new medical knowledge and prescribe older medications with higher abuse potential without regard to safe prescribing practices. *Duped* refers to physicians who fall for patient scams. *Dishonest* refers to the small number of physicians who divert controlled substances for their own proprietary gain. *Disabled* physicians include those who prescribe substances for themselves because of their own abuse problems and those who prescribe for family members with similar problems.

Other physician characteristics include holding the view that prescribing a medication is the best response to all patient complaints (ie, "medication mania"), having the urge to help patients with all problems (ie, "hypertrophied enabling"), and lacking the ability to say no to patients (ie, "confrontation phobia").[11] These issues increase the tendency to prescribe, the frequency of prescribing for vague indications, and the difficulty physicians have in extracting themselves from controlled-drug prescribing once they have begun.

Table 4. Strategies to prevent prescription drug abuse

Screen for alcohol and drug abuse before prescribing controlled substances
Be knowledgeable about controlled substances
Be familiar with anxiety, depression, and pain syndromes
Document all prescription drugs in medical record
Adopt safe prescribing practices
Use controlled-substance contracts

Prevention

A number of strategies can help prevent prescription drug abuse (table 4). First, physicians need to screen patients for substance use disorders before prescribing controlled substances. A full discussion of screening strategies is beyond the scope of this article, and there are excellent reviews on this topic.[13,14] The CAGE questionnaire is useful in screening for alcohol prob-

continued

CAGE: questionnaire to identify persons with alcohol use disorders (may be adapted to identify drug use disorders)

C ut down. Have you ever tried to cut down on your drinking?

A nnoyed. Have you ever been annoyed by criticism of your drinking?

G uilty. Have you ever felt guilty about your drinking?

E ye opener. Have you ever had an "eye opener"?

Any positive response should prompt a more thorough assessment.

lems and can be adapted to screening for drug problems as well (see box on this page). Physicians should inquire about problems with substances and review records of prior treating physicians, if available.

When initiating a controlled prescription, physicians should carefully document the diagnosis, rationale, and anticipated time course and symptom end point. Controlled substances should not be prescribed before the patient is examined and a diagnosis reached, and the refill policy should be stated up front. Physicians need to feel empowered to say no when a request for medications is inappropriate—even if it conflicts with the natural desire to relieve the patient's suffering. Finally, prescribing any psychotropic medication for the physician or for the physician's family members must be avoided.

Table 5. Safe prescribing practices for controlled substances

Ensure that there is clear clinical indication for the drug
Define the therapeutic end point
State your refill policy up front
Avoid prescribing multiple substances
Avoid giving multiple refills without office visits

Recommended prescribing practices

Several guidelines for office practice can limit prescription drug abuse (table 5). Prescription pads should be locked away to limit patient access, and a group practice policy should be formulated about after-hours prescriptions of controlled substances. Telling patients about such practice rules at the time controlled substances are prescribed can help avoid conflicts in the future.

Meticulous records need to be kept of all controlled substances that are prescribed, both during routine office hours and after hours. Use of a controlled-drug refill chart can be helpful. The advent of computerized medical records may make documentation of refills easier, especially for physicians who work at multiple practice sites.

Many pain management programs routinely use "pain contracts," although their efficacy for limiting prescription drug abuse is not well established. To our knowledge, there are no standard or validated contracts for patients using controlled substances. Fishman and colleagues[15] reviewed opioid contracts from 39 academic pain centers and found substantial diversity. If used, the contract should include clear expectations of proper medication use, methods for monitoring appropriate use (eg, pill counts, toxicology screens), and the consequences of improper use (eg, the medication may be discontinued).

Because the contract is intended to improve the therapeutic alliance between patients and physicians, a less pejorative term (eg, patient agreement form) is

recommended. Other physicians may prefer to use an informed consent form, which educates the patient about the controlled substance and the risk of physical and psychologic dependence. Examples of consent forms for ongoing opioid therapy can be obtained from professional organizations.[16]

Interventions

Patients who meet the criteria for prescription drug misuse or inappropriate use need to be educated, which includes explicit instructions about how to take medicine and clear expectations that the medication will be used only for the condition for which it was prescribed. Patients who do not change their behavior after such education are moving into the area of abuse.

If abuse is detected, it is important to remain professional, empathetic, and nonjudgmental. The patient's primary complaint should be the focus of physician statements (eg, "I know you have been suffering with your back pain and I want to help you, but I am concerned that your use of narcotics has become a problem"). Clinical behaviors related to the catastrophic use of controlled prescription drugs are rare but should precipitate immediate cessation of prescribing (see table 1). These include commission of felonies (eg, altering a prescription, selling controlled drugs), intentional or unintentional overdoses, and threatening of the physician or staff.

Use of consultants

Experts in prescribing controlled substances and in the treatment of substance use disorders can be invaluable when abuse is suspected. Patients at risk for abuse may benefit from consultation with a pain management expert, who may be able to develop an opioid-sparing strategy. Direct communication with a consultant can allow a physician to maintain his or her role as the patient's primary physician while obtaining help from an expert. If controlled substances are to be continued, the primary care physician and consultant need to decide which one of them will be responsible for prescribing.

Controlled drugs for noncancer pain syndromes

Use of controlled substances for chronic noncancer pain syndromes is controversial. A full discussion of this topic is beyond the scope of this article; however, physicians who engage in such practices need to closely follow guidelines from their state medical board and keep meticulous records. Some states have strict mandates that must be met before controlled substances can be prescribed on an ongoing basis. More extensive initial screening for chemical dependency disorders and monitoring efforts are indicated. Portenoy[17] has outlined the issues relevant to this topic in a recent review.

Conclusion

Prescription drug abuse occurs in a small but significant number of patients for whom controlled substances are prescribed. Physicians need to maintain a balance that allows for appropriate relief of patient suffering without undue risk. Several patient and physician factors increase the risk for prescription drug abuse. Recognition of these factors and

continued

J. Harry Isaacson, MD, John A. Hopper, MD, Daniel P. Alford, MD, MPH, and Ted Parran, MD

Dr Isaacson is associate professor of medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland. Dr Hopper is assistant professor, departments of internal medicine, pediatrics, psychiatry, and behavioral neuroscience, Wayne State University School of Medicine, Detroit. Dr Alford is assistant professor, section of general internal medicine, Boston University School of Medicine. Dr Parran is associate clinical professor of medicine and family medicine, Case Western Reserve University School of Medicine.

Correspondence: J. Harry Isaacson, MD, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, 9500 Euclid Ave, NA 24, Cleveland, OH 44195. E-mail: isaacsj@ccf.org.

implementation of prevention strategies can allow physicians to prescribe controlled substances in a safe, effective manner. PCM

References

1. **Finch J.** Prescription drug abuse. Prim Care 1993;20(1):231-9
2. **National Institute on Drug Abuse.** NIDA InfoFacts: prescription pain and other medications. Available at: http://www.drugabuse.gov/infofacts/painmed.html. Accessed Jun 3, 2005
3. **Simoni-Wastila L.** The use of abusable prescription drugs: the role of gender. J Womens Health Gend Based Med 2000;9(3):289-97
4. **Regier DA, Farmer ME, Rae DS, et al.** Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. JAMA 1990;264(19):2511-8
5. **US Department of Health and Human Services.** Prescription drugs abuse and addiction. National Institute on Drug Abuse research report series. Available at: http://www.nida.nih.gov/ResearchReports/Prescription/prescription5.html. Accessed May 6, 2004
6. Emergency department trends from the drug abuse warning network, final estimates 1994-2001. Rockville, Md: Substance Abuse and Mental Health Services Administration, Office of Applied Studies, 2002:111
7. **Vastag B.** Mixed message on prescription drug abuse. JAMA 2001;285(17):2183-4
8. **Reid MC, Engles-Horton LL, Weber MB, et al.** Use of opioid medications for chronic noncancer pain syndromes in primary care. J Gen Intern Med 2002;17(3):173-9
9. **Joranson DE, Gilson AM.** Policy issues and imperatives in the use of opioids to treat pain in substance abusers. J Law Med Ethics 1994;22(3):215-23
10. **Federation of State Medical Boards of the United States.** Model guidelines for the use of controlled substances for the treatment of pain. Available at: http://www.medsch.wisc.edu/painpolicy/domestic/model.htm. Accessed May 6, 2004
11. **Parran T Jr.** Prescription drug abuse: a question of balance. Med Clin North Am 1997;81(4):967-78
12. **Smith DE,** Seymore RB. Prescribing practices: the educational alternative for the misprescriber. Proceedings of the White House Conference on Prescription Drug Abuse, Washington, DC, 1980
13. **Weaver MF, Jarvis MA, Schnoll SH.** Role of the primary care physician in problems of substance abuse. Arch Intern Med 1999;159(9):913-24
14. **Isaacson JH, Schorling JB.** Screening for alcohol problems in primary care. Med Clin North Am 1999;83(6):1547-63
15. **Fishman SM, Bandman TB, Edwards A, et al.** The opioid contract in the management of chronic pain. J Pain Symptom Manage 1999;18(1):27-37
16. **American Academy of Pain Medicine.** Consent for chronic opioid therapy. Available at: http://www.painmed.org/productpub/statements/pdfs/opioid_consent_form.pdf. Accessed May 6, 2004
17. **Portenoy RK.** Opioid therapy for chronic nonmalignant pain: a review of the critical issues. J Pain Symptom Manage 1996;11(4):203-17

Homeless Chronicity and Health-Related Quality of Life Trajectories Among Adults With Addictions

Stefan G. Kertesz, MD, MSc, Mary Jo Larson, PhD, MPA,† Nicholas J. Horton, ScD,‡ Michael Winter, MPH,§ Richard Saitz, MD, MPH,¶ and Jeffrey H. Samet, MD, MA, MPH‖*

Background: New federal initiatives target funds toward chronically homeless as distinct from other homeless persons. Few data exist, however, to substantiate the implications of chronic homelessness for major health outcomes.

Objectives: Using data from a 2-year cohort of addicted persons, we tested whether changes in mental and physical health-related quality of life (HRQOL) differed according to homeless chronicity.

Methods: Using self-reported homelessness, we classified subjects as chronically homeless (CH; n = 60), transitionally homeless (TRANS; n = 108), or as housed comparison subjects (HSD; n = 106). The Short Form-36 Health Survey, administered at baseline and 2 follow-ups over a period of 2 years, provided a Mental Component Summary (MCS) and a Physical Component Summary (PCS) for HRQOL. Mixed model linear regression was used to test the association between housing status, MCS, and PCS. Additional models assessed whether medical, psychiatric, addiction, and social support measures could account for HRQOL differences.

Results: All subjects had low MCS scores at study entry (mean, 31.2; SD, 12.6). However, there was a significant housing status-by-time interaction ($P = 0.01$). At final follow-up, CH and TRANS subjects had lower adjusted MCS scores than HSD subjects (33.4, 38.8, and 43.7 for the 3 groups, respectively; all $P \leq 0.01$). By contrast, housing status and PCS were not significantly associated ($P = 0.19$). Medical, psychiatric, addiction, and social support variables had significant associations with MCS, and their inclusion in the regression reduced the apparent effect of housing status on MCS.

Conclusions: Chronic homelessness was associated with especially poor mental but not physical HRQOL over time. These findings reinforce a new typology of homelessness.

Key Words: homelessness, quality of life, substance abuse, longitudinal models

(*Med Care* 2005;43: 574–585)

Contemporary policy concerning the homeless has embraced a new typology for this population, based on the chronicity of homelessness. A statistical analysis of nightly shelter use in 2 cities identified persons with prolonged or recurrent episodes of homelessness as "chronically" or "episodically" homeless, respectively, and distinguished these persons from "transitionally homeless" persons with infrequent episodes.[1] Because the chronically and episodically homeless are more likely to have mental illness and addiction, and to account for a greater proportion of service use, some have proposed that interventions target these groups.[1,2]

Recent federal, state, and metropolitan initiatives to end homelessness embrace this typology.[1,3–5] The United States Congress appropriated $35 million to assist chronically homeless adults in Fiscal Year (FY) 2003 and proposed double that amount for FY2005.[6,7] Additionally, funding from the Department of Housing and Urban Development prioritizes housing the chronically homeless.[8]

From the *Department of Medicine, Division of Preventive Medicine, University of Alabama at Birmingham School of Medicine, Birmingham, Alabama and the Deep South Center on Effectiveness, Birmingham Veterans Affairs Medical Center, Birmingham, Alabama; the †Institute for Health Services Research & Policy, New England Research Institutes, Watertown, Massachusetts; the ‡Department of Mathematics, Smith College, Northampton, Masschusetts; the §Data Coordinating Center, Boston University School of Public Health, Boston, Massachusetts; the ¶Section of General Internal Medicine, Boston University School of Medicine/Boston Medical Center, and Department of Epidemiology and the Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, Massachusetts; and the ‖Section of General Internal Medicine, Boston University School of Medicine/Boston Medical Center, and Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts.

Primary grant support for this study came from the National Institute on Drug Abuse (K23-DA15847 and 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 findings were presented at national meetings of the Society of General Internal Medicine (Vancouver, Canada, May 2, 2003), the College on Problems of Drug Dependence (Bal Harbor, FL, June 17, 2003), the International Society of Urban Health (New York, October 17, 2003), and the Association for Medical Education and Research on Substance Abuse (Baltimore November 8, 2003).

Reprints: Stefan G. Kertesz, MD, MSc, University of Alabama at Birmingham School of Medicine, Division of Preventive Medicine, 1530 3rd Avenue South MT608, Birmingham, AL 35294. E-mail: skertesz@uab.edu.

ISSN: 0025-7079/05/4306-0574

The federal definition for "chronically homeless" (which combines Kuhn and Culhane's research categories of *chronically* and *episodically* homeless) is an unaccompanied single adult with a disabling condition, which may include addiction, who has experienced homelessness continuously for a year or more, or 4 times during a period of 3 years.[6]

Current understandings of chronic homelessness rely principally on an analysis of shelter and service utilization from 2 cities, where chronically homeless persons represented 10% of persons experiencing homelessness but accounted for 50% of shelter use.[1] Comparative data characterizing health differences between chronically and nonchronically homeless persons, however, are sparse. Kuhn et al used records and limited self-report data to show higher levels of mental, medical, and substance use problems among the chronically homeless compared with other homeless persons.[1] Their study did not profile health status prospectively and did not apply standard measures of health-related quality of life (HRQOL). One study showed that a HRQOL measure (Short Form-12) can be applied to homeless persons, but it did not include longitudinal follow-up or consider homeless chronicity.[9]

If homeless chronicity does signal differences in HRQOL over time, then additional attention to the subgroup of chronically homeless adults could be important to future services research and may support targeting interventions according to homeless chronicity.

To study the relationship between homeless chronicity and health status over time, we used data from a longitudinal cohort of persons with addictions to test the hypothesis that chronically homeless persons have persistently worse mental and physical HRQOL than transitionally homeless or housed persons. We also attempted to clarify the factors that might account for HRQOL in this cohort.

METHODS

We analyzed prospectively collected data from the Health Evaluation and Linkage to Primary Care trial (HELP), which was conducted among urban alcohol- and drug-dependent persons. The trial intervention, a facilitated referral to primary care, resulted in 69% of the intervention group obtaining a primary care visit, versus 53% of controls ($P = 0.0003$).[10]

Study Sample

The trial from which these data are drawn commenced in an urban, short-term (4-to 6-day) inpatient detoxification unit. Trial eligibility criteria included being older than 17 years of age and having an indication of alcohol, heroin, or cocaine as the substance of first or second choice. Exclusions were having a primary care provider and seeing that provider on at least 1 occasion in the preceding 2 years; being pregnant; having a Mini-Mental State Examination[11] score of less than 21; lacking fluency in either English or Spanish; having fewer than 3 contacts available to facilitate follow-up; or having plans to leave the Boston area within 24 months. Of 642 eligible subjects, 470 (73%) consented to participate.

Subjects randomly assigned to the HELP intervention received a 90-minute clinical session with a physician, nurse, and social worker before leaving the detoxification unit, along with an individualized future appointment for primary medical care. Control subjects did not receive this intervention. Baseline interviews occurred from June 1997 to March 1999. Subjects provided information regarding demographic, social and health status, and they were sought for additional interviews at 6-month intervals over the ensuing 2 years.

Subjects eligible for this prospective cohort analysis of homeless chronicity and HRQOL had completed at least 2 of 4 follow-up assessments (6, 12, 18, and/or 24 months) representing 301 of 470 (64%) subjects. From this group we excluded subjects who did not respond to questions regarding residential status or HRQOL at baseline ($n = 8$), did not indicate residential status at a follow-up interview ($n = 4$), or who did not offer complete responses for other variables in our regression analyses ($n = 15$), leaving 274 (58%) of the total cohort for analysis. We compared these 274 subjects to those not included ($n = 196$) to assess potential for selection bias.

Housing Status

The main predictor variable, "housing status," was devised as 3 categories: chronically homeless, transitionally homeless, and housed. We used baseline and follow-up observations to identify a subgroup whose residential history approximated the federal definition for chronic homelessness.[6] The federal requirement for a "disabling condition" was met by addiction sufficiently severe that subjects had sought detoxification (an entry criterion to the original trial). The federal definition for chronic homelessness incorporates persons with *recurrent episodes* of homelessness over a period of years. At baseline and follow-up, subjects reported the number of nights in "an overnight shelter" and "on the street, without shelter" in the previous 6 months or since the last interview. The sum of street/shelter nights, adjusted to a 6-month equivalent (eg, 30 nights over the course of 12 months equaled 15 nights over 6 months), represented nights "literally homeless"[12] per 6 months. A subject approximated the federal definition by reporting 1 or more homeless nights at all 3 assessments (most had more than 1 night homeless; see Results section).

Subjects reporting no homeless nights at any interval were considered "housed," whereas subjects reporting homeless nights during 1 or 2 assessments only were considered "transitionally homeless." To assure that housing status was based on the same number of observations per subject, we

restricted analysis to the first 3 available interviews per subject.

Outcome Variables

Outcomes were from subjects' responses to the Medical Outcomes Study 36-item Short Form-36 (SF-36) version 1, a reliable, validated questionnaire reflecting diverse aspects of HRQOL,[13–15] including a range of mental and physical health experiences pertaining to the previous 4 weeks. Successful and clinically meaningful applications of the SF-36 (or shorter variants like the MOS-SF-20 or the SF-12) to substance abusers and the homeless[9] support its validity in this population.[16–18]

Main analytic outcomes were mental and physical HRQOL over the course of 3 assessments, as reflected in the Mental Component Summary (MCS) and the Physical Component Summary (PCS) of the SF-36. Scores range from 0 (worst) to 100 (best), with 50 being the norm for the US population, 10 the standard deviation, and differences of 5 considered clinically and socially meaningful.[19] We also investigated 8 summary scales derived from the SF-36 (physical function, role physical, pain, general health, vitality, social function, role-emotional, and mental health); these scales also range from 0 to 100 but have larger standard deviations and were not normed to a US population sample. This study's interest in modeling HRQOL longitudinally meant that each subject contributed 3 time-separated values for MCS, for PCS and for the 8 scales.

Explanatory Variables

We compared housing status groups according to a range of demographic, behavioral, and health variables measured throughout the study. These variables derived from review of the health outcomes and homeless literature,[20–22] consultation with experts in health-related quality of life data, and clinical experience caring for homeless and housed persons with addictions.

Sociodemographic variables included age, sex, race/ethnicity, marital status, and employment and were assessed at baseline. Two self-report baseline measures represented psychiatric status: ever been prescribed psychiatric medication for mental illness and ever considered or attempted suicide. Medical status was specified as 2 time-varying scores: (1) count of episodic medical conditions from a list of 23 conditions (eg, pneumonia, a gunshot wound) and (2) cumulative count of chronic conditions from a list of 13 conditions[23] (eg, diabetes, asthma; the full list is shown in Appendix A). Substance use severity was specified as 3 time-varying measures, the drug and alcohol composite scores of the Addiction Severity Index (ASI/drug and ASI/alcohol),[24] and a validated score for addictive consequences.[25] Social support was from validated measures of perceived support from family (PSS-Fa) and friends (PSS-Fr).[26]

Analysis

The 3 housing status groups were compared with respect to the range of the aforementioned demographic, health, and social parameters. To assess the relationship between housing status and HRQOL over time, we prepared *core* and *fully adjusted* multivariable models with the HRQOL measures (MCS or PCS) as outcomes. For both models, we used a general linear model for correlated data,[27,28] which permits multiple observations on each subject over time while adjusting for the correlated nature of these data. In this approach, we modeled the outcome (eg, MCS or PCS) on the basis of 3 observations per subject provided by 274 subjects. For example, a subject "observation" at time 0 (baseline) includes both fixed (eg, age, gender, race/ethnicity) and time-varying (eg, addiction severity) characteristics as well as an indication of time (eg, time = 0 for data from the baseline interview). Subsequent observations of the same subject (at time 1 and time 2) retain the same values for fixed characteristics (eg, sex) while permitting variation in others (eg, addiction severity).

The *core* model offered the most direct contrast of the 3 housing status groups over time, with statistical adjustment only for age, gender, race/ethnicity (white, black, Hispanic, other), time and randomization group. To further demonstrate how residential status related to HRQOL, we separately modeled each of the 8 SF-36 scales in relation to the predictors included in the *core* model.

The *fully adjusted* model was developed to explore whether putative explanatory variables would attenuate any HRQOL differences between housing status groups in the *core* model. For these *fully adjusted* models, we added the following covariates to the *core* model: lifetime history of receipt of psychiatric medication for a mental condition, lifetime history of suicidal ideation or attempt, number of episodic medical conditions in the previous 6 months, cumulative number of chronic medical conditions, ASI/drug, ASI/alcohol, PSS-Fa, and PSS-Fr.

Because data came from a trial of facilitated referral to primary care, sensitivity analyses included repetition of the *core* model with terms for the interaction of housing status by trial arm and the *core and fully adjusted* models with terms for actual receipt of primary care (categorized as 0, 1, or 2+ visits) and any substance abuse treatment (from administrative records).

The Fisher exact test and Wilcoxon rank sum tests were used to compare the 3 groups at baseline and to compare those cohort subjects included (n = 274) and excluded (n = 196) from analysis. We used PROC MIXED for longitudinal regression models, generating model-adjusted MCS and PCS scores for each of the 3 housing status groups at baseline

(time = 0) and at the first and second available follow-up interviews (time = 1 or time = 2, respectively). We fitted all models with terms for the interaction of time-by-housing status. Contrast tests permitted post-hoc pairwise comparisons. We used SAS statistical software version 8.2.[29]

RESULTS

Subject Characteristics

Among the 274 subjects in this analysis, 60 (22%) were chronically homeless (ie, reported homeless episodes before all research interviews, including baseline and 2 follow-ups), 106 (39%) were stably housed, and 108 (39%) were transitionally homeless (ie, reported homelessness before 1 or 2 but not all 3 interviews) during the 2-year study time frame. Among subjects who reported any homeless nights in an interval, the median number of nights in the previous 6 months was 24 (interquartile range, 6–95). As shown in Table 1, chronically homeless subjects were older, less likely to be black, and less likely to be married. They had nonsignificantly greater numbers of acute and chronic medical conditions. They were more likely to report alcohol as substance of choice, had higher alcohol but not drug severity scores, and more drug and alcohol use consequences. They also had greater psychiatric morbidity and scored lower on social support. The 3 groups did not differ with respect to the proportion randomly assigned to the trial intervention (P = 0.66).

The 3 prospectively defined housing status groups differed with respect to homeless experience before study entry (Table 1). Chronically homeless subjects experienced a mean of 15 months homeless during the 5 years before study entry, compared with 8 months and 0.2 months for transitionally homeless and housed subjects, respectively ($P < 0.001$).

Characteristics of Subjects Included Versus Those Excluded From Analytic Cohort

To assess for selection bias, we compared the 274 subjects who formed the basis for this analysis to the 196 subjects who were not included primarily because of lack of follow-up for at least 2 follow-up interviews. The 2 groups did not differ for most comparisons and notably did not differ regarding history of homelessness at baseline or baseline SF-36 MCS or PCS scores. Of 22 other characteristics, included subjects differed by race/ethnicity ($df = 3$, $P = 0.01$ for comparison across 4 race/ethnicity groups); included subjects were somewhat more likely to be black (53% versus 37%) and less likely to be white (32% versus 43%) compared with excluded subjects. Included subjects were somewhat more likely to identify cocaine (38% versus 27%) and somewhat less likely to identify alcohol (36% versus 44%) as their substance of choice ($df = 2$, $P = 0.05$). Similarly, included subjects had higher drug severity scores (mean ASI/drug 0.3 versus 0.2, $P = 0.01$).

Health-Related Quality of Life

At study entry, subjects had low MCS scores (unadjusted mean, 31.2; SD 12.6), regardless of housing status, and PCS scores (unadjusted mean, 47.7; SD 10.5) that were slightly lower than the US norm of 50. Baseline unadjusted MCS scores were 30.7, 30.5, and 32.3, and PCS scores were 45.8, 47.2, and 49.4 for chronically homeless, transitionally homeless and housed individuals, respectively.

The core longitudinal regression models, which included only age, sex, race/ethnicity, randomization group, and time as covariates, identified a clinically and statistically significant relationship between housing status and mental but not physical HRQOL over time (Figs. 1 and 2). For MCS, we observed an interaction of housing status with time ($F[4,265] = 3.3$, $P = 0.01$), along with a significant main effect of housing status ($df = 2$, $P < 0.0001$). MCS scores at baseline, adjusting for other model variables, were similarly low for chronically homeless, transitionally homeless, and housed subjects (adjusted scores 27.7, 27.7, and 29.8, respectively, $P > 0.15$ for all pairwise contrasts) However, at final follow-up, the chronically homeless and transitionally homeless had lower adjusted scores than housed subjects (adjusted means 33.4, 38.8, and 43.7, respectively; pairwise contrasts: $P < 0.0001$ for chronically homeless and $P = 0.005$ for transitionally homeless subjects versus housed). In the core model for physical HRQOL, the main effect of housing status was not significant ($df = 2$, $P = 0.19$), and there was no interaction of housing status with time ($F[4,265] = 0.41$, $P = 0.80$; Fig. 2).

Core longitudinal regression models designed to assess the relationship between housing status and the 8 individual scales from the SF-36 were broadly consistent with the results for the MCS and PCS (shown in Appendix B). Significant main effects of housing status were noted for the following scales: Role–Physical ($P = 0.001$), Bodily Pain ($P = 0.002$), General Health ($P = 0.009$), Vitality ($P = 0.01$), Social Function ($P < 0.0001$), Role–Emotional ($P < 0.0001$), and Mental Health ($P < 0.0001$). Chronically homeless subjects obtained poorer scores than the other groups over time on all but 1 subscale. Significant or near-significant interactions of housing status with time were observed for Physical Function ($P = 0.08$), Social Function ($P = 0.07$), Role–Emotional ($P = 0.08$), and Mental Health scales ($P = 0.02$).

As reviewed above and in Table 1, the housing status groups differed in many ways. Our fully adjusted statistical model added 8 potential explanatory variables in a longitudinal analysis for the outcome of MCS. All 8 variables were significantly associated with MCS over time (ie, receipt of psychiatric medication, lifetime history of suicidal attempt/ideation, ASI/drug and ASI/alcohol composite scores, numbers of episodic and chronic medical conditions, and perceived social support from family and friends, see Table 2).

TABLE 1. Comparison of Chronically Homeless, Transitionally Homeless, and Housed* Subjects at Time of Entry to a Residential Detoxification Unit

	Chronically Homeless, n = 60	Transitionally Homeless, n = 108	Housed, n = 106	*P* Value
Demographics				
Age, years	38.2 (SD, 7.9)	36.1 (SD, 8.1)	35.2 (SD, 6.8)	0.03
Male, %	80	79	72	0.38
Race/ethnicity, %				0.08
Black	40	55	59	
White	48	30	26	
Hispanic	10	8	9	
Other	2	7	5	
Married, %	5	5	13	0.05
Employed, full time, %	33	40	47	0.29
Highest yearly income in last 5 years, %				0.79
<$5000	14	13	15	
$5000–10,999	26	18	24	
$11,000–$19,999	29	27	25	
$20,000+	31	42	36	
Nonwork income, %†	43	29	34	0.16
Housing status				
Months homeless, previous 5 years	15.0 (SD, 17.1)	8.1 (SD, 14.4)	0.2 (SD, 0.8)	<0.0001
Mean nights homeless, previous 6 months	69 (SD, 68.8)	27.1 (SD, 48.5)	0 (SD, 0)	<0.0001
Medical status				
No. chronic medical conditions	1.0 (SD, 1.1)	0.8 (SD, 1.2)	0.7 (SD, 0.8)	0.17
No. episodic medical conditions, previous 6 months	0.9 (SD, 1.2)	0.95 (SD, 1.4)	0.6 (SD, 0.8)	0.10
No. hospital admissions, previous 6 months	0.23 (SD, 0.7)	0.19 (SD, 0.4)	0.09 (SD, 0.3)	0.20
No. emergency department visits, previous 6 months	1.5 (SD, 2.4)	0.7 (SD, 1.1)	0.5 (SD, 0.9)	0.0005
Medical severity score‡	0.43 (SD, 0.3)	0.39 (SD, 0.4)	0.32 (SD, 0.4)	0.05
Substance abuse				
Substance of choice, %§				0.005
Alcohol	55	36	25	
Cocaine	23	38	45	
Heroin	22	26	29	
Alcohol severity¶	0.61 (SD, 0.3)	0.47 (SD, 0.3)	0.38 (SD, 0.3)	0.0001
Drug severity¶	0.23 (SD, 0.2)	0.28 (SD, 0.1)	0.27 (SD, 0.1)	0.25
Drug and alcohol consequences‖	37.9 (SD, 6.7)	36.9 (SD, 6.8)	34.1 (SD, 6.8)	<0.0001
Psychiatric				
Suicide ideation/attempt, ever, %	40	31	24	0.08
Psychiatric medication, ever, %	30	21	20	0.30
Depressive symptoms at baseline**	36 (SD, 13.2)	33 (SD, 12.0)	32 (SD, 12.4)	0.04
Physical or sexual abuse, ever, %††	80	67	71	0.22
Social support				
Social support from friends‡‡	6.1 (SD, 4.0)	6.5 (SD, 3.9)	7.6 (SD, 4.1)	0.04
Social support from family‡‡	5.4 (SD, 4.8)	6.4 (SD, 4.6)	7.9 (SD, 4.7)	0.002

(Continued)

TABLE 1. *(Continued)*

	Chronically Homeless, n = 60	Transitionally Homeless, n = 108	Housed, n = 106	*P* Value
Health-related quality of life				
SF-36 PCS scale scores	45.6 (SD, 9.6)	47.2 (SD, 12.0)	49.4 (SD, 9.1)	0.08
SF-36 MCS scale scores	30.7 (SD, 12.8)	30.5 (SD, 11.9)	32.3 (SD, 13.2)	0.59

*Subjects were designated "chronically homeless" if they reported homelessness at each of 3 assessments, spaced at least 6 months apart, over the course of 2 years. Subjects reporting homelessness at one or 2 of 3 assessments were designated "transitionally homeless," whereas subjects reporting no homelessness were considered "housed" (see Methods section).

†Nonwork income included receipt of any federal or state monetary benefit (eg, Social Security, unemployment insurance) or child support.

‡Medical severity is from the medical composite score of the Addiction Severity Index, on a scale of 0 to 1, with 1 indicating greatest medical severity.[24]

§Substance of choice was defined on basis of subject's self-report at time of initial screening (see Methods section).

¶Drug and alcohol composite scores from the Addiction Severity Index, on a scale of 0 to 1, with 1 indicating greatest severity.[24]

‖Drug and Alcohol consequences computed from the Inventory of Drug and Alcohol Consequences.[25]

**Depressive symptoms, using the Center for Epidemiologic Studies Depression Scale.[50]

††Past history of abuse was based on the subject's response to questions seeking a lifetime history of exposure to physical or sexual abuse.

‡‡Perceived social support from friends and family.[26]

Comparing the core and fully adjusted models, inclusion of the 8 additional variables attenuated the statistical association between housing status and MCS. Specifically, in the core model, chronic homelessness (relative to housed status) was associated with a difference in MCS of −5.3 (95% CI −9.4 to −1.2). In the fully adjusted model (Table 2), chronic homelessness was associated with an MCS difference of −2.3 (95% CI −5.9 to +1.3). Graphic representation of the fully adjusted model suggested that there remained differences between the groups' mean MCS scores over time (Fig. 3), suggesting that some health status differences were not fully explained by model variables.

In sensitivity analyses, the interaction of trial randomization group and housing status was nonsignificant (F [2,263] = 0.05, P = 0.95), suggesting that the implications of chronic homelessness were similar regardless of trial arm. Repetition of the core and fully adjusted model for MCS with terms for receipt of primary care and addiction treatment attenuated but did not eliminate the relationship between housing status and MCS. This was also the case using models treating time as a continuous as opposed to an ordinal variable (data not shown).

DISCUSSION

This study showed that poor mental HRQOL is the norm for individuals entering a publicly funded, inner-city detoxification unit, and that the chronically homeless (22% of this sample) had markedly worse mental HRQOL over 2 years after detoxification compared with transitionally homeless and housed subjects. Physical HRQOL did not tend to differ by housing status. Although the literature has docu-

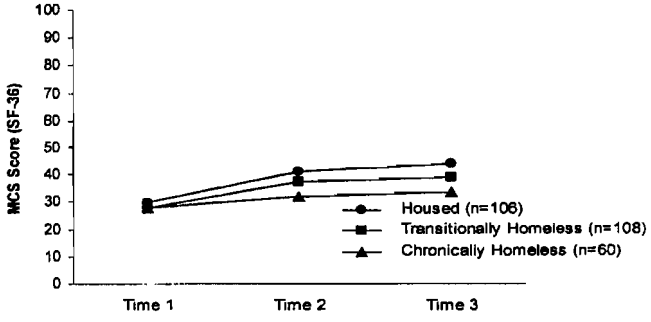


FIGURE 1. Core model mental health-related quality of life (MCS score of the SF-36) among 274 subjects recruited at detoxification and followed up to 2 years. Core model MCS scores are adjusted for age, sex, race/ethnicity, and randomization group, including a time-by-housing status group interaction term. In this mixed linear regression model, $P < 0.0001$ for the difference between the 3 housing status groups, and $P = 0.01$ for the interaction of time and housing status group.

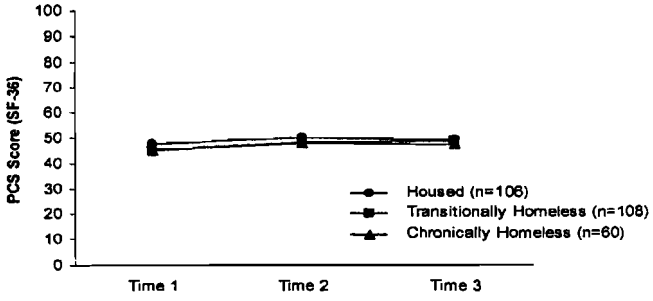


FIGURE 2. Core model physical health-related quality of life (PCS score of the SF-36) among 274 subjects recruited at detoxification and followed up to 2 years. Core model PCS scores are adjusted for age, sex, race/ethnicity, and randomization group, including a time-by-housing status group interaction term. In this mixed linear regression model, $P = 0.19$ for the difference between the 3 housing status groups.

TABLE 2. Multivariable Longitudinal Regression Model of Mental Health-Related Quality of Life (MCS) Among Subjects Recruited at Detoxification and Followed Up to 2 Years

Variable	Estimate (SE)	*F* Value	*P* Value*
Housing status		F (2,263) = 3.04	0.05
Chronically homeless	−2.3 (1.8)		
Transitionally homeless	—		
Housed	3.1 (1.5)		
Time		F (2,263) = 14.77	<0.0001
Time = 0 (Baseline)	−5.2 (1.4)		
Time = 1 (1st Follow-up)	−0.7 (1.2)		
Time = 2 (2nd Follow-up)	—		
Interaction terms		F (4,263) = 2.14	0.08
Time = 0, chronically homeless	2.5 (2.2)		
Time = 1, chronically homeless	−0.99 (1.96)		
Time = 0, housed	−3.1 (1.8)		
Time = 1, housed	−1.6 (1.7)		
Age	0.2 (0.1)	F (1,263) = 5.23	0.02
Sex		F (1,263) = 13.43	0.0003
Male	4.3 (1.2)		
Race/ethnicity		F (3,263) = 1.26	0.29
Black	2.6 (2.2)		
White	2.2 (2.3)		
Hispanic	−0.2 (2.6)		
Other	—		
Pyschiatric			
Lifetime suicidal ideation/attempt	−2.3 (1.1)	F (1,263) = 3.97	0.05
Ever received psychiatric medication	−2.9 (1.3)	F (1,263) = 5.31	0.02
Medical			
No. episodic medical conditions	−0.8 (0.4)	F (1,263) = 4.77	0.03
No. chronic medical conditions	−0.9 (0.3)	F (1,263) = 8.69	0.004
Addiction			
Drug severity (ASI/Drug)	−27.2 (3.1)	F (1,263) = 75.94	<0.0001
Alcohol severity (ASI/Alcohol)	−3.3 (1.4)	F (1,263) = 5.36	.02
Social support			
Support from family	0.4 (0.1)	F (1,263) = 18.46	<0.0001
Support from friends	0.4 (0.1)	F (1,263) = 13.60	0.0003
Randomization group		F (1,263) = 0.35	0.56
Control	−0.6 (0.97)		
Intercept	31.9 (3.6)		

*For variables composed of multiple categories, *P* values reflect type 3 tests of fixed effects.

mented health vulnerabilities associated with homelessness in cross-sectional samples,[30–36] to our knowledge, this study is unique in correlating HRQOL with homelessness over time and strengthened by inclusion of a housed comparison group. Federal, state, and metropolitan plans[3–6,37] to address homelessness refer explicitly to homeless chronicity, and these findings help to substantiate the validity of this typology. The findings underscore the importance of chronic homelessness in accentuating the burden of mental health-related problems and social distress.

This cohort was drawn from a randomized trial of facilitated referral to primary care,[10] but our findings were not altered by inclusion of terms representing trial arm or receipt of primary care. The baseline MCS score in this study sample was 31.2 (SD, 12.6), which is consistent with other persons entering detoxification[18] and similar to persons with clini-

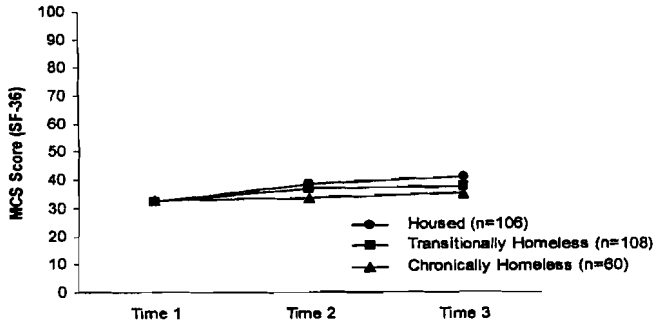


FIGURE 3. Fully adjusted mental health-related quality of life (MCS score of the SF-36) among 274 subjects recruited at detoxification and followed up to 2 years. Fully adjusted MCS scores are adjusted for age, sex, race/ethnicity, randomization group, lifetime history of receipt of psychiatric medication for a mental condition, lifetime history of suicidal ideation or attempt, number of episodic medical conditions in the previous 6 months, cumulative number of chronic medical conditions, ASI/drug, ASI/alcohol, PSS-Fa, and PSS-Fr, including a time-by-housing status group interaction term. In this mixed linear regression model, $P = 0.05$ for the difference between housing status groups and $P = 0.08$ for the interaction of group and time.

cally diagnosed depression[14] or adjustment disorder with depressed mood.[38] Because people typically seek help when addictive consequences are at their worst,[39] the subjects' low baseline MCS scores are unsurprising. Improvement of scores after detoxification, noted in another published report,[22] may or may not reflect treatment effects and could be expected for any disorder marked by cyclic exacerbations. Although this study found post-treatment improvement in mental HRQOL scores, the chronically and transitionally homeless subjects failed to match the improvement observed among housed subjects and, in fact, all 3 groups differed in post-hoc contrast tests. Thus, detoxification alone did not improve the mental health problems of these homeless and addicted adults.

Other published studies have demonstrated poor mental HRQOL among addicted persons,[17,40,41] but these studies have not focused on the implications of homelessness and, with one exception,[16] the studies did not address change over time.

Previous studies applying variants of the Short Form Health Survey to the homeless reported cross-sectional results and were less able to characterize the associations between homeless chronicity and mental HRQOL.[9,42] Gelberg and Linn[42] found associations between mental distress and numerous health indicators, but duration of homelessness was not statistically related to psychological distress. The discrepancy with the present report probably relates to their use of a cross-sectional sample. Cross-sectional samples may overrepresent the long-term homeless[2] and cannot consistently include meaningful numbers of transitionally homeless or housed individuals.

Our fully adjusted statistical model showed that several distinct vulnerabilities helped to account for inferior mental HRQOL among chronically homeless subjects. These vulnerabilities included psychiatric illness, medical problems, addiction severity, and poor social support. Residual confounding is possible because this study was neither designed nor powered to test the universe of potential vulnerabilities. None of the identified vulnerabilities is unexpected in the homeless[31,34,42–44] and all have previously been shown to be associated with decrements in HRQOL in studies of housed persons.[14,17,20,45] Our study underscores the complex determinants of HRQOL, particularly the mental component, among chronically homeless substance abusers and highlights the insufficiency of any attempt to construe the vulnerability of this population as resulting from addiction alone.

The question of whether homelessness itself causes poor health is unlikely to be resolved by observational data.[46] Because homelessness impedes self-care,[47] a degree of causality is somewhat plausible in light of reports that housing the homeless was associated with reduced subsequent health service use.[48]

Some limitations merit acknowledgment. This study's definition for "chronic homelessness," like the federal government's,[6] combines Kuhn and Culhane's categories of "chronic homelessness" (single long spells) and "episodic homelessness" (recurrent spells).[1] Future policy and research may "lump" categories similarly, as a reliable distinction between Kuhn's definitions for chronic and episodic homelessness may require nightly observational data from multiple shelters for several years coupled with the assumption that a person is not homeless when he or she is not in a shelter. We suspect that differentiation of subjects with long, continuous homeless episodes could demonstrate differences even more extreme than reported here.

Additionally, although paralleling the federal definition, this study could incur misclassification in that homelessness occurring 3 times during a 2-year period qualified as "chronically homeless," a reflection of our sample size and study duration, whereas the federal method references 4 episodes in 3 years. On this basis, we acknowledge a degree of arbitrariness to this study's "transitionally homeless" designation. Robust differences in homeless experience over the 5 years prior to baseline (Table 1) suggest, however, that our housing status groups were quite distinct.

To characterize implications of housing status, we relied on a sample of addicted persons without primary care. Although this could affect generalizability, 48% of currently homeless persons have experienced problems of substance abuse in the past year,[32] and federally funded programs of supportive housing typically recruit clients from addiction

treatment, suggesting that these findings do speak to the chronically homeless population currently targeted for intervention. Our finding that homeless chronicity had a strong association with mental HRQOL, even within a population of substance abusers and after statistical adjustment for addiction severity, mitigates against the notion that substance abuse alone accounts for the mental health status of chronically homeless individuals.

Any study attempting to prospectively follow inner-city homeless and housed substance abusers is susceptible to irregular follow-up, with the resulting potential for selection bias. We believe selection bias is unlikely to account for the finding of poor HRQOL in this sample, in part because the baseline findings are consistent with a cross-sectional assessment from the same treatment system,[40] and because included subjects did not differ significantly from excluded subjects in regard to homelessness, the outcome variables (ie, MCS and PCS), and most other characteristics, and the study retained a cohort of 60 chronically homeless persons.

This study's strengths include the size of its cohort and the availability of valid measures for HRQOL, as well as social support, physical health, psychiatric status and addiction severity. Use of longitudinal observations permitted us to take into account the reality that medical and addiction severity are not stable characteristics. Inclusion of a measure of social support incorporates the impact of social issues on personal health, a fact emphasized in the social epidemiology literature.[49]

This study has implications for policymakers, clinicians, and researchers. For policymakers, these findings highlight the extreme vulnerability of the chronically homeless when compared with other addicted persons who are transitionally homeless or housed, and in that way support the new federal policy distinction as meaningful.

For clinicians meeting new clients in detoxification or comparable settings, a multidimensional sociomedical evaluation of all clients may be impractical. Questions about the number and frequency of homeless episodes, however, may suggest the range of potential vulnerabilities. Plans for care of chronically homeless persons entering treatment facilities likely require attention to medical conditions, psychiatric comorbidity, addictive behavior, social support, and housing.

The robust difference in mental health prognosis between chronically as opposed to transitionally homeless individuals may be relevant to health services research. Should future studies of health care access and utilization consistently take homeless chronicity into account? Our findings suggest an affirmative answer to this empiric question that awaits confirmation from future studies.

ACKNOWLEDGMENTS

The authors thank the staff of the HELP study for their help in the conduct of this research and Ashley Mullins for assistance with production of this manuscript and its illustrations.

REFERENCES

1. Kuhn R, Culhane DP. Applying cluster analysis to test a typology of homelessness by pattern of shelter utilization: results from the analysis of administrative data. *Am J Community Psychol.* 1998;26:207–232.
2. Phelan JC, Link BG. Who are "the homeless"? Reconsidering the stability and composition of the homeless population. *Am J Public Health.* 1999;89:1334–1338.
3. Chicago Continuum of Care. Getting Housed, Staying Housed: A Collaborative Plan to End Homelessness. Available at: http://www.ich.gov/slocal/plans/chicago.pdf. Accessed March 25, 2005.
4. Commission on Homelessness (Atlanta). Blueprint to End Homelessness in Atlanta in Ten Years. Available at: http://www.unitedwayatlanta.org/homeless/docs/AtlantaHomeless.pdf. Accessed March 25, 2005.
5. Governor's Executive Commission for Homeless Services Coordination (Commonwealth of Massachusetts). Housing the Homeless: A More Effective Approach. Available at: http://www.state.ma.us/homelesscommission/pdf/docs/finalechrpt.pdf. Accessed March 25, 2005.
6. United States Department of Housing and Urban Development. Notice of Funding Availability (NOFA) for the Collaborative Initiative to Help End Chronic Homelessness. Available at: http://www.hud.gov/offices/cpd/homeless/apply/2002nofa/index.cfm. Accessed March 25, 2005.
7. Ending chronic homelessness. *New York Times.* March 13, 2002.
8. United States Department of Housing and Urban Development. Notice of HUD's Fiscal Year (FY), 2004 Notice of Funding Availability (NOFA), Policy Requirements and General Section to the SuperNOFA for HUD's Discretionary Programs. Vol 69. Washington, DC: Federal Register; 2004:26941–27291.
9. Larson CO. Use of the SF-12 instrument for measuring the health of homeless persons. *Health Serv Res.* 2002;37:733–750.
10. Samet JH, Larson MJ, Horton NJ, et al. Linking alcohol and drug dependent adults to primary medical care: a randomized controlled trial of a multidisciplinary health evaluation in a detoxification unit (The Health Evaluation and Linkage to Primary care [HELP] Study). *Addiction.* 2003;98:509–516.
11. 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.
12. Rossi PH, Wright JD, Fisher GA, et al. The urban homeless: estimating composition and size. *Science.* 1987;235:1336–1341.
13. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31:247–263.
14. Ware JE, Kosinski M, Keller SD. *SF-36 Physical & Mental Health Summary Scales: A User's Manual.* Boston: The Health Institute; 1994.
15. Ware JE, Kosinski M, Keller SD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30:473–483.
16. Falck RS, Wang J, Siegal HA, et al. Longitudinal application of the Medical Outcomes Study 36-Item Short-Form health survey with not-in-treatment crack-cocaine users. *Med Care.* 2000;38:902–910.
17. Daeppen JB, Krieg MA, Burnand B, et al. MOS-SF-36 in evaluating health-related quality of life in alcohol-dependent patients. *Am J Drug Alcohol Abuse.* 1998;24:685–694.
18. Stein MD, Mulvey KP, Plough A, et al. The functioning and well being of persons who seek treatment for drug and alcohol use. *J Subst Abuse.* 1998;10:75–84.
19. Ware JE, Snow KK, Kosinski M, et al. *SF-36 Health Survey Manual and Interpretation Guide.* Boston: The Health Institute; 1993.
20. Spitzer RL, Kroenke K, Linzer M, et al. Health-related quality of life in primary care patients with mental disorders. Results from the PRIME-MD 1000 Study. *JAMA.* 1995;274:1511–1517.
21. Johnson JG, Spitzer RL, Williams JB, et al. Psychiatric comorbidity, health status, and functional impairment associated with alcohol abuse and dependence in primary care patients: findings of the PRIME MD-1000 study. *J Consult Clin Psychol.* 1995;63:133–140.

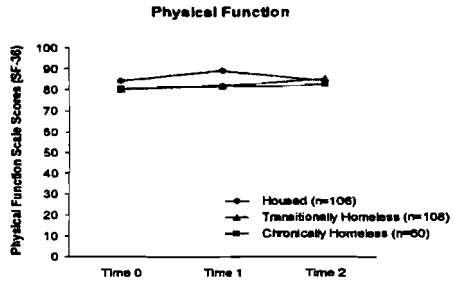
22. Garg N, Yates WR, Jones R, et al. Effect of gender, treatment site and psychiatric comorbidity on quality of life outcome in substance dependence. *Am J Addict.* 1999;8:44–54.
23. De Alba I, Samet JH, Saitz R. Burden of medical illness in drug- and alcohol-dependent persons without primary care. *Am J Addict.* 2004;13: 33–45.
24. McLellan AT, Kushner H, Metzger D, et al. The Fifth Edition of the Addiction Severity Index. *J Subst Abuse Treat.* 1992;9:199–213.
25. 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.
26. Procidano ME, Heller K. Measures of perceived social support from friends and from family: three validation studies. *Am J Community Psychol.* 1983;11:1–24.
27. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med.* 1997;16:2349–2380.
28. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics.* 1982;38:963–974.
29. SAS Institute Inc. *SAS/STAT Software: Changes and Enhancements, Release 8.2.* Cary, NC: SAS Institute Inc.; 2001.
30. Gelberg L, Linn LS. Assessing the physical health of homeless adults. *JAMA.* 1989;262:1973–1979.
31. Breakey W, Fischer P, Kramer M, et al. Health and mental health problems of homeless men and women in Baltimore. *JAMA.* 1989;262: 1352–1357.
32. Burt MR, Aron LY, Douglas T, et al. *Homelessness: Programs and the People They Serve. Technical Report of Findings of the National Survey of Homeless Assistance Providers and Clients.* Washington, DC: Urban Institute; 1999.
33. Ferenchick GS. The medical problems of homeless clinic patients: a comparative study. *J Gen Intern Med.* 1992;7:294–297.
34. Fischer PJ, Breakey WR. The epidemiology of alcohol, drug, and mental disorders among homeless persons. *Am Psychol.* 1991;46:1115–1128.
35. Langnase K, Muller MJ. Nutrition and health in an adult urban homeless population in Germany. *Public Health Nutr.* 2001;4:805–811.
36. Gelberg L, Linn LS, Usatine RP, et al. Health, homelessness, and poverty. A study of clinic users. *Arch Intern Med.* 1990;150:2325–2330.
37. The Secretary's Work Group on Ending Chronic Homelessness. *Ending Chronic Homelessness: Strategies for Action.* Washington, DC: Department of Health and Human Services; 2003.
38. Jones R, Yates WR, Williams S, et al. Outcome for adjustment disorder with depressed mood: comparison with other mood disorders. *J Affect Disord.* 1999;55:55–61.
39. Tucker JA, Gladsjo JA. Help-seeking and recovery by problem drinkers: characteristics of drinkers who attended Alcoholics Anonymous or formal treatment or who recovered without assistance. *Addict Behav.* 1993;18:529–542.
40. Smith KW, Larson MJ. Quality of life assessments by adult substance abusers receiving publicly funded treatment in Massachusetts. *Am J Drug Alcohol Abuse.* 2003;29:323–335.
41. Ryan CF, White JM. Health status at entry to methadone maintenance treatment using the SF-36 health survey questionnaire. *Addiction.* 1996; 91:39–45.
42. Gelberg L, Linn LS. Psychological distress among homeless adults. *J Nerv Ment Dis.* 1989;177:291–295.
43. Bahr HM. *Skid Row: An Introduction to Disaffiliation.* New York: Oxford University Press; 1973.
44. Wright JD, Weber E. *Homelessness and Health.* New York: McGraw-Hill, Inc; 1987.
45. Sherbourne CD, Meredith LS, Rogers W, et al. Social support and stressful life events: age differences in their effects on health-related quality of life among the chronically ill. *Qual Life Res.* 1992;1:235–246.
46. Hwang SW. Is homelessness hazardous to your health? Obstacles to the demonstration of a causal relationship. *Can J Public Health.* 2002;93: 407–410.
47. Gelberg L, Gallagher TC, Andersen RM, et al. Competing priorities as a barrier to medical care among homeless adults in Los Angeles. *Am J Public Health.* 1997;87:217–220.
48. Culhane D, Metraux S, Hadley T. Public service reductions associated with placement of homeless persons with severe mental illness in supportive housing. *Housing Policy Debate.* 2002;13:107–162.
49. Berkman LF, Kawachi I. *Social Epidemiology.* New York: Oxford University Press; 2000.
50. Radloff L. A self-report depression scale for research in the general population. *Appl Psych Meas.* 1977;1:385–404.

APPENDIX A. Episodic and Chronic Medical Conditions Solicited from All Subjects at Baseline and Follow-up Interviews

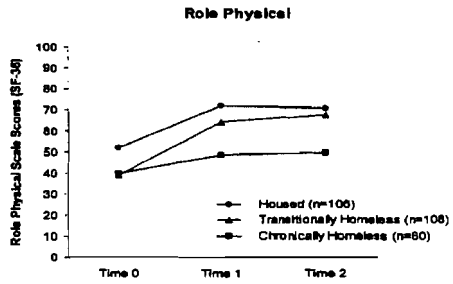
Episodic Medical Conditions*	Chronic Medical Conditions†
Skin infections like cellulitis or an abscess	Seizures, epilepsy or convulsions
Pneumonia (includes PCP)	Asthma, emphysema or chronic lung disease
Septic arthritis (joint infection requiring antibiotics)	Heart attack (myocardial infarction)
TB (tuberculosis)	Heart failure (congestive heart failure)
Endocartitis (heart infection)	Other heart disease (requiring medication)
An ulcer (peptic, stomach, or intestinal/duodenal)	High blood pressure
Pancreatitis	Ongoing or chronic liver disease (ie, cirrhosis or "fatty liver"; chronic hepatitis B or C)
Abdominal or stomach pain requiring an overnight hospital stay	Kidney failure
Vomiting (throwing up) blood	Chronic arthritis or osteoarthritis (lasting more than 3 months; degenerative joint disease)
Hepatitis	Peripheral neuropathy (constant numbness, tingling, or burning in the feet)
Blood clots in the legs or lungs	Cancer
Osteomyelitis (bone infection)	Diabetes
Chest pain while using cocaine, resulting in emergency room visit or hospital stay	Stroke (cerebrovascular accident)
Jaundice (turning yellow)	
Low back pain lasting more than 3 months that required medical attention	
Seizures or convulsions	
Drug or alcohol overdose (requiring medical attention right away)	
A gunshot wound (been shot)	
A stab wound (been stabbed or cut)	
Any accidents or falls requiring medical attention	
Fractures (broken) or dislocations to bones or joints	
An injury from a road traffic accident such as a car or motorcycle	
A head injury	

*At baseline and at follow-up, subjects were asked whether they had experienced any of the episodic medical conditions during the previous 6 months or since last meeting with the research assistant.

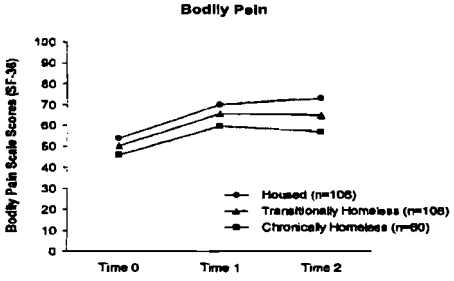
†At baseline, subjects were asked whether they had ever been told by a doctor that they had any of the listed chronic medical conditions. At follow-up, subjects were asked whether they had been told by a doctor during that they had any of the listed chronic medical conditions the previous 6 months or since the last time they met with the research assistant.



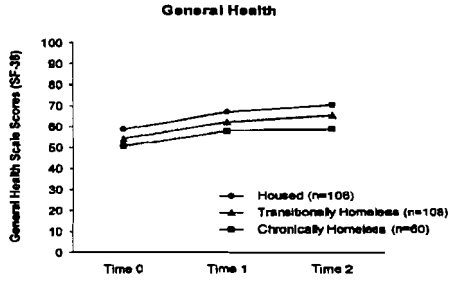
In this mixed linear regression model, p=.16 for the difference between the 3 housing status groups, and p=.08 for the interaction of time and housing status group.



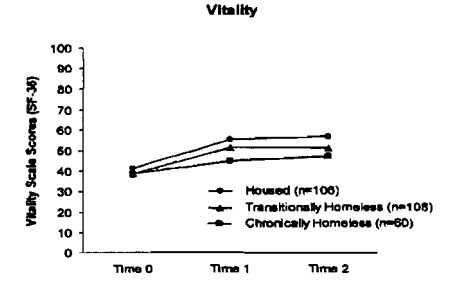
In this mixed linear regression model, p=.001 for the difference between the 3 housing status groups, and p=.14 for the interaction of time and housing status group.



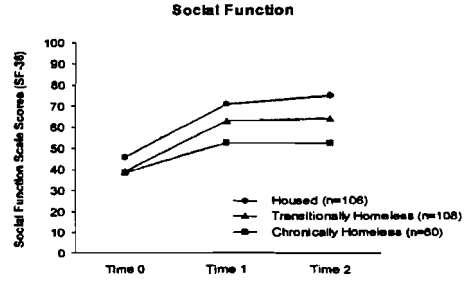
In this mixed linear regression model, p=.16 for the difference between the 3 housing status groups, and p=.08 for the interaction of time and housing status group.



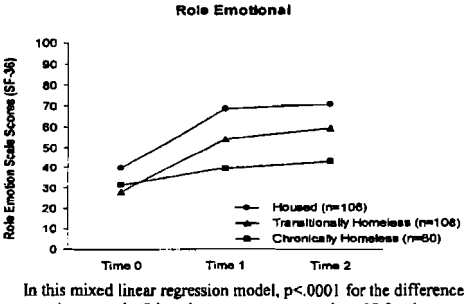
In this mixed linear regression model, p=.001 for the difference between the 3 housing status groups, and p=.14 for the interaction of time and housing status group.



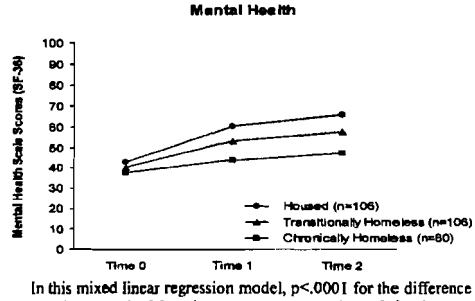
In this mixed linear regression model, p=.01 for the difference between the 3 housing status groups, and p=.34 for the interaction of time and housing status group.



In this mixed linear regression model, p<.0001 for the difference between the 3 housing status groups, and p=.07 for the interaction of time and housing status group.



In this mixed linear regression model, p<.0001 for the difference between the 3 housing status groups, and p=.08 for the interaction of time and housing status group.



In this mixed linear regression model, p<.0001 for the difference between the 3 housing status groups, and p=.02 for the interaction of time and housing status group.

Appendix B. Core model individual SF-36 scale scores for 274 subjects recruited at detoxification and followed up to 2 years. Core model SF-36 scores are adjusted for age, sex, race/ethnicity, and randomization group, including time-by-housing status group interaction term.

Health Utility Ratings for a Spectrum of Alcohol-Related Health States

Kevin L. Kraemer, MD, MSc,† Mark S. Roberts, MD, MPP,*†‡ Nicholas J. Horton, ScD,§ Tibor Palfai, PhD,‖ Jeffrey H. Samet, MD, MA, MPH,¶ Naomi Freedner, MPH,¶ Nicole Tibbetts, BA,** and Richard Saitz, MD, MPH¶††*

Background: Preference-based utility ratings for health conditions are important components of cost–utility analyses and population burden of disease estimates. However, utility ratings for alcohol problems have not been determined.

Objectives: The objectives of this study were to directly measure utility ratings for a spectrum of alcohol-related health states and to compare different methods of utility measurement.

Design, Setting, and Subjects: The authors conducted a cross-sectional interview of 200 adults from a clinic and community sample.

Methods: Subjects completed computerized visual analog scale (VAS), time tradeoff (TTO), and standard gamble (SG) utility measurement exercises for their current health, a blindness scenario, and for 6 alcohol-related health state scenarios presented in random order. The main outcome measures were the utility ratings, scaled from 0 to 1, and anchored by death (0) and perfect health (1).

Results: The 200 subjects were middle-aged (mean, 41 ± 14 years), 61% women, and racially diverse (48% black, 43% white). Utility ratings decreased as the severity of the alcohol-related health state increased, but differed significantly among the VAS, TTO, and SG methods within each health state. Adjusted mean (95% confidence interval) utility ratings for alcohol dependence (VAS, 0.38 [0.34–0.41]; TTO, 0.54 [0.48–0.60]; SG, 0.68 [0.63–0.73]) and alcohol abuse (VAS, 0.53 [0.49–0.56]; TTO, 0.71 [0.65–77]; SG, 0.76 [0.71–0.81]) were significantly lower than utility ratings for non-drinking, moderate drinking, at-risk drinking, current health, and blindness.

Conclusions: Utility ratings for alcohol-related health states decrease as the severity of alcohol use increases. The low utility ratings for alcohol abuse and alcohol dependence are similar to those reported for other severe chronic medical conditions.

Key Words: alcohol, drinking, preferences, utilities, utility assessment

(*Med Care* 2005;43: 541–550)

From the *Center for Research on Health Care, Division of General Internal Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; the †Department of Health Policy and Management, Graduate School of Public Health, and the ‡Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania; the §Department of Mathematics, Smith College, Northampton, Massachusetts; the ‖Department of Psychology, Boston University, Boston, Massachusetts; the ¶Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, Massachusetts; **DM-Stat, Inc., Medford, Massachusetts; and the ††Department of Epidemiology and the Center to Prevent Alcohol Problems Among Young People, Boston University School of Public Health, Boston, Massachusetts.

This research was supported by grants R01AA12617 and K23AA00235 from the National Institute of Alcohol Abuse and Alcoholism, National Institutes of Health.

Preliminary findings from this work were presented at the Annual Meeting of the Society of General Internal Medicine in May 2003 in Vancouver, British Columbia, Canada, and at the Annual Scientific Meeting of the Research Society on Alcoholism in June 2003 in Fort Lauderdale, Florida.

Reprints: Kevin L. Kraemer, MD, MS., Center for Research on Health Care, Division of General Internal Medicine, University of Pittsburgh School of Medicine, 230 McKee Place, Suite 600, Pittsburgh, PA 15213. E-mail: kraemerkl@upmc.edu.

ISSN: 0025-7079/05/4306-0541

Cost–utility analysis (CUA) is a useful methodologic tool for comparing alternative health programs and for informing resource allocation decisions.[1–3] However, CUA has not been widely applied to alcohol prevention and treatment services[4,5] despite the large societal and healthcare burden from alcohol problems in the United States and internationally.[6–8] This shortcoming limits decision-makers' ability to choose among multiple effective alternatives for addressing alcohol-related illness. In addition, the funding of effective alcohol-targeted programs could suffer because public- and private-sector decision-making agencies increasingly use CUA to allocate resources for new pharmaceutical agents and other services.[9]

One barrier to the application of CUA to alcohol studies is the absence of data regarding societal and patient preference-based utility ratings for alcohol-related health states and outcomes. Preference-based utility ratings give a quantitative estimate of how much an individual values a health state relative to reference health states, typically death and full

health.[10] Importantly, utility ratings provide the quality-of-life adjustment weight needed to calculate quality-adjusted life-years (QALYs),[11,12] the effectiveness measure recommended for CUA studies by the Panel on Cost-Effectiveness in Health and Medicine.[1] To date, utility ratings and QALY estimates are largely unknown for alcohol-related health states. Major "societal cost of alcohol" studies,[6–8] international methodologic guidelines for measuring alcohol costs,[13] and the World Health Organization (WHO) Burden of Disease project[14,15] either do not include utilities and preferences for alcohol-related health states or estimate them for alcohol-attributable disease (eg, cirrhosis) but not for specific alcohol use patterns or diagnoses.[16–18] As a consequence, prior economic studies of alcohol programs have not adjusted for quality of life[19–22] or have limited the utility assessment to loss from alcohol-related injury.[23] Better estimates of utility ratings for alcohol problems could improve QALY estimates and increase the rigor of CUA for alcohol prevention and intervention services.

Our specific objectives in this study were to directly measure utility ratings for a spectrum of alcohol-related health states and to compare standard methods for measuring utility ratings. We hypothesized that: 1) utility ratings would decrease as the severity of the alcohol-related health state increased; and 2) utility ratings for specific health states would differ by measurement method.

METHODS

Study Design, Setting, and Subjects

The study was a cross-sectional, single interview of 100 primary care clinic attendees in Boston, Massachusetts, and 100 individuals from a community sample in Pittsburgh, Pennsylvania. Subjects were 18 years or older and English-speaking. There was no alcohol use or alcohol problem requirement for inclusion in the study. In Boston, potential subjects were approached in the waiting room of the Primary Care Clinic at Boston Medical Center, an urban academic medical center. Any adult in the waiting area was eligible, including patients or individuals accompanying them. In Pittsburgh, subjects were recruited through electronic and print advertisements within the University of Pittsburgh community and through advertisements in a major Pittsburgh newspaper. Research assistants conducted interviews with participating subjects in private areas of the clinic in Boston or in a private interview room in Pittsburgh. Subjects were paid for completing the 30- to 60-minute interview. The study protocol was approved by the Institutional Review Boards of Boston Medical Center and the University of Pittsburgh.

Utility Rating Exercises

We sequentially presented 1 actual health state and 7 hypothetical health state scenarios to subjects and asked them to imagine themselves with that health state. The actual health state was the subject's own current health. The hypothetical states were blindness and 6 alcohol-related health states (Figure 1): nondrinking, moderate drinking, at-risk drinking, alcohol abuse, alcohol dependence, and alcohol dependence in recovery. The health state of blindness was included to introduce subjects to the utility elicitation exercises with a chronic disability they could comprehend and evaluate and for which published utility ratings exist.[24–27] The 6 alcohol-related health states covered a broad spectrum of alcohol use behaviors and problems and were designed through a group process of drafting and revision with reference to standard definitions.[28]

The utility rating exercises were conducted with a laptop computer using Impact 3.0, a web-based, graphic interface, utility measurement software program.[29,30] Study subjects completed visual analog scale (VAS), time tradeoff (TTO), and standard gamble (SG) exercises[31–33] for each of the 8 health state scenarios. We presented the alcohol-related health state scenarios in random order to assess for potential bias from order effects. The 3 measurement methods were always performed in the same order (VAS first, TTO second, and SG third) for each health state before moving to the next health state.

The VAS was a vertically oriented scale anchored by perfect health (100) at the top and death (0) at the bottom. Subjects indicated their preference for a given health state by choosing a point on the scale between perfect health and death. Subjects rated just 1 health state at a time on the VAS rather than plotting all health states simultaneously. The utility was calculated as the subject's choice divided by 100. For example, the utility was 0.75 if the subject placed a mark at 75 on the 100-unit scale.

For the TTO, we asked subjects to choose between 20 years of life with the health state in question (eg, alcohol dependence) and the same or lesser amount of time in perfect health. The length of time in perfect health was alternated until the subject was indifferent (unwilling to give up additional time) between X years in perfect health and 20 years with the described health state. The utility was then calculated as X years divided by 20 years. For example, the utility was 0.75 if the subject was willing to give up 5 of 20 years in the described health state to have 15 years of perfect health (utility = 15 of 20 = 0.75).

For SG, we asked subjects to choose between a "sure thing" of living with the health state in question and a "gamble" in which there was some chance for perfect health but also a chance of death. The relative probability of perfect health (P) and death (1—P) was alternated until the rater was indifferent between the "sure thing" and the "gamble." The utility was calculated as the lowest probability of perfect health that the rater was willing to gamble for. For example, the utility for a health state was 0.75 if the subject accepted the gamble when the

Alcohol-Related Health States for Utility Measurement

1. Nondrinker
Imagine that you do not drink alcohol. Alcohol has never harmed your health, mood, social life or family life.

2. Moderate drinker
Imagine that you drink alcohol. You often enjoy a drink after work and on Saturdays you typically have a couple of drinks in the evening with your closest friends. You find that drinking alcohol helps you relax and makes social occasions more enjoyable. Your use of alcohol has never harmed your health, mood, social life or family life.

3. At-risk drinker
Imagine that you drink alcohol. Although you don't drink very often at home, when you go out with your friends, you have about 5 or 6 drinks. Usually you drink on weekend nights, but in the summer you drink about 3 times per week. Drinking has never harmed your health, mood, social life or family life. You have taken a few chances that you would not take if you were sober, such as getting rides home from friends who have been drinking. You haven't missed any work, although you are less productive at work the days after you have been drinking.

4. Alcohol Abuse
Imagine that you drink alcohol. Your friend thinks you drink too much and the two of you argue about your drinking frequently. Sometimes you have driven drunk, and several times you have been late for work the morning after you've been drinking. Sometimes after drinking you feel a burning in your stomach that lasts for days. You continue to drink even though you think alcohol might be causing some problems for you.

5. Alcohol Dependence
Imagine you drink alcohol. You need to drink to get rid of the shakes, to calm your nerves, and to get any sleep. You need to drink a lot just to feel the effects. Even though you know alcohol is hurting you, you can't seem to stop. You miss important family events because of your drinking. Your doctor has told you that drinking has damaged your liver. Several times in the past year drinking has caused indigestion, upper stomach pain, nausea, and vomiting.

6. Alcohol Dependence in Recovery
Imagine that you used to drink alcohol. You don't drink anymore, but sometimes you think about drinking. Several years ago you used to disappoint the people you were closest to (spouse, friends, relatives) by being drunk when they needed you. In the morning you felt shaky and anxious and needed a drink to calm yourself. You had trouble falling and staying asleep every night. You couldn't seem to control how much and how often you drank. You knew drinking was hurting you but you couldn't stop. Since you no longer drink, drinking isn't causing you problems these days.

Other Health State Scenarios for Utility Measurement

1. Current Health
We now want to ask you several questions about how you feel about your current state of health. When answering these questions remember to think about how you feel now in general, but not necessarily at this exact moment.

2. Blindness
We humans are visual creatures, but in some cases people are for one reason or another unable to see. This may be a genetic problem, a physical problem or the result from some injury. Blindness can be devastating, but is also a health condition that one can overcome mentally, if not physically. We would like to ask you a few questions regarding this condition.

FIGURE 1. Text of health state scenarios for utility measurement exercises.

probability of perfect health was 0.75 (probability of death, 0.25) but preferred the "sure thing" of remaining in the described health state when the gamble's probability of perfect health was 0.74 (probability of death, 0.26).

Other Assessments

We collected demographic data on age, gender, ethnicity, race, education, employment, health insurance, living situation, and marital status. Medical comorbidity was assessed by a validated self-report instrument.[34] Health-related quality of life of the subjects over the prior 4 weeks was assessed by the 12-item Short Form Health Survey (SF-12) and then summarized as Physical Component Summary (PCS-12) and Mental Component Summary (MCS-12) scores.[35] Health status was also measured by the EuroQoL instrument (EQ-5D), a prescored, generic health classification measure.[36,37] Numeracy (numerical ability) was measured with a validated 3-item questionnaire,[38] which uses word problems to: 1) estimate the number of times "heads" will come up from 1000 tosses of a fair coin; 2) estimate a percentage when given the denominator and numerator; and 3) estimate the numerator when given a percentage and denominator.

We asked subjects about family history of alcohol problems and assessed personal alcohol use with the 10-item Alcohol Use Disorders Identification Test (AUDIT)[39] and alcohol quantity–frequency questions. Subjects with an AUDIT score greater than or equal to 8 were considered to have "at-risk drinking" and indicated their readiness to change their alcohol use by completing the "Readiness Ladder," a vertical scale anchored by "no thought of changing" at 0 and "taking action to change" at 10. A full diagnostic assessment to assess alcohol abuse and alcohol dependence in our subjects was beyond the scope of this study and would have added to the interview burden.

Analysis

The main outcomes for analysis were the measured utility ratings, scaled continuously from 0 to 1, and anchored by death at 0 and perfect health at 1. Because the utility ratings were not normally distributed either before or after logarithmic transformation, we used both parametric (*t* tests)

and nonparametric methods (Kruskal-Wallis tests) to test for statistically significant differences in utility ratings between health states and between utility measurement methods. We report both medians and means for the utility ratings to give the reader an appreciation for the distribution of the data. Because tests of significance for differences between utility ratings were similar regardless of whether *t* tests or nonparametric tests were used, we report *P* values for the means only and assumed normality for the modeling strategy.

We used the general linear model for correlated data[40,41] to test for the effects of health state scenario on the utility rating. This type of regression analysis was necessary because the utility ratings were not independent observations but instead were correlated data with multiple observations per subject. The general linear model accommodated the correlated outcomes, allowed us to explore and control for other factors related to utility ratings, and provided flexibility in examining within and between health state effects, within and between subject group effects, and within and between utility measurement method effects.

In the general linear models, the utility ratings for the alcohol-related health states of each subject were treated as repeated-measures dependent variables. Because preliminary analyses indicated a significant interaction between the health state scenarios and utility measurement method, models were fit separately for each utility measurement method and adjusted for study site, gender, race, age, education, numeracy, health status, alcohol use, and the order in which the scenario was presented to the subject. Model covariates included variables that differed significantly between the 2 study sites. We used the general linear models for correlated data to generate parameter estimates and significance levels for the difference in utility rating between the reference state (nondrinking) and each of the 5 remaining alcohol-related health states. We generated predicted mean utility values and 95% confidence intervals for each of the 6 alcohol-related health states by substituting the mean value for each covariate into the regression equations and predicting the mean outcomes. This was done with the LSMEANS procedure in SAS.

All analyses were performed with SAS software. All *P* values were 2-tailed and considered significant at $P < 0.05$.

RESULTS

Subject Characteristics

The 200 subjects were middle-aged (mean, 41 ± 14 years; range, 18–87 years), 61% women, racially diverse (48% black, 43% white), and a majority (63%) had at least a high school education (see Table 1). Only 4% had an annual income greater than $40,000 per year. Subjects reported an average alcohol intake of 5.9 standard drinks per week, but the standard deviation was large (25 drinks) and 30% reported at least 1 binge (5 or more drinks on 1 occasion for men; 4 or more drinks on 1 occasion for women) in the 4 weeks before the interview. Forty-six (23%) subjects met criteria for at-risk drinking (defined as an AUDIT score ≥8) and 131 (66%) subjects had a family history of alcohol problems. Subjects from the primary care clinic sample were significantly older, more often a racial minority, and of lower socioeconomic status, health status, and numeracy than the community sample. The numeracy of the overall sample was poor: 58 (29%) answered 0 of 3 questions correctly, 66 (33%) answered 1 correctly, 47 (24%) answered 2 correctly, and only 28 (14%) answered all 3 correctly. Clinic sample subjects reported significantly less binge alcohol behavior and more readiness to change alcohol use than the community sample subjects.

Utility Ratings

Mean and median utility ratings for all 8 health state scenarios are shown in Table 2. For each measurement method, utility ratings decreased as the severity of the alcohol problem progressed from nondrinking to alcohol dependence. Mean and median VAS and mean TTO and SG utility ratings dropped immediately after the nondrinking health state, whereas TTO and SG median utility ratings were similar for nondrinking, moderate drinking, and at-risk drinking before decreasing for alcohol abuse and alcohol dependence. The utility ratings for the alcohol dependence in recovery health state tended to be similar to the ratings for at-risk drinking. In addition, the utility ratings for the blindness health state scenario were similar to published utilities for blindness (ranges: VAS 0.34–0.85, TTO 0.37–0.68, SG 0.41–0.80)[24–27] but were significantly higher than the utility ratings for alcohol dependence ($P < 0.0001$ for each of VAS, TTO, and SG) and similar to those for alcohol abuse ($P > 0.05$ for each of VAS, TTO, and SG). Utility ratings for the alcohol-related health states did not differ significantly between subjects with and without at-risk drinking (AUDIT ≥8) behavior or between subjects with and without a family history of alcohol problems. The 3 utility measurement methods yielded different utility ratings. Nearly all pairwise comparisons of means and medians between VAS, TTO, and SG utility ratings were significant at the $P < 0.05$ level for each health state (Table 2).

To explore the relationship of numeracy to utility ratings, we calculated the mean VAS, TTO, and SG utility estimates for each strata of numeracy (0 correct, 1 correct, 2 correct, or 3 correct) for 2 of the health states (moderate drinking and alcohol dependence). Utilities tended to be lower for those with poor numeracy, but this trend was not consistent across utility elicitation methods and the 2 health states. In addition, we assessed how respondents in 2 drinking groups, moderate drinkers (classified as drinks per week >0 and AUDIT <8; n = 80) and alcohol dependence (classified as AUDIT ≥20; n = 10), compared his or her own health

TABLE 1. Characteristics of Subjects by Total Sample and by Study Site

Characteristic	Total (n = 200)	Boston (n = 100)	Pittsburgh (n = 100)	*P* Value*
Age, mean ± SD	41.1 ± 13.9	45.9 ± 12.8	36.3 ± 13.4	<0.0001
	(range 18 –87)	(range 24 –87)	(range 18 –74)	0.0819
Gender, n (%)				
Female	122 (61)	67 (67)	55 (55)	
Male	78 (39)	33 (33)	45 (45)	
Race, n (%)				<0.0001
Black	95 (47.5)	62 (62)	33 (33)	
White	86 (43)	21 (21)	65 (65)	
American Indian/Alaskan native	4 (2)	4 (4)	0 (0)	
Other	15 (7.5)	13 (13)	2 (2)	
Marital status, n (%)				0.0318
Single	114 (57)	52 (52)	62 (62)	
Married	42 (21)	18 (18)	24 (24)	
Divorced	23 (11.5)	13 (13)	10 (10)	
Separated	9 (4.5)	8 (8)	1 (1)	
Widowed	12 (6)	9 (9)	3 (3)	
Education, n (%)				<0.0001
Greater than high school degree	126 (63)	55 (55)	71 (71)	
High school degree or less	74 (37)	45 (45)	29 (29)	
Annual income (n = 199), n (%)				0.0376
0–$10,000	94 (47.2)	57 (57)	37 (37)	
$10,001–20,000	36 (18.1)	19 (19)	17 (17)	
$20,001–30,000	40 (20.1)	14 (14)	26 (26)	
$30,001–40,000	21 (10.6)	7 (7)	14 (14)	
>$40,000	8 (4.02)	3 (3)	5 (5)	
Health status				
SF-12 Physical Component Summary, mean ± SD	46.8 ± 11.6	41.5 ± 12.2	52.0 ± 8.1	<0.0001
SF-12 Mental Component Summary, mean ± SD	46.6 ± 11.4	44.6 ± 12.1	48.6 ± 10.3	0.0114
EQ-5D Utility Score, mean ± SD	0.79 ± 0.26	0.76 ± 0.29	1.0 ± 0.19	<0.0001
Comorbidity score, mean ± SD	0.61 ± 1.20	0.94 ± 1.44	0.29 ± 0.80	0.0001
Alcohol use				
AUDIT score, mean ± SD	5.4 ± 6.9	5.0 ± 7.4	5.8 ± 6.3	0.4203
AUDIT ≥8, n (%)	46 (23.1)	19 (19.2)	27 (27)	0.1914
Drinks per week, mean ± SD	5.9 ± 25.0	6.5 ± 34.4	5.3 ± 8.6	0.7367
Binge, in past month, n (%)	59 (29.7)	22 (22.2)	37 (37)	0.0225
Readiness to change (if AUDIT ≥8, n = 46), mean ± D	4.8 ± 4.1	7.4 ± 3.2	3.0 ± 3.7	0.0001
Family history of alcohol problems (n = 190), n (%)	131 (69)	73 (73.4)	58 (64.6)	0.1889
Numeracy				
Number correct, mean ± SD	1.23 ± 1.02	0.71 ± 0.80	1.74 ± 0.96	<0.0001
Two or more correct responses, n (%)	75 (37.7)	15 (15.2)	60 (60.0)	<0.0001

*P values represent tests for significance between Boston and Pittsburgh subjects. Mean values were compared with Student t tests and categorical variables with χ^2 or Fisher exact test.

SD indicates standard deviation; SF-12, 12-item Short Form Health Survey; AUDIT, Alcohol Use Disorders Identification Test.

TABLE 2. Mean ± SD and Median (Interquartile Range) Utility Ratings for Each Health State Scenario and for Each Utility Measurement Method*

	Visual Analog Scale Mean ± SD Median (IQR)	Time Tradeoff Mean ± SD Median (IQR)	Standard Gamble Mean ± SD Median (IQR)
Alcohol-related health state scenarios			
Nondrinking	0.94 ± 0.09	0.97 ± 0.13	0.93 ± 0.15
	1.0 (0.90–1.0)	1.0 (1.0–1.0)	1.0 (0.90–1.0)
Safe drinking	0.85 ± 0.17	0.94 ± 0.20	0.88 ± 0.22
	0.90 (0.80–1.0)	1.0 (1.0–1.0)	1.0 (0.90–1.0)
At-risk drinking	0.72 ± 0.24	0.84 ± 0.30	0.82 ± 0.27
	0.80 (0.60–0.90)	1.0 (0.85–1.0)	0.96 (0.77–1.0)
Alcohol abuse	0.52 ± 0.23	0.72 ± 0.35	0.75 ± 0.29
	0.50 (0.33–0.70)	0.88 (0.53–1.0)	0.89 (0.61–0.97)
Alcohol dependence	0.36 ± 0.22	0.54 ± 0.37	0.67 ± 0.29
	0.40 (0.20–0.50)	0.70 (0.05–0.85)	0.73 (0.55–0.90)
Alcohol dependence, in recovery	0.71 ± 0.24	0.86 ± 0.25	0.83 ± 0.24
	0.80 (0.53–0.90)	1.0 (0.85–1.0)	0.90 (0.77–0.99)
Other health state scenarios			
Current health	0.76 ± 0.22	0.93 ± 0.17	0.83 ± 0.25
	0.85 (0.70–0.90)	1.0 (1.0–1.0)	0.91 (0.78–1.0)
Blindness	0.55 ± 0.26	0.76 ± 0.32	0.74 ± 0.28
	0.60 (0.40–0.75)	0.90 (0.70–1.0)	0.81 (0.60–0.96)

*Pairwise comparisons of the means between utility measurement methods (VAS versus TTO, VAS versus SG, TTO versus SG) were significant at the $P < 0.01$ level for each health state except for VAS versus SG for nondrinking ($P = 0.41$), VAS versus SG for safe drinking ($P = 0.09$), TTO versus SG for at-risk drinking ($P = 0.16$), TTO versus SG for alcohol abuse ($P = 0.08$), and TTO versus SG for blindness ($P = 0.36$). P values (not shown) were similar in nonparametric comparisons of the medians.

SD indicates standard deviation; IQR, interquartile range; VAS, visual analog scale; TTO, time tradeoff; SG, standard gamble.

utility rating to the utility rating they gave to the hypothetical alcohol-related health state closest to their own. In general, no significant differences between subjects' actual health state and the corresponding hypothetical state were observed. The exceptions were moderate drinkers who rated their own health as lower (mean utility difference of −0.07, $P = 0.006$ on paired t test) on VAS and subjects with alcohol dependence who rated their own health as higher (mean utility difference +0.175, $P = 0.044$) on SG.

General Linear Models

Results of the 3 general linear models, 1 for each utility measurement method, are shown in Figure 2 and Table 3. Figure 2 depicts estimated means and 95% confidence intervals of utility ratings for the 6 alcohol-related health states derived from the models and adjusted for multiple covariates. The alcohol-related health state scenario was independently associated ($P < 0.001$) with utility ratings for each measurement method. For each utility measurement method, contrasts in estimated utility ratings between adjacent alcohol-related health states (eg, nondrinking versus moderate drinking, moderate drinking versus at-risk drinking, at-risk drinking versus alcohol abuse, and alcohol abuse versus alcohol dependence) were significant at $P < 0.001$ except for the contrasts of moderate drinking versus at-risk drinking (VAS, $P = 0.49$; TTO, $P = 0.56$; SG, $P = 0.69$).

For each utility measurement method, mean utility ratings were slightly, but significantly, lower for any given alcohol-related scenario if the scenario was presented first (estimated mean utility difference compared with presentation in the sixth [last] position: VAS, −0.051, $P < 0.001$; TTO, −0.048, $P = 0.02$; SG, −0.034, $P = 0.026$). The clinic sample subjects provided minimally, but significantly, lower utility ratings (estimated utility difference compared with Pittsburgh site: −0.06, $P = 0.009$) for the SG method only. Subjects' mental and physical health status, as summarized by the MCS-12 and PCS-12, were positively correlated with utility ratings for the VAS (0.01 utility increase per 10-unit increase in MCS-12, $P = 0.047$; 0.02 utility increase per 10-unit increase in PCS-12, $P = 0.003$) but not for TTO or SG. Subjects' current drinking habits, age, gender, race, education, and numeracy were not independent correlates of the utility ratings for alcohol-related health states (all P values > 0.05).

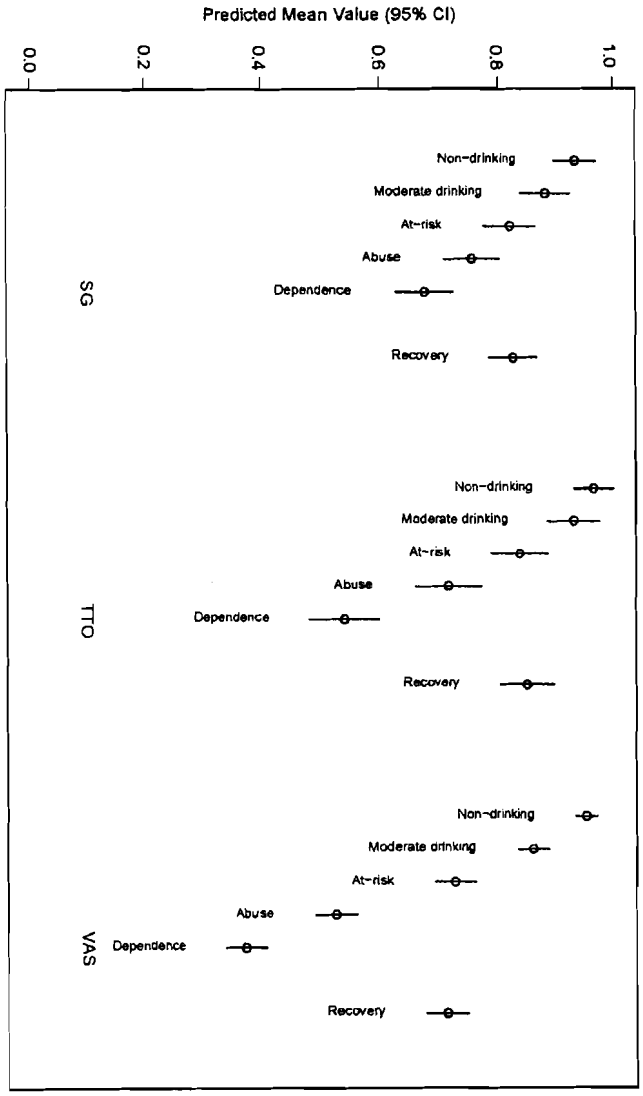


FIGURE 2. Predicted mean utility estimates (95% confidence intervals) from the general linear models for each alcohol-related health state and each utility measurement method (visual analog scale [VAS], time tradeoff [TTO], standard gamble [SG]).

In general linear models stratified by numeracy, subjects with poor numeracy tended to rate the scenarios lower than those with good numeracy. However, the utility rating differences between numeracy categories were significantly different for VAS but not for TTO and SG. In general linear models stratified by study site, subjects from Boston tended to rate the health states lower for each utility elicitation method. However, the only significant differences between sites were for at-risk drinking in the VAS and TTO models and for moderate drinking, at-risk drinking, and alcohol abuse in the SG model.

DISCUSSION

Our findings suggest utility ratings for a spectrum of alcohol-related health states can be estimated with standard utility measurement methods. The observed utility ratings appropriately decreased as the severity of the alcohol-related health states increased. In fact, utility ratings for the 2 most severe alcohol-related scenarios, alcohol abuse and alcohol dependence, were similar or lower than the utility ratings for the blindness scenario and similar to reported utility ratings for other severe chronic diseases such as major depression (utility = 0.44, by SG), congestive heart failure (utility = 0.46, by Health Utility Index), and moderate–severe osteoarthritis of the hip (utility = 0.50, by SG).[24] We found no consistent modifiers of utility ratings other than severity of the health state and the order of scenario presentation. Surprisingly, subjects' sociodemographic characteristics, personal alcohol use, and family history of alcohol problems did not independently correlate with the utility ratings. As expected, the VAS, TTO, and SG methods of measurement yielded different utility ratings for the alcohol-related health states.[32,42,43]

Few prior studies have attempted to estimate or directly measure utility ratings for alcohol disorders. The World Health Organization's Global Burden of Disease Project[44,45] used secondary data and expert opinion to assign disability weights to 22 indicator conditions through person tradeoff valuation and then extrapolated to other conditions. By extrapolation, "alcohol use disorders" received a disability weight of 0.180 (in which 0 = perfect health and 1 = death), equivalent to a utility of 0.82 (1—0.18 = 0.82). In recent work describing a societal utility index for substance abuse,[46] health state descriptions for alcohol and drug use states were presented to 143 respondents (primarily white, college-educated women who worked in substance abuse clinics) for TTO exercises. On average, respondents were willing to trade 2.34 years out of 10 years of life left (equivalent to a utility of 0.766) to avoid alcohol problems. Although the utility ratings for "alcohol use disorders" from these prior studies were higher than our observed values for alcohol abuse and alcohol dependence, the utility raters and breadth of alcohol-related health states in both prior studies were quite different from our study.

To our knowledge, no prior studies have reported estimates of utility ratings for moderate drinking behaviors and at-risk drinking. One surprising result from our study is the relatively low utility assessed to at-risk drinking, a scenario in which no significant adverse consequences have occurred. It is possible some subjects were imagining the potential future consequences (eg, car crash, missed work, hangover, medical problem) of the at-risk drinking scenario. In addition, stigma, negative perceptions, and religious beliefs about alcohol consumption may play a role in the low utility values for alcohol abuse and alcohol dependence and explain why some participants gave low utility ratings to

TABLE 3. Results of General Linear Models for Correlated Data*

Variable	Model 1—Visual Analog Scale (n = 199)			Model 2—Time Tradeoff (n = 199)			Model 3—Standard Gamble (n = 199)		
	Utility Estimate	Standard Error	*P* Value	Utility Estimate	Standard Error	*P* Value	Utility Estimate	Standard Error	*P* Value
Model intercept (constant)	0.792	0.056	<0.0001	0.956	0.062	<0.0001	0.914	0.083	<0.0001
Alcohol-related health state scenario									
Nondrinking (ref)	0	—	—	0	—	—	0	—	—
Moderate drinking	−0.092	0.012	<0.0001	−0.034	0.013	0.010	−0.052	0.013	<0.0001
At-risk drinking	−0.228	0.017	<0.0001	−0.128	0.022	<0.0001	−0.114	0.018	<0.0001
Alcohol abuse	−0.426	0.017	<0.0001	−0.252	0.026	<0.0001	−0.179	0.018	<0.0001
Alcohol dependence	−0.578	0.016	<0.0001	−0.425	0.027	<0.0001	−0.260	0.019	<0.0001
Alcohol dependence, in recovery	−0.241	0.016	<0.0001	−0.115	0.018	<0.0001	−0.108	0.014	<0.0001
Order of scenario presentation									
1	−0.051	0.015	<0.001	−0.048	0.021	0.020	−0.034	0.015	0.026
2	−0.015	0.014	0.264	−0.016	0.017	0.337	0.001	0.013	0.945
3	0.014	0.013	0.266	0.005	0.016	0.769	0.017	0.013	0.208
4	0.011	0.013	0.412	0.001	0.020	0.958	0.015	0.015	0.301
5	−0.012	0.011	0.295	−0.010	0.019	0.602	0.008	0.012	0.481
6 (ref)	0	—	—	0	—	—	0	—	—
Study site									
Boston	0.014	0.015	0.356	−0.012	0.020	0.569	−0.060	0.023	0.009
Pittsburgh (ref)	0	—	—	0	—	—	0	—	—
Age	0.0005	0.0004	0.244	−0.0007	0.0008	0.364	0.00004	0.0009	0.958
Gender									
Female	−0.002	0.012	0.899	−0.014	0.015	0.333	−0.006	0.019	0.753
Male (ref)	0	—	—	0	—	—	0	—	—
Race									
American Indian/Alaskan	0.025	0.025	0.310	−0.007	0.065	0.914	0.044	0.067	0.512
Black	−0.011	0.020	0.583	0.021	0.051	0.677	−0.001	0.058	0.983
White	−0.003	0.025	0.916	0.023	0.047	0.619	0.012	0.055	0.825
Other (ref)	0	—	—	0	—	—	0	—	—
Education									
Greater than high school degree	−0.005	0.0005	0.728	0.018	0.021	0.389	0.007	0.021	0.721
High school degree or less (ref)	0	—	—	0	—	—	0	—	—
SF-12 Mental Component Summary	0.001	0.0005	0.047	0.0003	0.0007	0.655	−0.001	0.001	0.302
SF-12 Physical Component Summary	0.002	0.0006	0.003	0.0003	0.0009	0.787	0.002	0.001	0.073
Personal alcohol use									
At-risk	0.009	0.016	0.589	0.021	0.022	0.327	−0.018	0.024	0.452
Low risk	0.026	0.014	0.071	0.012	0.026	0.644	−0.022	0.030	0.471
None (ref)	0	—	—	0	—	—	0	—	—
Numeracy (answered 2 or more of 3 questions correctly)									
Yes	0.008	0.017	0.656	0.013	0.017	0.464	0.003	0.025	0.912
No (ref)	0	—	—	0	—	—	0	—	—

*A separate model was fit for each of the 3 utility measurement methods (VAS, TTO, and SG). The outcomes are the utility ratings for the 6 alcohol-related health states, entered as repeated measures. Each model reflects the 6 utility observations for each of 199 subjects. Models are adjusted for order of scenario presentation, study site, age, gender, race, education, health status, personal alcohol use, and numeracy.

moderate drinking and at-risk drinking. This will be an important area for future research.

As expected, utility ratings for the alcohol-related health states differed significantly among the VAS, TTO, and SG measurement methods. Because the cognitive tasks required for each measurement method are different and the VAS requires no tradeoff, VAS values tend to be lower than TTO or SG valuations.[47] Although we cannot state definitively which utility value (VAS versus TTO versus SG) is the most appropriate for a specific alcohol-related health state, a reasonable and conservative approach is to use our SG estimates for current modeling efforts. The SG estimates are reasonable because SG is based directly on the fundamental axioms of utility theory,[48] and they have the least risk of overstating the disutility of alcohol problems. For alcohol dependence, an alternative approach is to use the utility difference between our alcohol dependence and alcohol dependence in recovery health states and to model this as the "utility loss" to the condition.

Our study has several limitations. First, the subjects did not represent a true random societal sample. Half of the subjects were a convenience sample recruited from the primary care clinic waiting room at an urban academic hospital and the other half were self-selected individuals from the community responding to an advertisement. We pooled the clinic and community samples because neither sample was a random sample, which needed to be kept pure. The diversity of the pooled sample ensured a wide range of preferences and perspectives, which we felt was important for this initial work. Future research should focus on repeating our measures in a random societal sample.

Second, the poor numeric ability of our sample is a potential threat to the validity of the measured utility ratings. However, numeracy was not a significant influence on utility rating in our general linear models. Further analyses indicated a tendency for lower utility ratings from those with low numeracy but, interestingly, this was most pronounced for VAS, the technique thought to require the least amount of numeric ability.

Third, all alcohol-related health states were presented as hypothetical scenarios to approximate "societal" values. We chose this approach because CUA models using a societal perspective are our first planned applications for the utility estimates. In general, the societal perspective is the proper perspective to take when societal resources must be allocated within budget constraints to maximize utility.[1]

Fourth, we did not assess subsets of disease severity within the diagnoses of alcohol abuse and alcohol dependence. Certainly, scenarios depicting a different level of severity for each of these may have yielded different utility ratings. In addition, the health state scenarios for alcohol abuse and dependence did not include text lauding the positive or enjoyable aspects of alcohol use.[49] We do not know if inclusion of positive language would have yielded higher utility ratings.

Lastly, we did not have subjects perform a simple rank ordering, by preference, of the health states before the utility exercises began. This prevented consistency checks of the rank ordering of health states by each utility measurement method[50] and prevented exploration of whether the graduated nature of the alcohol-related health states forced subjects to rate some health states (eg, nondrinking and moderate drinking) farther apart than their true preferences. We believe the random presentation order of the alcohol-related health states made this latter possibility less likely.

Despite these limitations, we believe our results represent an important step in developing utility weights for a broad spectrum of alcohol use behaviors and consequences and have practical implications for future alcohol treatment and policy changes. Such utility weights will be useful for CUA models that address alcohol prevention and treatment services and for population burden of disease calculations. For example, the low ratings we observed for at-risk drinking, alcohol abuse, and alcohol dependence suggest that failure to adjust for these "disutilities" in CUA models may underestimate the benefit of alcohol prevention and treatment programs, and potentially result in the underutilization and/or under-funding of such programs. In addition, our observed low utilities for the alcohol abuse and alcohol dependence health states suggest the true global burden of alcohol use disorders may be higher than the previously derived 58.3 million disability-adjusted life years (DALYs) from the WHO Global Burden of Disease project[15] and support arguments for greater resource allocation to alcohol-related programs.

In summary, alcohol-related health state utility ratings reflect the severity of alcohol use disorders and should be used in the future to estimate the effectiveness of alcohol treatment and prevention services. Future research in this area should focus on the utility rating differences among individuals with and without specific alcohol diagnoses, the impact of individuals' alcohol problems on the utility of their significant others (eg, spouse, partner, or close family member, coworkers), and the role of negative perceptions and stigma.

ACKNOWLEDGMENTS

The authors thank Julie Hirsch (Boston) and Susan Richey (Pittsburgh) for assistance with subject recruitment and data collection.

REFERENCES

1. Gold MR, Russell LB, Seigel JE, et al., eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.
2. Weinstein M, Stason W. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med*. 1977;13:716–721.
3. Drummond MF, O'Brien BJ, Stoddart GL, et al. *Methods for the*

Economic Evaluation of Health Care Programmes, 2nd ed. Oxford: Oxford University Press; 1997.
4. French MT. Economic evaluation of alcohol treatment services. *Recent Dev Alcohol.* 2001;15:209–228.
5. French MT. The role of health economics in substance abuse research. Recent advances and future opportunities. Overview. *Recent Dev Alcohol.* 2001;15:201–208.
6. Harwood HJ. *Updating Estimates of the Economic Costs of Alcohol Abuse in the United States: Estimates, Update Methods and Data.* National Institute of Alcohol Abuse and Alcoholism; 2000.
7. Single E, Robson L, Xie X, et al. *The Costs of Substance Abuse in Canada.* Ottawa, Ontario, Canada: Canadian Center on Substance Abuse; 1996.
8. Collins D, Lapsley H. *The Social Costs of Drug Abuse in Australia in 1988 and 1992, monograph no. 30.* Canberra: Commonwealth Department of Human Services and Health, Australian Government Publishing Service; 1996.
9. Glennie JL, Torrance GW, Baladi JF, et al. The revised Canadian Guidelines for the Economic Evaluation of Pharmaceuticals. *Pharmacoeconomics.* 1999;15:459–468.
10. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ.* 1986;5:1–30.
11. Gold MR, Stevenson D, Fryback DG. HALYS and QALYS and DALYS, oh my: similarities and differences in summary measures of Population Health. *Annu Rev Public Health.* 2002;23:115–134.
12. Johannesson M, Jonsson B, Karlsson G. Outcome measurement in economic evaluation. *Health Econ.* 1996;5:279–296.
13. Single E, Collins D, Easton B, et al. *Proposed International Guidelines for Estimating the Costs of Substance Abuse.* Ottawa, Ontario, Canada: Canadian Center on Substance Abuse; 1995.
14. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet.* 1997;349:1436–1442.
15. Murray CJ, Lopez AD. *The Global Burden of Disease.* Boston: Harvard School of Public Health; 1996.
16. Room R, Graham K, Rehm J, et al. Drinking and its burden in a global perspective: policy considerations and options. *Eur Addict Res.* 2003;9: 165–175.
17. Rehm J, Room R, Monteiro M, et al. Alcohol as a risk factor for global burden of disease. *Eur Addict Res.* 2003;9:157–164.
18. Ezzati M, Hoorn SV, Rodgers A, et al. Estimates of global and regional potential health gains from reducing multiple major risk factors. *Lancet.* 2003;362:271–280.
19. Holder HD, Blose JO. The reduction of health care costs associated with alcoholism treatment: a 14-year longitudinal study. *J Stud Alcohol.* 1992;53:293–302.
20. Fleming MF, Mundt MP, French MT, et al. Benefit-cost analysis of brief physician advice with problem drinkers in primary care settings. *Med Care.* 2000;38:7–18.
21. Fleming MF, Mundt MP, French MT, et al. Brief physician advice for problem drinkers: long-term efficacy and benefit-cost analysis. *Alcohol Clin Exp Res.* 2002;26:36–43.
22. Wutzke SE, Shiell A, Gomel MK, et al. Cost effectiveness of brief interventions for reducing alcohol consumption. *Soc Sci Med.* 2001;52:863–870.
23. Miller TR, Levy DT. Cost-outcome analysis in injury prevention and control: eighty-four recent estimates for the United States. *Med Care.* 2000;38:562–582.
24. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. *Med Care.* 2000;38:583–637.
25. Salomon JA, Murray CJ. Estimating health state valuations using a multiple-method protocol. In: Murray CJ, Salomon JA, Mathers CD, et al., eds. *Summary Measures of Population Health.* Geneva: World Health Organization; 2002:487–499.
26. Miller TR. *Databook on Nonfatal Injury: Incidence, Costs, and Consequences.* Urban Institute Press; 1995.
27. Miller TR. Valuing non-fatal quality of life losses with quality-adjusted life years: the health economist's meow. *Journal of Forensic Economics.* 2000;13:145–168.
28. *Diagnostic and Statistical Manual of Mental Disorders,* 4th ed. Washington, DC: American Psychiatric Association; 1994.
29. Lenert LA, Michelson D, Flowers C, et al. IMPACT: an object oriented graphical environment for construction of multimedia patient interviewing software. *Proc Annu Symp Comput Appl Med Care.* 1995:319–323.
30. Lenert LA. The reliability and internal consistency of an Internet-capable computer program for measuring utilities. *Qual Life Res.* 2000; 9:811–817.
31. Furlong W, Feeny D, Torrance G, et al. *Guide to Design and Development of Health-State Utility Instrumentation,* vol 90–9. Hamilton, Ontario: McMaster University; 1990.
32. Froberg DG, Kane RL. Methodology for measuring health-state preferences—I: measurement strategies. *J Clin Epidemiol.* 1989;42:345–354.
33. Froberg DG, Kane RL. Methodology for measuring health-state preferences—II: scaling methods. *J Clin Epidemiol.* 1989;42:459–471.
34. Katz JN, Chang LC, Sangha O, et al. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care.* 1996;34: 73–84.
35. Salyers MP, Bosworth HB, Swanson JW, et al. Reliability and validity of the SF-12 health survey among people with severe mental illness. *Med Care.* 2000;38:1141–1150.
36. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001;33:337–343.
37. Johnson JA, Coons SJ, Ergo A, et al. Valuation of EuroQOL (EQ-5D) health states in an adult US sample. *Pharmacoeconomics.* 1998;13:421–433.
38. Woloshin S, Schwartz LM, Moncur M, et al. Assessing values for health: numeracy matters. *Med Decis Making.* 2001;21:382–390.
39. Saunders JB, Aasland OG, Babor TF, et al. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption—II. *Addiction.* 1993;88:791–804.
40. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med.* 1997;16:2349–2380.
41. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics.* 1982;38:963–974.
42. Read JL, Quinn RJ, Berwick DM, et al. Preferences for health outcomes. Comparison of assessment methods. *Med Decis Making.* 1984;4: 315–329.
43. Stiggelbout AM, Kiebert GM, Kievit J, et al. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Med Decis Making.* 1994;14:82–90.
44. Murray CJ, Lopez AD. Evidence-based health policy—lessons from the Global Burden of Disease Study. *Science.* 1996;274:740–743.
45. Murray CJ, Lopez AD. Quantifying disability: data, methods and results. *Bull World Health Organ.* 1994;72:481–494.
46. Bury-Maynard D. *Developing a Utility Index for Substance Abuse* [Dissertation]. Advanced Studies in Social Welfare, Brandeis University; 1999.
47. Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the measurement of preferences for health states? *Med Decis Making.* 2001;21:329–334.
48. Von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press; 1944.
49. Froberg DG, Kane RL. Methodology for measuring health-state preferences—III: population and context effects. *J Clin Epidemiol.* 1989;42: 585–592.
50. Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life. *Med Care.* 2000;38(suppl):II 138–150.

Available online at www.sciencedirect.com
SCIENCE DIRECT
Drug and Alcohol Dependence 79 (2005) 251–256
DRUG and ALCOHOL DEPENDENCE
www.elsevier.com/locate/drugalcdep

Short communication

Alcohol use and HIV risk behaviors among HIV-infected hospitalized patients in St. Petersburg, Russia

Evgeny M. Krupitsky [a], Nicholas J. Horton [b], Emily C. Williams [c,e,f], Dmitri Lioznov [a], Maria Kuznetsova [a], Edwin Zvartau [a], Jeffrey H. Samet [c,d,*]

[a] *St. Petersburg Scientific-Research Center of Addictions and Psychopharmacology, St. Petersburg State Pavlov Medical University, St. Petersburg, 197089, Russia*
[b] *Smith College, Department of Mathematics, Northampton, MA 01063, USA*
[c] *Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA 02118, USA*
[d] *Department of Social and Behavioral Sciences, Boston University School of Public Health, 91 East Concord Street, Suite 200, Boston, MA 02118, USA*
[e] *Department of Medicine, University of Washington School of Medicine, Seattle, WA 98195, USA*
[f] *Health Services Research and Development and Center of Excellence in Substance Abuse Treatment and Education, VA Puget Sound Health Care System, Seattle, WA 98108, USA*

Received 14 May 2004; received in revised form 3 January 2005; accepted 3 January 2005

Abstract

Purpose: Russia has high per capita alcohol consumption and an injection-drug-use-driven HIV epidemic. However, the role of alcohol in the spread of HIV infection in Russia is largely unexplored. Thus, we assessed recent alcohol use and associated HIV risk behaviors among HIV-infected persons in St. Petersburg, Russia.
Methods: We recruited HIV-infected hospitalized patients from the Botkin Infectious Disease Hospital between June 2001 and March 2002. Interviewers assessed alcohol and drug use with the addiction severity index (ASI) and sex- and drug-risk behaviors with the risk assessment battery (RAB). Lifetime abuse or dependence diagnoses for alcohol and drugs were established by a physician with addiction medicine training.
Results: Among 201 subjects, diagnoses of abuse or dependence (AB/DEP) were common: 9% (19/201) had only alcohol AB/DEP; 39% (78/201) had alcohol and drug AB/DEP; 47% (95/201) had only drug AB/DEP; and 4% (9/201) had no diagnosis of alcohol or drug AB/DEP. Sex- and drug-risk behaviors varied significantly by substance use diagnosis. Subjects with any alcohol AB/DEP had higher sex-risk RAB scores than those with drug only AB/DEP (6.1 versus 3.9, $p < .0001$). Among subjects with any diagnosis of drug AB/DEP, having in addition an alcohol diagnosis was associated with unclean needle use in the last six months (33% (26/78) versus 21% (20/95), $p = 0.08$).
Conclusions: Lifetime alcohol diagnoses of abuse or dependence were present in nearly one-half of hospitalized HIV-infected patients in St. Petersburg, Russia and were associated with significantly higher sex-risk behaviors and borderline significantly higher drug-risk behaviors. As HIV infection spreads rapidly in Russia and Eastern Europe, these data support the need for HIV risk-reduction interventions in alcohol abusing populations and raise the potential of benefit by addressing alcohol use in HIV-infected populations.

Keywords: HIV risk; Alcohol abuse; Russia

1. Introduction

Over the past five years, reports of HIV infection in Russia have increased dramatically, with an estimated one million people already infected (Hamers and Downs, 2003). This represents an approximate 10-fold increase from 130,000 infections reported in 1999 (Grisin and Wallander, 2002; Stephenson, 2000). According to forecasts, there may be approximately five million HIV-infected individuals in the Russian Federation by 2007 (Anonymous, 2002). The majority of HIV infection in Russia is currently among injec-

* Corresponding author. Tel.: +1 617 414 7399; fax: +1 617 414 4676.
E-mail address: jsamet@bu.edu (J.H. Samet).

0376-8716/$ – see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.drugalcdep.2005.01.015

tion drug users (IDUs) (Dehne et al., 2000; Krupitsky et al., 2004).

Injection drug use is increasingly common in Russia. The number of drug-dependent persons rose 10-fold from 1986 to 1998 and exceeded two million people (Koshkina, 2000). However, even more common than drug use in Russia is alcohol consumption. In fact, in Russia, alcohol consumption per capita is among the highest in the world, particularly for men (Nemtsov, 2000). Average alcohol consumption for Russian men has increased from 29 g per day in 1992 to 45 g per day in 2002 (Zohoori et al., 2003). Estimates of lifetime prevalence of alcohol dependence in the former Soviet Union may be as high as 69% in men (Pakriev et al., 1998) versus 18% (Grant, 1997) in US men.

Alcohol use has been shown to impact HIV infection with increased transmission risk and possible disease progression. Several studies have demonstrated that people with heavy alcohol use tend to engage in riskier behaviors, such as sex with multiple partners, unprotected vaginal and anal intercourse, and injection drug use (Bagnall et al., 1990; Fenaughty and Fisher, 1998; Halpern-Felsher et al., 1996). In Asia, drinking alcohol is often associated with the high-risk behavior of engaging in sex, usually unprotected, with a commercial sex worker (Fordham, 1995; Gibney et al., 2003; Kim et al., 1998; MacQueen et al., 1996; Poudel et al., 2004; Wee et al., 2004).

In the United States, HIV infection has been examined in substance abuse treatment settings where its prevalence is increased among alcohol-dependent persons (Avins et al., 1994; Mahler et al., 1994). Alcohol abuse in combination with other drug use may lead to even more frequent occurrences of risky sex- and drug-use behaviors. Increased alcohol consumption is associated with sexual HIV risk-taking behavior among female drug users (Rees et al., 2001). Among IDUs, alcohol use is associated with increased sex-risk behavior (Stein et al., 2001). Specifically, among HIV-infected individuals with a history of alcohol problems, at-risk drinking was associated with inconsistent condom use among active IDUs (Ehrenstein et al., 2004). There are mixed results concerning the impact of alcohol use on risky injection drug behavior (Rees et al., 2001; Stein et al., 2000). Reasons underlying the association between alcohol use and high-risk behaviors for HIV have been described and include decreased inhibitions and risk perception (Cooper, 2002; Fromme et al., 1999), belief that alcohol enhances sexual arousal (George et al., 2000), deliberate use of alcohol to excuse high-risk behavior (Dermen et al., 1998), and the indirect association that bars are common places to meet potential sexual partners (Purcell et al., 2001).

Further, recent reports in animals raise the possibility that alcohol consumption plays a permissive role for HIV replication, possibly resulting in higher HIV viral loads which would be associated with higher transmission risk (Stoltz et al., 2002). Finally, among HIV-infected patients with a history of alcohol problems who are receiving antiretroviral treatment, alcohol consumption was associated with higher HIV viral loads and lower CD4 cell counts, markers of HIV disease progression (Samet et al., 2003).

In light of Russia's epidemic spread of HIV, high alcohol consumption, and ongoing injection drug use, we sought to clarify if alcohol use among HIV-infected Russians exacerbated unsafe sex- and drug-use behaviors. Thus, we examined HIV risk behaviors and alcohol and drug use among hospitalized HIV-infected persons in St. Petersburg, Russia.

2. Methods

2.1. Study design

For 201 HIV-infected inpatients at an infectious disease hospital in St. Petersburg, Russia, researchers administered a survey and abstracted data from medical records in order to assess drug and alcohol use and severity, HIV risk behaviors, lifetime substance abuse, and dependence diagnoses. This study was approved by the Institutional Review Boards of Boston Medical Center and St. Petersburg Pavlov State Medical University.

2.2. Data collection

Data were collected between June 2001 and March 2002 at the Botkin Infectious Disease (ID) Hospital, an inpatient facility founded in 1882 for adult patients with infectious diseases in St. Petersburg, Russia. The 1200-bed hospital, staffed primarily by ID specialists, is the largest inpatient facility of its kind in Northwest Russia and is divided into departments (e.g. HIV/AIDS, hepatitis, and food-borne diseases) consisting of 40–50 beds each. It also has a surgery and maternity ward serving infected persons. In the 1990s and early 2000s, Botkin Hospital was the only inpatient site for HIV-infected individuals in St. Petersburg. All individuals admitted to the three departments that care for HIV-infected patients were eligible and asked to join the study. Most of the patients from these departments were admitted for initial presentation or exacerbations of viral hepatitis; 84% (201/240) of the approached patients agreed to participate. Study participants provided written informed consent prior to data collection. An interviewer assessed subjects in person using a standardized instrument to ascertain information including the following: demographics, HIV risk behaviors, alcohol consumption, and addiction severity. Diagnoses of lifetime alcohol or drug abuse or dependence were made through clinical assessment. Laboratory tests performed as part of clinical care were recorded. We obtained available serology results, liver function tests, and history of disease from medical records. All other data collected on subjects were obtained specifically for research purposes.

2.3. Measures of substance use and substance abuse and dependence

Within the first week of inpatient stay, subjects were evaluated for lifetime alcohol or drug abuse or dependence. Clin-

ical diagnoses were made via assessment by an infectious disease specialist with training in addiction medicine (Maria Kuznetsova, MD) using criteria from the diagnostic and statistical manual of mental disorders–fourth edition (DSM-IV) (American Psychiatric Association, 1994). Additionally, interviewer assessment included standardized questions on alcohol and heroin use including the Michigan alcohol screening test (MAST)(Selzer, 1971), the time line follow back (TLFB) assessment (Sobell and Sobell, 1992), the addiction severity index (ASI) (McLellan et al., 1992), and the risk assessment battery (RAB) (Navaline et al., 1994), instruments with well-documented reliability and validity. For purposes of analysis, substance use diagnoses of abuse and dependence were combined resulting in four groups: alcohol only, alcohol and drug, drug only, and no diagnosis.

2.4. Primary outcome measure: HIV sex- and drug-risk behaviors

Estimates of sex- and drug-risk behaviors were derived from the risk assessment battery (RAB) (Navaline et al., 1994). This instrument sums the scores for individual items, for totals of 35 (sex-risk) and 25 (drug-risk). Scores are derived for both sex- and drug-risk behaviors via a series of questions including inquiries about numbers of sexual partners, usage of condoms, and sharing of needles.

Interviews were conducted in Russian. Standard survey elements already translated into Russian were used (i.e., MAST, ASI, TLFB, RAB). Other questions were translated from English to Russian, back-translated into English to check for accuracy, and then corrected.

2.5. Analysis

Fisher's exact (for categorical outcomes) and Kruskal–Wallis tests (for continuous outcomes) were used to compare subject substance use diagnosis with subject characteristics including risk behaviors for HIV infection. Reported p-values are two-tailed, and a p-value less than 0.05 was considered statistically significant. A multivariable linear regression was fit to predict RAB sex-risk subscale scores, as a function of diagnosis group, gender, and age. A logistic regression model, also controlling for gender and age, was fit to predict any needle sharing, for subjects with any drug diagnosis and injection use in the past six months. All analyses were carried out using SAS/STAT version 8.2 (SAS Institute, 2001).

3. Results

3.1. Subject characteristics

The characteristics of the 201 HIV-infected subjects are outlined in Table 1. Three-fifths (62%) were male, with a mean age of 27 years. The most common lifetime substance use diagnosis of abuse or dependence was drug only (47%), alcohol and drug (39%), and alcohol only (9%). Nine subjects (4%) had no substance use diagnosis. Assessments of past 30-day use of alcohol among those with any alcohol diagnosis ($n=97$) revealed a mean of 28 g/day, the equivalent of approximately 2.5 standard drinks/day.

3.2. Risk behaviors

Risk assessments yielded a sex-risk subscale mean score of 5.0 and a drug-risk subscale score of 4.3. More than half (103/201) of all subjects reported greater than two sexual partners in the past six months, and 66% (132/201) reported inconsistent condom use, with 23% (47/201) of all subjects reporting condom use none of the time. Forty-seven percent (94/201) of all subjects reported injecting drugs in the last six months with 23% (46/201) reporting using others' needles and 35% (71/201) reporting sharing their own needles.

Results from data stratified by lifetime substance use diagnosis are displayed in Table 2. Notable among the results are significant differences among substance abuse diagnostic groups for age ($p<0.0001$), current work ($p=0.002$) and antibodies to the hepatitis C virus ($p<0.0001$).

An unadjusted comparison between subjects with any diagnosis of alcohol abuse or dependence (i.e., alcohol only combined with alcohol and drug subjects) and those with only drug diagnoses reveals significantly higher sex-risk subscale scores (6.1 versus 3.9, $p<0.0001$).

Table 1
Demographic and substance abuse characteristics of HIV-infected persons in an infectious disease hospital in St. Petersburg, Russia ($n=201$)

Characteristic	Total cohort ($n=201$)
Female	76 (38%)
Age	26.6 (8.17)
Current work	34 (17%)
Hep C ($n=200$)	186 (93%)
Hep B ($n=123$)	58 (47%)
Grams ethanol/day ($n=98$)	28.19 (32.31)
MAST	6.58 (5.00)
Addiction severity index	
Medical status	0.72 (0.34)
Employment	0.71 (0.29)
Alcohol use	0.15 (0.22)
Drug use	0.10 (0.16)
Legal (law)	0.14 (0.27)
Family (social)	0.31 (0.28)
Psychiatric	0.56 (0.24)
Risk assessment battery	
Sex-risk subscale	4.98 (2.96)
Always condom/no sex	69 (34%)
Two or more partners	103 (51%)
Drug-risk subscale	4.31 (5.54)
Injected drugs[a]	94 (47%)
Used others' needles[a]	46 (23%)
Shared their needles	71 (35%)

[a] Refers to the 6 months prior to assessment.

Table 2
Demographic and substance abuse characteristics of HIV-infected persons in an infectious disease hospital in St. Petersburg, Russia stratified by alcohol and drug abuse or dependence diagnoses ($n = 201$)

Characteristic	Abuse or dependence diagnosis [n(%) or mean (S.D.)]				
	Alcohol only ($n = 19$)	Alcohol + drug ($n = 78$)	Drug only ($n = 95$)	No diagnosis ($n = 9$)	p-value
Female	7 (37%)	21 (27%)	43 (45%)	5 (56%)	0.05
Age	36 (10.9)	27.3 (6.9)	23.3 (6.0)	35.9 (9.1)	<0.0001
Current work	6 (32%)	15 (19%)	12 (13%)	1 (11%)	0.0016
Hep C ($n = 200$)	14 (74%)	73 (95%)	92 (97%)	6 (67%)	<0.0001
Hep B ($n = 123$)	7 (41%)	29 (60%)	20 (39%)	2 (29%)	0.13
Grams ethanol/day ($n = 98$)	43.85 (54.48)	24.68 (23.26)	2.37 (3.55)	2.92 (5.28)	<0.0001
MAST	11.79 (4.77)	9.54 (4.57)	3.45 (2.71)	3.11 (2.71)	<0.0001
Addiction severity index					
Medical status	0.82 (0.25)	0.71 (0.37)	0.69 (0.33)	0.76 (0.34)	0.29
Employment	0.86 (0.25)	0.68 (0.31)	0.70 (0.28)	0.77 (0.23)	0.07
Alcohol use	0.43 (0.32)	0.23 (0.22)	0.04 (0.06)	0.02 (0.04)	<0.0001
Drug use	0.00 (0.0)	0.09 (0.14)	0.13 (0.17)	0.00 (0.0)	<0.0001
Legal (law)	0.20 (0.34)	0.17 (0.30)	0.12 (0.24)	0.00 (0.0)	0.08
Family (social)	0.37 (0.27)	0.32 (0.30)	0.29 (0.27)	0.25 (0.29)	0.56
Psychiatric	0.68 (0.21)	0.58 (0.23)	0.51 (0.25)	0.59 (0.16)	0.03
Risk assessment battery					
Sex-risk subscale	5.95 (2.63)	6.17 (3.13)	3.86 (2.41)	4.89 (2.76)	<0.0001
Always condom/no sex	4 (21%)	23 (29%)	37 (39%)	4 (44%)	0.31
2 or more partners	13 (68%)	52 (67%)	35 (37%)	3 (33%)	0.0003
Drug-risk subscale	0.00 (0.0)	4.53 (5.68)	5.38 (5.67)	0.22 (0.67)	<0.0001
Injected drugs[a]	0 (0%)	39 (50%)	54 (57%)	1 (11%)	<0.0001
Used others' needles[a]	0 (0%)	26 (33%)	20 (21%)	0 (0%)	0.004
Shared their needles	0 (0%)	29 (37%)	42 (44%)	0(0%)	0.0003

[a] Refers to the 6 months prior to assessment.

In multivariable linear regression, females had borderline significantly higher sex-risk scores (predicted scores 0.76 units higher than men, $p = 0.07$), while age was not a significant predictor ($p = 0.78$). Diagnosis group was a significant predictor of RAB sex-risk scores ($F(3,195) = 11.36$, $p < 0.0001$). There was a significant difference between the predicted RAB sex-risk subscale score for subjects with any alcohol diagnosis compared to those with drug-only diagnoses ($F(1,195) = 22.1$, $p < .0001$).

Additionally, in an assessment of use of unclean needles, a comparison between drug users with alcohol diagnoses (26/78, 33%) and drug users without (20/95, 21%) indicated that more subjects with alcohol diagnoses used unclean needles in the last six months ($p = 0.08$). While controlling for sex, age, and diagnosis (drug and alcohol diagnosis versus drug-only diagnosis, 1 df), multivariate logistic regression indicated that younger age was a significant predictor of needle sharing (OR = 1.07, 95% CI = 1.01–1.15 per year, $p = 0.03$); gender was not a significant predictor (OR for females relative to males = 0.9, 95% CI = 0.4–1.9). Diagnosis of drug and alcohol versus drug-only diagnosis was associated with increased odds of any sharing (OR = 2.5, 95% CI = 1.2–5.1, $p = 0.02$).

3.3. Laboratory data

Among 123 subjects with recorded serology test results, 47% (58) were hepatitis B surface antigen positive. When stratified by lifetime substance use diagnoses of abuse or dependence, positive results for hepatitis B antigens were detected in 41% (7/17), 60% (29/48), 40% (20/50) and 29% (2/7) of the alcohol only, alcohol and drug, drug-only and no-diagnosis groups, respectively.

Among 200 subjects with available data, hepatitis C antibodies were detected in 93% (186). When stratified by substance use diagnoses, hepatitis C antibodies were detected in 74% (14/19), 95% (73/77), 97% (92/95) and 67% (6/9) of the alcohol-only, alcohol and drug, drug-only and no-diagnosis groups, respectively.

4. Discussion

Among hospitalized HIV-infected patients in Russia, reports of alcohol use and high-risk behaviors for HIV transmission are common. While a majority of subjects were diagnosed with drug abuse or dependence, almost half (48%) of all subjects had lifetime diagnoses of alcohol abuse or dependence. Further, a majority of the total population reported inconsistent condom use and/or having two or more sexual partners in the last six months. This extent of risky sex is particularly disturbing in that for women, sex-risk behavior may be more significant to HIV seroconversion than drug-risk behavior (Strathdee et al., 2001). Strathdee et al. (2001) also found that among female IDUs sex risks (e.g., recent STD and sex trade) were more commonly associated

with HIV seroconversion than drug-related risk behaviors. Forty-seven percent of all subjects reported injecting drugs in the last six months, and almost half of these reported using others' needles. The marked substance abuse and risk behaviors in this population of HIV-infected individuals are alarming.

The finding of an association between HIV risk behaviors and an alcohol abuse or dependence diagnosis is notable in that Russia's HIV epidemic has been nearly totally attributed to injection drug use (Dehne et al., 2000). We found that sex-risk was greater, as reflected in the RAB sex-risk subscores, with diagnoses of alcohol abuse or dependence. Also, an alcohol diagnosis was associated with increased odds of needle sharing. These findings are of particular importance given the fact that in Russia alcohol use is widespread, alcohol dependence is common, and HIV infection is epidemic. The small number of subjects diagnosed with "alcohol only" lifetime abuse or dependence is unique in two ways. First, these subjects were significantly older than the subjects with other substance use diagnoses. Second, co-infection with hepatitis C was common among a surprising number (74%) of alcohol-only subjects. In fact, co-infection with hepatitis B and, particularly C, was common in the entire cohort. This abnormal prevalence is likely a result of selection bias as reasons for hospitalization included acute hepatitis. Because the interview did not assess whether or not subjects ever used injection drugs, we cannot speculate on the means of transmission of hepatitis C to those with alcohol-only diagnoses. This may be a limitation of these data and implicate the need for further study of this population.

Our study has other noteworthy limitations. These data represent a cross-sectional perspective of the substance abuse and risk behaviors of a population of HIV-infected inpatients with substantial co-morbidity. Thus, inferences that may be drawn regarding the influence of alcohol on HIV risk behaviors in general populations are limited. However, as illness severity is typically greater in hospitalized patients, and sicker patients have been shown to have less drug- and sex-risk behaviors (Collins et al., 2001) use of such patients is likely to provide a conservative estimate of HIV risk. Additionally, because our population has already been infected with HIV, our data cannot address the relationship between HIV infection and use of alcohol or drugs. Further, the risk behaviors of those already infected with HIV may differ from those among non-infected individuals in Russia.

Despite these limitations, our data show that lifetime diagnoses of alcohol abuse or dependence were present in nearly one-half of hospitalized HIV-infected patients in St. Petersburg, Russia and were associated with significantly higher sex-risk behaviors and a trend toward higher drug-risk behaviors. As HIV spreads rapidly in Russia and Eastern Europe, addressing alcohol use in HIV-infected persons holds potential to decrease the transmission of HIV by lowering the prevalence of high sex- and drug-risk behaviors.

Acknowledgements

We gratefully acknowledge the intellectual contributions of Seville Meli, MPH and Naomi Freedner, MPH. This project was supported by a National Institute on Alcohol Abuse and Alcoholism Supplement to R01-AA11785.

References

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. American Psychiatric Association, Washington D.C.

Anonymous 2002. Report of the Epidemiological Service of the Ministry of Health Care of the Russian Federation.

Avins, A.L., Woods, W.J., Lindan, C.P., Hudes, E.S., Clark, W., Hulley, S.B., 1994. HIV infection and risk behaviors among heterosexuals in alcohol treatment programs. JAMA 271, 515–518.

Bagnall, G., Plant, M., Warwick, W., 1990. Alcohol, drugs and AIDS-related risks: results from a prospective study. AIDS Care 2, 309–317.

Collins, R.L., Kanouse, D.E., Gifford, A.L., Senterfitt, J.W., Schuster, M.A., McCaffrey, D.F., Shapiro, M.F., Wenger, N.S., 2001. Changes in health-promoting behavior following diagnosis with HIV: prevalence and correlates in a national probability sample. Health Psychol. 20, 351–360.

Cooper, M.L., 2002. Alcohol use and risky sexual behavior among college students and youth: evaluating the evidence. J. Stud. Alcohol Suppl., 101–117.

Dehne, K.L., Pokrovskiy, V., Kobyshcha, Y., Schwartlander, B., 2000. Update on the epidemics of HIV and other sexually transmitted infections in the newly independent states of the former Soviet Union. AIDS 14 (Suppl. 3), 75–84.

Dermen, K.H., Cooper, M.L., Agocha, V.B., 1998. Sex-related alcohol expectancies as moderators of the relationship between alcohol use and risky sex in adolescents. J. Stud. Alcohol 59, 71–77.

Ehrenstein, V., Horton, N.J., Samet, J.H., 2004. Inconsistent condom use among HIV-infected patients with alcohol problems. Drug Alcohol Depend. 73, 159–166.

Fenaughty, A.M., Fisher, D.G., 1998. High-risk sexual behavior among drug users: the utility of a typology of alcohol variables. Sex. Transm. Dis. 25, 38–43.

Fordham, G., 1995. Whisky, women and song: men, alcohol and AIDS in northern Thailand. Aust. J. Anthropol. 6, 154–177.

Fromme, K., D'Amico, E.J., Katz, E.C., 1999. Intoxicated sexual risk taking: an expectancy or cognitive impairment explanation? J. Stud. Alcohol 60, 54–63.

George, W.H., Stoner, S.A., Norris, J., Lopez, P.A., Lehman, G.L., 2000. Alcohol expectancies and sexuality: a self-fulfilling prophecy analysis of dyadic perceptions and behavior. J. Stud. Alcohol 61, 168–176.

Gibney, L., Saquib, N., Metzger, J., 2003. Behavioral risk factors for STD/HIV transmission in Bangladesh's trucking industry. Soc. Sci. Med. 56, 1411–1424.

Grant, B.F., 1997. Prevalence and correlates of alcohol use and DSM-IV alcohol dependence in the United States: results of the National Longitudinal Alcohol Epidemiologic Survey. J. Stud. Alcohol 58, 464–473.

Grisin, S., Wallander, C. 2002. Russia's HIV/AIDS crisis: confronting the present and facing the future. CSIS HIV/AIDS Task Force.

Halpern-Felsher, B.L., Millstein, S.G., Ellen, J.M., 1996. Relationship of alcohol use and risky sexual behavior: a review and analysis of findings. J. Adolesc. Health. 19, 331–336.

Hamers, F.F., Downs, A.M., 2000. HIV in central and eastern Europe. Lancet 361, 1035–1044.

Kim, J., Celentano, D.D., Crum, R.M., 1998. Alcohol consumption and sexually transmitted disease risk behavior: partner mix among male Korean university students. Alcohol. Clin. Exp. Res. 22, 126–131.

Koshkina, E.A., 2000. The prevalence of the use of narcotics and other psychoactive substances in Russia today. Zh. Mikrobiol. Epidemiol. Immunobiol., 15–19.

Krupitsky, E., Zvartau, E., Karandashova, G., Horton, N.J., Schoolwerth, K.R., Bryant, K., Samet, J.H., 2004. The onset of HIV infection in the Leningrad region of Russia: a focus on drug and alcohol dependence. HIV Med. 5, 30–33.

MacQueen, K.M., Nopkesorn, T., Sweat, M.D., Sawaengdee, Y., Mastro, T.D., Weniger, B.G., 1996. Alcohol consumption, brothel attendance, and condom use: normative expectations among Thai military conscripts. Med. Anthropol. Q. 10, 402–423.

Mahler, J., Yi, D., Sacks, M., Dermatis, H., Stebinger, A., Card, C., Perry, S., 1994. Undetected HIV infection among patients admitted to an alcohol rehabilitation unit. Am. J. Psychiatry 151, 439–440.

McLellan, A., 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.

Navaline, H.A., Snider, E.C., Petro, C.J., Tobin, D., Metzger, D., Alterman, A.I., Woody, G.E., 1994. Preparations for AIDS vaccine trials. An automated version of the risk assessment battery (RAB): enhancing the assessment of risk behaviors. AIDS Res. Hum. Retroviruses 10 (Suppl. 2), 281–283.

Nemtsov, A.V., 2000. Estimates of total alcohol consumption in Russia, 1980–1994. Drug Alcohol Depend. 58, 133–142.

Pakriev, S., Vasar, V., Aluoja, A., Shlik, J., 1998. Prevalence of ICD-10 harmful use of alcohol and alcohol dependence among the rural population in Udmurtia. Alcohol Alcohol 33, 255–264.

Poudel, K.C., Jimba, M., Okumura, J., Joshi, A.B., Wakai, S., 2004. Migrants' risky sexual behaviours in India and at home in far western Nepal. Trop. Med Int. Health. 9, 897–903.

Purcell, D.W., Parsons, J.T., Halkitis, P.N., Mizuno, Y., Woods, W.J., 2001. Substance use and sexual transmission risk behavior of HIV-positive men who have sex with men. J. Subst. Abuse. 13, 185–200.

Rees, V., Saitz, R., Horton, N.J., Samet, J., 2001. Association of alcohol consumption with HIV sex- and drug-risk behaviors among drug users. J. Subst. Abuse Treat. 21, 129–134.

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? Alcohol Clin. Exp. Res. 27, 862–867.

SAS Institute Inc., 2001. SAS/STAT® Software: Changes and Enhancements, Release 8.2. SAS Institute Inc., Cary, NC.

Selzer, M.L., 1971. The Michigan alcoholism screening test: the quest for a new diagnostic instrument. Am. J. Psychiatry 127, 1653–1658.

Sobell, L.C., Sobell, M.B., 1992. Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: Litten, R., Allen, J. (Eds.), Measuring Alcohol Consumption. The Human Press Inc.

Stein, M.D., Anderson, B., Charuvastra, A., Friedmann, P.D., 2001. Alcohol use and sexual risk taking among hazardously drinking drug injectors who attend needle exchange. Alcohol Clin. Exp. Res. 25, 1487–1493.

Stein, M.D., Hanna, L., Natarajan, R., Clarke, J., Marisi, M., Sobota, M., Rich, J., 2000. Alcohol use patterns predict high-risk HIV behaviors among active injection drug users. J. Subst. Abuse Treat. 18, 359–363.

Stephenson, J., 2000. HIV/AIDS surging in Eastern Europe. JAMA 284, 3113–3114.

Stoltz, D.A., Nelson, S., Kolls, J.K., Zhang, P., Bohm, R.P., Murphey-Corb, M., Bagby, G.J., 2002. 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. 26, 527–534.

Strathdee, S.A., Galai, N., Safaiean, M., Celentano, D.D., Vlahov, D., Johnson, L., Nelson, K.E., 2001. Sex differences in risk factors for HIV seroconversion among injection drug users: a 10-year perspective. Arch. Intern Med. 161, 1281–1288.

Wee, S., Barrett, M.E., Lian, W.M., Jayabaskar, T., Chan, K.W., 2004. Determinants of inconsistent condom use with female sex workers among men attending the STD clinic in Singapore. Sex Transm. Infect. 80, 310–314.

Zohoori, N., Blanchette, D., Popkin, B., 2003. Monitoring health conditions in the Russian Federation: The Russian Longitudinal Monitoring Survey 1992–2002. Report submitted to the U.S. Agency for International Development. Carolina Population Center, University of North Carolina at Chapel Hill, NC.

AIDS Care, July 2005; 17(5): 566–578

Routledge
Taylor & Francis Group

Physical and sexual violence and health care utilization in HIV-infected persons with alcohol problems

J. M. LIEBSCHUTZ[1], J. L. GEIER[1], N. J. HORTON[2], C. H. CHUANG[1], & J. H. SAMET[1,3]

[1]*Clinical Addictions Research and Education Unit, Section of General Internal Medicine, Boston University School of Medicine,* [2]*Department of Mathematics, Smith College, Northampton, MA, and* [3]*Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, USA*

Abstract

We examined interpersonal violence and its association with health care utilization and substance use severity among a cohort of 349 HIV-infected men and women with histories of alcohol problems assessed biannually up to 36 months. Data included demographics, lifetime interpersonal violence histories, age at first violence exposure, recent violence (prior six months), substance use severity and health care utilization (ambulatory visits, Emergency Department (ED) visits, hospitalizations) and adherence to HIV medication. Kaplan-Meier survival curves estimated the proportion of subjects experiencing recent violence. Generalized estimating equation regression models evaluated the relationship between recent violence, utilization and substance use severity over time, controlling for demographics, CD4 counts and depressive symptoms. Subject characteristics included: 79% male; mean age 41 years; 44% black, 33% white and 23% other. Eighty percent of subjects reported lifetime interpersonal violence: 40% physical violence alone, and 40% sexual violence with or without physical violence. First violence occurred prior to age 13 in 46%. Twenty-four (41%) of subjects reported recent violence by 24 and 36 months, respectively. In multivariate analyses, recent violence was associated with more ambulatory visits, ED visits and hospitalizations and worse substance use severity, but not medication adherence. Due to the high incidence and associated increased health care services utilization, violence prevention interventions should be considered for HIV-infected patients with a history of alcohol problems.

Introduction

Interpersonal violence is intentional use of force (or threat of force), and includes both physical and sexual violence. It is experienced by a majority of people in the United States at some point in their lives. A survey of 16,000 men and women found that 52% of women and 66% of men had experienced physical assault at some point in their lives, while 18% of women and 3% of men had experienced rape (Tjaden & Thoennes, 2000). Studies of the prevalence of childhood sexual violence, including national and local probability samples,

Correspondence: Jane Liebschutz, MD MPH, Boston Medical Center, 91 East Concord Street, Suite 200 Boston, MA 02118, USA. Tel: +1 (617) 414 7399. Fax: +1 (617) 414 4676. E-mail: jliebs@bu.edu

ISSN 0954-0121 print/ISSN 1360-0451 online
DOI: 10.1080/09540120512331314358

have shown that approximately 25% of adult women and 16% of adult men report childhood violence (Johnsen & Harlow, 1996).

Among a variety of populations studied, individuals who report lifetime histories of interpersonal violence are more likely to report a number of associated findings compared to those who do not report such histories. These findings include: increased somatic symptoms, increased medical illnesses, worse self-rated health status, more sexually transmitted diseases, more substance use and abuse, increased mental illness, increased health care utilization (ED visits and hospitalizations) and more subsequent violence (Bergman et al., 1992; Coker et al., 2000, 2002; Eisenman et al., 2003; Felitti et al., 1998; Frayne et al., 1999; Johnsen & Harlow, 1996; Kilpatrick et al., 1997; Koss et al., 1991; Liebschutz et al., 2002; Liebschutz et al., 1997, 2000, 2003; Windle, 1994). Although most of these studies focused on female victims of interpersonal violence, a few extended these findings to men (Clark et al., 2001; Eisenman et al., 2003; Liebschutz et al., 2002). A consistent feature of most of these studies is the cross-sectional association of lifetime history of interpersonal violence with worse consequences, physical health, mental health and risky behaviors (Felitti et al., 1998; Liebschutz et al., 2002; McCauley et al., 1998; Plichta, 1992; Zierler et al., 1991). Only one longitudinal study of 3,006 women assessed over two years found that exposure to violence during this period was associated with an increase in substance use and abuse (Kilpatrick et al., 1997). In addition, substance abuse and high risk sexual behaviors confer an increased risk of experiencing subsequent interpersonal violence (Clark et al., 2001; Kilpatrick et al., 1997; Liebschutz et al., 2002). HIV-infected individuals with drug abuse and risky sexual behaviors are at particularly high risk for having experienced violence. The prevalence and consequences of violence exposure have been studied most carefully among HIV-infected women and less so among HIV-infected men. Among HIV-infected women, 66–68% report physical and 32–46% sexual violence as adults; 41% report physical and 31–41% sexual violence as children; and 21–28% report interpersonal violence in the prior year (Cohen et al., 2000; Gielen et al., 2000; Kimerling et al., 1999; Morrill et al., 2001; Vlahov et al., 1998). Few studies examine exposure to violence among HIV-infected men, although studies of men at high risk for HIV (men who have sex with men (MSM) and injection drug users) demonstrate high violence exposure of all types: childhood, physical and sexual violence (Jinich et al., 1998; Liebschutz et al., 2002; Zierler et al., 1991, 2000). Among a nationally representative probability sample of 2,864 HIV-infected adults, 20.5% of the women, 11.5% of the MSM and 7.5% of the heterosexual men reported physical harm by a partner or someone close to them since the HIV diagnosis, a period of one to six years in the majority of the sample (Zierler et al., 2000).

The thrust of the research on violence among HIV-infected persons has been the relationship of violence to risk transmission. The impact of violence exposure on physical health, mental health, substance abuse and health care utilization in this population has received less attention (Eisenman et al., 2003; Liebschutz et al., 2000). Studies have been limited by their cross-sectional design to examine the impact of recent violence compared to childhood or lifetime violence. Prospectively collected data on violence incidence in an HIV-infected population, male or female, has not been published. Existing studies utilizing cross-sectional violence exposure have not examined the effect of lifetime or recent violence exposure on HIV-medication adherence. Because HIV-infected individuals often have multiple comorbidities, it is difficult to attribute any observed associations to the violence or the comorbid problems. If recent violence is shown to be important in its relationship to

comorbidities, interventions to prevent violence might help improve quality and appropriateness of medical care.

To address these questions, we examined interpersonal violence in a longitudinal cohort of HIV-infected men and women with a history of alcohol problems and the association of violence exposure with health care utilization, substance abuse severity and adherence to HIV medications. We report data on males separately because of lack of such published data. We define interpersonal violence broadly: physical and/or sexual violence perpetrated by strangers, acquaintances and intimates at any age.

We hypothesized that exposure to interpersonal violence in this cohort would be high, and that those who reported prior interpersonal violence, childhood violence and recent violence would have higher utilization, worse substance abuse severity and lower HIV medication adherence than those who did not report these exposures.

Methods

Subjects

The HIV-ALC (HIV-Alcohol Longitudinal Cohort) study recruited HIV-infected individuals with a history of alcohol problems with the primary aim of evaluating the effect of alcohol use on HIV disease. A randomized controlled trial was conducted for 151 members of this cohort to test a behavioral intervention to improve adherence to HIV medication (Samet et al., 2002). With the approval of the Institutional Reviews Boards of Boston Medical Center and Beth Israel Deaconess Medical Center, individuals were asked to participate if they had a lifetime history of alcohol problems, (defined as two or more positive responses to the CAGE questionnaire (Ewing, 1984; Samet et al., 2004) and were HIV-infected. Those patients recruited from the Boston Medical Center HIV Diagnostic Evaluation Unit (Samet et al., 1995) who did not meet CAGE criteria were eligible if one of the two regular attending physicians made a clinical diagnosis of alcohol abuse or dependence. Other eligibility requirements included fluency in English or Spanish, Mini-Mental State Examination score greater or equal to 21 (Folstein et al., 1975) and no plans to move from the Boston area in the next two years.

Multiple methods of recruitment were utilized during the data collection period from July 1997 through July 2001. Primary enrollment, 56% of subjects, was from the HIV Diagnostic Evaluation Unit, an intake clinic for HIV-infected patients initiating medical care. Additional subjects were recruited as follows: 17% from posted flyers at homeless shelters and HIV/AIDS social service agencies in the Boston area, 13% from Boston Medical Center's Primary Care Practices, 5% from a respite facility for homeless persons, 4% from a methadone clinic, 4% from referrals by friends and 2% from the Beth Israel Deaconess Medical Center.

Data collection

After obtaining informed consent, a research associate interviewed subjects using a standardized instrument to ascertain baseline information including the following: demographics, exposure to interpersonal violence, alcohol and drug use, health care utilization in the preceding six months, adherence to HIV medications and depressive symptoms. Subsequent research interviews were attempted at six-month intervals through July 2001, a maximum period of 30 months. We attempted to obtain CD4 cell counts and HIV RNA levels on all subjects. Laboratory tests performed within six months of the

interview as part of clinical care were recorded. If not available through routine clinical care, blood samples were obtained and tested for CD4 cell count and HIV RNA using the Boston Medical Center Clinical Laboratory. Interviews were conducted both in English and in Spanish. The Spanish interview utilized standardized scales when available; the remainder of the interview was translated from English into Spanish, back-translated to assure accuracy, and then corrected.

Key variables

Interpersonal violence. For physical violence, the subjects were asked, 'Have you ever been physically abused or assaulted ... (for example: kicked, hit, choked, shot, stabbed, burned or held at gunpoint)?' 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 a violence history was reported, subjects were asked their age at the time of the first violence and whether any violence occurred within the past six months. The questions about violence in the past six months were asked at each subsequent research interview to all subjects.

Three variables were created to describe interpersonal violence: lifetime violence, recent violence and childhood sexual violence. Recent violence referred to violence occurring in the past six months. For lifetime violence and recent violence, three mutually exclusive categories were defined: no violence, physical violence only (without sexual violence) and sexual violence (with or without physical violence). Childhood sexual violence was defined as sexual violence occurring prior to age 13.

Outcome variables

Health care utilization. Three variables were used to define health care utilization in the prior six months: the number of ED visits, the number of hospitalizations and the number of ambulatory care visits.

Alcohol and drug severity. Alcohol and drug use severity were assessed using the Addiction Severity Index (ASI), an assessment instrument with well-documented reliability and validity, scored 0–1, with higher scores indicating increased severity (McLellan et al., 1985).

Medication adherence. Adherence was determined with the AIDS Clinical Trials Group (ACTG) Questionnaire for Adherence to Anti-Retroviral Medications(Chesney et al., 2000). Subjects reported the names of their antiretroviral medications, as well as the number of doses and the total number of pills prescribed daily. The three-day self-reported number of pills missed was computed for each HIV medication and the outcome was dichotomized (100% adherent versus less than fully adherent). The 30-day self-reported number of pills missed was also dichotomized ($\geq$95% adherent versus $\leq$95% adherent).

Other independent variables. Additional variables were included in our analysis including: gender, ethnicity (black, white, Hispanic or other), homelessness, CD4 count, any

use of drugs or alcohol in the past 30 days, depressive symptoms, participation in the randomized controlled trial and medication status. Homelessness was defined as having spent at least one night either on the street or in a shelter in the six months preceding the interview (Kertesz et al., 2003). Depressive symptoms were measured by the Center for Epidemiological Studies-Depression Scale (CES-D; Radloff, 1977), a continuous scale to measure depressive symptoms. The question for HIV medication status was, 'Are you currently taking any HIV medications (such as AZT or protease inhibitors)?' Health insurance status was measured, but 99% of all subjects had access to private, Medicaid or a special publicly-funded health insurance for HIV-infected individuals such that all had coverage for medications, ambulatory visits and hospitalizations.

Data analysis

Descriptive statistics were used to characterize the study population. We used the Kaplan-Meier survival estimator to calculate the proportion of subjects reporting recent violence over time. Since recent violence was defined as any physical or sexual violence in the six months prior to each interview, time 0 was considered to be six months prior to the first interview. Thus, although there were 30 months of follow-up, the survival estimator calculated 36 months of possible exposure to violence. Generalized estimating equation regression models were used to evaluate the relationship between recent violence and ambulatory visits, ED visits, hospitalizations, adherence to HIV medication and substance abuse severity over time while controlling for age, gender, education, race, childhood violence, recent substance use, medication status, CD4 count, participation in the randomized control trial for adherence, depressive symptoms and homelessness.

Models substituting lifetime violence (sexual and physical) for childhood violence exposure were also conducted. The results were similar in direction and magnitude, and are not reported separately. The logit link (Poisson regression) was used for the number of ambulatory visits, ED visits and hospitalizations per six-month period. The logit link was used for the dichotomous adherence outcomes and the identity link (linear regression) for the substance use severity indices. Only the 250 subjects taking HIV medication were used in the calculations involving adherence measures. An identity working correlation matrix was specified and empirical standard error estimates reported.

Stratified analyses by gender were performed. They showed no differences in the directions of the findings, although the statistical significance of the results were attenuated by the lower numbers in the separate gender groups. Because recent violence was a low frequency event, the female models evaluating effect of recent violence on hospitalizations, ED visits, substance abuse severity and adherence did not include enough observations to be conducted separately. Only the total group and male model results are reported.

Results

Subject characteristics

Characteristics of the 349 subjects in this cohort include the following, as outlined in Table I: mean age 41 years old; male 79%; ethnic minorities 77%; high school graduates 60%; and homeless 29%. Fifty-nine percent of the sample reported a history of injection

Table I. Subject characteristics at baseline*.

	Total $N=349$	Males $n=276$ (79%)	Females $n=73$ (21%)
Age mean years (SD)	40.6 (7.3)	41 (7.3)	38.8 (7.3)
Race			
Black	154 (44)	116 (42)	38 (52)
White	116 (33)	94 (34)	22 (30)
Hispanic	75(21)	63(23)	12(16)
Other	4(1)	3(1)	1(1)
High school graduate	210 (60.2)	164 (59.4)	46 (63)
Homelessness	101 (29)	88 (32)	13 (18)
HIV infection risk factor			
Injection drug use	205 (59)	166 (60)	39 (54)
MSM	65 (19)	65 (24)	
Heterosexual /blood	78 (24)	45 (16)	33 (46)
Alcohol use			
Abstinent	201 (58)	153 (55)	48 (66)
Moderate	82 (24)	69 (25)	13 (18)
Hazardous	66 (19)	54 (20)	12 (16)
CD4 count mean, mm^3 (SD)	401 (278)	390 (261)	444 (335)

*n (%) unless noted otherwise.

drug use, while 19% reported MSM and 22% reported heterosexual contact as their mode of HIV transmission. Alcohol and drug use in the last 30 days was reported by 47% of the subjects, while 19% reported hazardous drinking as defined by the National Institute on Alcohol Abuse and Alcoholism guidelines (National Institute on Alcohol Abuse and Alcoholism, 1995). Of the 349 subjects, 219 (63%) had at least one follow-up interview after baseline, with an average of three follow-up visits per subject. Previous work reported that time of recruitment into this study was the most important predictor of the number of completed visits ($p < 0.0001$) with subjects entering at the beginning of the recruitment completing, on average, three interviews more than subjects entering at the end of the recruitment (Ehrenstein et al., 2004).

Interpersonal violence prevalence

A lifetime history of interpersonal violence was exceedingly common, with 80% reporting such histories, half of those with physical violence only and half with sexual violence with or without physical violence. Forty-six percent reported occurrence of first physical or sexual violence before age 13, 17% between 13 and 17 and 17% after 17 (Table II). In 885 follow-up interviews, 9.7% reported violence in the prior six months. Survival analyses show that 24% (95% confidence intervals 21, 27) and 41% (95% CI 36, 46) of the sample reported at least one episode of violence by 24 and 36 months, respectively. When including only those subjects who reported a lifetime history of violence at baseline, 45% (95% CI 40, 50) reported at least one episode of subsequent violence by 36 months. Among men, 36% (95% CI 31, 41) reported at least one episode of recent violence by 36 months. Among women, 54% (95% CI 44, 64) reported at least one episode of recent violence by 36 months.

Table II. Exposure to interpersonal violence: baseline characteristics.

	Total $N=349$	Male $n=276$	Female $n=73$
Lifetime exposure			
None	67 (20)	58(22)	9 (12)
Physical only	139 (40)	128 (47)	11 (15)
Sexual +/− Physical	139 (40)	86 (32)	53 (73)
Age at first exposure			
Never	67 (20)	58 (22)	9 (12)
<13 years	156 (46)	120 (44)	36 (50)
13–17 yrs	59 (17)	41 (15)	18 (25)
18+years	60 (17)	51 (19)	9 (13)

*Numbers may not add up to 100% due to rounding.

Association of violence exposure with health care utilization

Among 885 follow-up interviews, the median number of ambulatory visits, ED visits and hospitalizations in the prior six months was four (range =0–48), 0 (range =0–15), and 0 (range =0–10), respectively, with 75% quartile numbers of six ambulatory visits, one ED visit and no hospitalizations. Lifetime and childhood violence were not associated with any difference in ambulatory visits, ED visits or hospitalizations (all p values >0.10). Recent violence (when controlling for lifetime and childhood violence) was independently associated with 1.45 more ambulatory visits ($p=0.01$), 1.65 more ED visits ($p=0.003$) and 1.65 more hospitalizations ($p=0.03$) in the prior six months. For the male subjects, recent violence was associated with 1.61 more ambulatory visits ($p=0.01$), 1.35 more ED visits ($p=0.15$) and 1.27 more hospitalizations ($p=0.34$) in the prior six months (see Table III). Other factors associated with increased ambulatory visits were female gender ($p=0.02$), younger age ($p=0.02$), white race ($p=0.03$), abstinence from drugs or alcohol ($p=0.01$), more depressive symptoms ($p=0.05$) and lower CD4 count ($p=0.04$). Other factors significantly associated with increased hospitalizations were black ethnicity ($p=0.004$) and lower CD4 count ($p=0.03$). The only other factor besides recent violence significantly associated with increased ED visits was increased depressive symptoms ($p=0.0001$).

Table III. Associations of recent physical and sexual violence with health care utilization and substance use severity[a].

	Total sample	Males only
Ambulatory visits[b]	1.45 (1.08, 1.95)	1.61 (1.12, 2.32)
ED visits[b]	1.65 (1.18, 2.30)	1.35 (0.89, 2.06)
Hospitalizations[b]	1.65 (1.03, 2.45)	1.27 (0.77, 2.11)
ASI Alcohol[c]	0.08 (0.03, 0.14)	0.11 (0.04, 0.17)
ASI Drug[c]	0.02 (−0.01, 0.04)	0.02 (−0.01, 0.05)

[a]From longitudinal Poisson regression models controlling for childhood violence, substance abuse, mental health symptoms, age, high school graduate, race, housing status, gender, and currently taking medication; [b]IRR is the estimated rate of utilization of those reporting recent violence versus those not reporting recent violence (95% CI); [c]differences in mean ASI scores comparing those reporting recent violence versus those not reporting recent violence. (95% CI).

Association of violence with substance use severity

In 888 follow-up interviews, the mean ASI alcohol score was 0.18 (SD = 0.21) and mean ASI-drug score was 0.11 (SD = 0.11). Neither lifetime violence nor childhood violence was associated with worse alcohol or drug use severity as measured by the ASI. Recent violence was associated with 0.08 higher scores on ASI-alcohol (worse alcohol use severity) ($p = 0.004$) and 0.02 higher scores on ASI-drug ($p = 0.08$). For the male analyses, recent violence was associated with 0.11 higher scores on ASI-alcohol (worse alcohol use severity) ($p = 0.001$) and 0.02 higher scores on ASI-drug ($p = 0.30$). Younger age ($p = 0.001$), black ethnicity ($p = 0.004$) and homelessness ($p = 0.04$) were associated with worse alcohol use severity. Black ethnicity ($p = 0.02$) and increased depressive symptoms ($p < 0.0001$) were associated with worse drug use severity.

Association of violence with adherence to HIV medication

Lifetime, childhood and recent violence were not significantly associated with any change in adherence to HIV medications in any of the models tested (three-day and 30-day, males only), all p values > 0.10. In both models, only increased depressive symptoms ($p < 0.003$) and use of alcohol or drugs in the last 30 days ($p = 0.001$) were significantly associated with decreased adherence to HIV medication.

Discussion

The overwhelming majority of this cohort of HIV-infected men and women with a history of alcohol problems experienced interpersonal violence. Half of those abused reported sexual violence. Close to half the sample experienced their first violence in childhood. Over the three-year follow-up, more than a third of men and half of women were exposed to subsequent violence. While the high level of violence experienced by women was expected, the men in this study experienced almost equally high levels of violence. These high exposures to violence have been suggested in other studies on HIV-infected persons (Gruskin et al., 2002; Zierler et al., 2000), although they were cross-sectional studies asking about recent violence and may not be as reliable in establishing incidence of violence exposure as was this prospective data collection. Our study's finding of such high violence exposure might, in part, be accounted for by the broader definition of violence (both known and unknown perpetrators) and by the eligibility criteria of past alcohol problems and high prevalence of prior drug abuse, a known risk factor for experiencing subsequent violence (Kilpatrick et al., 1997).

Given that violence was experienced by the majority of subjects, the consistent and significant association of recent violence with increased health utilization was impressive even after controlling for prior violence exposure. The association of increased health care utilization and interpersonal violence exposure is well documented among HIV-infected and non-HIV-infected women (Bergman et al., 1992; Eisenman et al., 2003; Frayne et al., 1999; Koss et al., 1991; Liebschutz et al., 2000; Plichta, 1992). Only one study evaluated the effect of intimate partner violence among HIV-infected men (Eisenman et al., 2003) and found that gay/bisexual men with violence victimization had increased ED visits (OR 1.74, 95% CI 1.20–2.52), whereas heterosexual men with violence victimization had increased outpatient mental health visits (OR 2.23, 95% CI 1.07–4.64) and a trend toward increased hospitalizations (OR 2.74, 95% CI 0.96–7.85). Among HIV-infected women in that study, only outpatient mental health visits were associated with violence victimization.

That study was limited in its assessment of violence (physically hurt by partner or someone close to them since diagnosis of HIV), which might explain the lack of association of violence exposure to utlization outcomes.

No study has been designed to look at causality of this association. One potential explanation for the increased utilization, that violent injuries account for the utilization, has not been shown in previous studies (Bergman et al., 1992; Koss et al., 1991; Liebschutz et al., 2000; Rosenberg et al., 2000; Schnurr et al., 2000). Furthermore, this explanation would not account for an increase in ambulatory visits. A second hypothesis, that violence exposure affects physical health which leads to more need for medical care, has been suggested by previous study findings. The bulk of the studies on this topic have looked at somatic symptoms, chronic pain and self-reported health status (Coker et al., 2000; Drossman et al., 1990, 1995; Frayne et al., 1999; Liebschutz et al., 1997, 2000; McCauley et al., 1998). A few studies have shown an association with physical illnesses such as pneumonias, or coronary events that would require ED visits or hospitalizations (Coker et al., 2000, 2002; Cokkinides et al., 1999; Frayne et al., 1999; Liebschutz et al., 1997, 2000). All studies were limited by use of self-report data or small sample sizes and most focused on women. Coker and colleagues found a 1.5–1.6 fold increase in chronic diseases (hypertension, heart disease, arthritis, obstructive pulmonary disease and cancer) among men and women, respectively, who reported a history of physical violence as part of a large national survey on interpersonal violence (Coker et al., 2002). In multiple logistic regression analyses controlling for severity of violence exposure, they found that severe psychological violence, not physical, was associated with a 1.6 fold increased risk of developing a chronic disease in women only. In a study using medical record reviews of 50 HIV-infected women, exposure to interpersonal violence was associated with more infectious illnesses and health care utilization, even while controlling for CD4 count (Liebschutz et al., 2000). A third possible explanation, increased mental illness leading to increased utilization, is theoretically supported by many studies. These show that violence exposure leads to worse mental health (Felitti et al., 1998; Goodman et al., 1993; Mullen et al., 1988; Resnick et al., 1997; Schelling et al., 1998). As well as studies showing an increase in health care utilization associated with Post-traumatic Stress Disorder, the mental illness most closely identified with interpersonal violence exposure (Calhoun et al., 2002; Marshall et al., 1997, 1998; Schelling et al., 1998). All utilization studies looked at lifetime or childhood exposure to interpersonal violence and not recent violence. The only study suggesting an association with recent interpersonal violence did not measure physical or sexual violence exposure, per se, but marital stress among cohabitating Swedish women with coronary heart disease. It showed a threefold increased risk of recurrent coronary events among those reporting marital stress (Orth-Gomer et al., 2000). The association between violence and increased health care utilization found in this and other studies warrants further examination through epidemiological and biological methods. In particular, studies of this relationship should include not only past exposure to violence but also recent exposure.

The relationship between recent violence and increased substance use severity found in this study is consistent with other published data. Based on a two-year follow-up of a national probability sample of 3,006 women, Kilpatrick and colleagues (1997) concluded that 'drug use leads to a vicious cycle in which substance use increases risk of future assault and assault increases risk of subsequent substance use'. They also concluded that alcohol use appeared to be a reaction to assault and not a predictor of subsequent violence. Lastly, they found a strong predictor of recurrent violence based on prior violence exposure. Our

study confirms these associations, although does not point to the direction of causality. Our finding that recent violence was only weakly associated with increased severity of drug use was likely due to the low incidence of recent drug use and of recent violence at any one point through the follow-up period.

The lack of association of adherence to HIV medication with prior or recent violence has not been reported elsewhere. Because of the smaller number of subjects taking HIV medications, our study may not have had sufficient sample size to detect such a difference. However, there was not even a trend toward increased or decreased adherence. Larger studies of adherence to HIV medication would be required to evaluate this question.

A unique aspect of this study is the longitudinal nature of assessing violence exposure in combination with longitudinal assessment of health care utilization and other outcomes. The study by Eisenman and colleagues prospectively examined health care utilization in an HIV-infected sample but used a single baseline question on violence exposure as the sole determinant of this status (Eisenman et al., 2003). In addition, our study included large numbers of heterosexual men, a group that has been understudied with respect to the associations between violence exposure and health outcomes. The study population is unique in that all subjects had a history of alcohol problems and a high proportion of the sample had a history of injection drug use. These characteristics provided a sample with a heavy burden of violence exposure, helping to tease out the effect of recent violence exposure on a number of outcomes.

The study's main limitation is the lack of consistent follow-up by subjects over the 30-month follow-up period. Analytically, this was handled by using the Generalized Estimating Equation which takes advantage of longitudinal data while weighing responses based on number of observations for any subject. Because the main cause for lack of follow-up was late recruitment into the study, this would not bias the results in either direction. Another limitation was the self-report nature of the utilization data. This method has been used in numerous studies and shown to be a valid method for assessing health care utilization (Cleary & Jette, 1984; Roberts et al., 1996). Another limitation was the lack of statistical power to stratify the men by their sexual orientation to compare to the HIV Costs and Service Utilization Study sample (Eisenman et al., 2003; Zierler et al., 2000). Lastly, it would have been useful to stratify by the frequency and severity of violence exposure, but we were not able to measure that with the instruments used.

Conclusion

Violence is very common among HIV-infected men and women with a history of alcohol problems, including high exposure to recurring violence. The recent violence exposure was associated with increased health care utilization and worse substance use severity. These findings suggest that attention to violence prevention measures in such populations of HIV-infected persons may be an opportunity to minimize health care utilization and potentially improve physical and mental health. Future research about the health and health services implications of interpersonal violence exposure should account for both recent and past violence exposure.

Acknowledgements

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. Jane Liebschutz, MD, MPH, was a Robert Wood Johnson Foundation Generalist Physician Faculty Scholar during this study.

The authors wish to thank Seville Meli for asssistance in all aspects of the study, and Nicole Tibbets and the staff at DM-Stat for data management and statistical support.

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, LICSW as well as research assistance provided by: Jacqueline Savetsky, MPH, Jessica Garcia, Laura Weinstein, Sharon Phillips, Vera Doljanskaia, MPH, Kathleen Doyle and Katie Schoolwerth. The authors likewise appreciate the contributions of Howard Libman MD, and the clinical staff at the Beth Israel Deaconess Medical Center.

References

Bergman, B., Brismar, B., & Nordin, C. (1992). Utilisation of medical care by abused women. *British Medical Journal*, *305*, 27–28.

Calhoun, P. S., Bosworth, H. B., Grambow, S. C., Dudley, T. K., & Beckham, J. C. (2002). Medical service utilization by veterans seeking help for posttraumatic stress disorder. *American Journal of Psychiatry*, *159*(12), 2081–2086.

Chesney, M. A., Ickovics, J. R., Chambers, D. B., Gifford, A. L., Neidig, J., & Zwickl, B, W. (2000). Patient Care Committee and Adherence Working Group of the Outcomes Committee of the Adult AIDS Clinical Trials Group (AACTG). Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: The AACTG Adherence Instruments. *AIDS*, *12*, 255–266.

Clark, H. W., Masson, C. L., Delucchi, K. L., Hall, S. M., & Sees, K. L. (2001). Violent traumatic events and drug abuse severity. *Journal of Substance Abuse Treatment*, *20*, 121–127.

Cleary, P. D., & Jette, A. M. (1984). The validity of self-reported physician utilization measures. *Medical Care*, *22*(9), 796–803.

Cohen, M., Deamant, C., Barkan, S., Richardson, J., Young, M., Holman, S., Anastos, K., Judith, C., & Melnick, S. (2000). Domestic violence and childhood sexual abuse in HIV-infected women and women at risk for HIV. *American Journal of Public Health*, *90*(4), 560–563.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., & Smith, P. H. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine*, *23*(4), 260–268.

Coker, A. L., Smith, P. H., Bethea, L., King, M. R., & McKeown, R. E. (2000). Physical health consequences of physical and psychological intimate partner violence. *Archives of Family Medicine*, *9*(5), 451–457.

Cokkinides, V. E., Coker, A. L., Sanderson, M., Addy, C., & Bethea, L. (1999). Physical violence during pregnancy: Maternal complications and birth outcomes. *Obstetrics and Gynecology*, *93*(5 Part 1), 661–666.

Drossman, D. A., Leserman, J., Nachman, G., Li, Z., Gluck, H., Toomey, T. C., & Mitchell, M. (1990). Sexual and physical abuse in women with functional or organic gastrointestinal disorders. *Annals of Internal Medicine*, *113*, 828–833.

Drossman, D. A., Talley, N. J., Leserman, J., Olden, K. W., & Barreiro, M. A. (1995). Sexual and physical abuse and gastrointestinal illness: Review and recomendations. *Annals of Internal Medicine*, *123*, 782–794.

Ehrenstein, V., Horton, N., & Samet, J. (2004). Inconsistent condom use among HIV-infected patients with alcohol problems. *Drug and Alcohol Dependency*, *73*, 159–166.

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*(2), 125–127.

Ewing, J. (1984). Detecting alcoholism: The CAGE Questionnaire. *Journal of the American Medical Association*, *252*(14), 1905–1907.

Felitti, V., Anda, R., Nordenberg, D., Wiliamson, D., Spitz, A., Edwards, V., Koss, M., & Marks, J. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) study. *American Journal of Preventive Medicine*, *14*(4), 245–258.

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.

Frayne, S. M., Skinner, K. M., Sullivan, L. M., Tripp, T. J., Hankin, C. S., Kressin, N. R., & Miller, D. R. (1999). Medical profile of women Veterans Administration outpatients who report a history of sexual assault occurring while in the military. *Journal of Women's Health*, *8*(6), 835–845.

Gielen, A. C., McDonnell, K. A., Burke, J. G., & O'Campo, P. (2000). Women's lives after an HIV-positive diagnosis: Disclosure and violence. *Maternal and Child Health Journal*, *4*(2), 111–120.

Goodman, L. A., Koss, M. P., & Russo, N. F. (1993). Violence against women: Mental health effects. Part II. Conceptualizations of posttraumatic stress. *Applied and Preventice Psychology*, *2*, 123–130.

Gruskin, L., Gange, S., Celentano, D., Schuman, P., Moore, J., Zierler, S., & Vlahov, D. (2002). Incidence of violence against HIV-infected and uninfected women: Findings from the HIV Epidemiology Research (HER) study. *Journal of Urban Health*, *79*(4), 512–524.

Jinich, S., Paul, J. P., Stall, R., Acree, M., Kegeles, S., Hoff, C., & Coates, T. J. (1998). Childhood sexual abuse and HIV risk-taking behavior among gay and bisexual men. *AIDS and Behavior*, *2*(1), 41–51.

Johnsen, L. W., & Harlow, L. L. (1996). Childhood sexual abuse linked with adult substance use, victimization, and AIDS-risk. *AIDS Education and Prevention*, *8*(1), 44–57.

Kertesz, S., Horton, N., Friedman, P., Saitz, R., & Samet, J. (2003). Slowing the revolving door: Stabilization programs reduce homeless persons' substance use after detoxification. *Journal of Substance Abuse Treatment*, *24*, 197–207.

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

Kimerling, R., Armistead, L., & Forehand, R. (1999). Victimization experiences and HIV infection in women: Associations with serostatus, psychological symptoms and health status. *Journal of Traumatic Stress*, *12*(1), 41–58.

Koss, M. P., Koss, P. G., & Woodruff, W. J. (1991). Deleterious effects of criminal victimization on women's health and medical utilization. *Archives of Internal Medicine*, *151*, 342–347.

Liebschutz, J. M., Feinman, G., Sullivan, L., Stein, M. D., & 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., Finley, E. P., Braslins, P. G., Christiansen, D., Horton, N. J., & Samet, J. H. (2003). Screening for sexually transmitted infections in substance abuse treatment programs. *Drug and Alcohol Dependency*, *70*(1), 93–99.

Liebschutz, J. M., Mulvey, K. P., & Samet, J. H. (1997). Victimization among substance-abusing women: Worse health outcomes. *Archives of Internal Medicine*, *157*, 1093–1097.

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*(3), 121–128.

Marshall, R. P., Jorm, A. F., Grayson, D. A., Dobson, M., & O'Toole, B. (1997). Help-seeking in Vietnam veterans: Post-traumatic stress disorder and other predictors. *Australian and New Zealand Journal of Public Health*, *21*(2), 211–213.

Marshall, R. P., Jorm, A. F., Grayson, D. A., & O'Toole, B. I. (1998). Posttraumatic stress disorder and other predictors of health care consumption by Vietnam veterans. *Psychiatric Services*, *49*(12), 1609–1611.

McCauley, J., Kern, D., Kolodner, K., Derogatis, L., & Bass, E. (1998). Relation of low-severity violence to women's health. *Journal of General Internal Medicine*, *13*, 687–691.

McLellan, A. T., Luborsky, L., Cassiola, J., Griffith, J., Evans, F., Bass, H., & O'Brien, C. P. (1985). New data from the Addictions Severity Index: Reliability and validity in three centers. *Journal of Nervous and Mental Diseases*, *173*, 412–423.

Morrill, A. C., Kasten, L., Urato, M., & Larson, M. J. (2001). Abuse, addiction, and depression as pathways to sexual risk in women and men with a history of substance abuse. *Journal of Substance Abuse*, *13*(1–2), 169–184.

Mullen, P. E., Romans-Clarkson, S. E., Walton, V. A., & Herbison, G. P. (1988). Impact of sexual and physical abuse on women's mental health. *The Lancet*, 841–845.

National Institute on Alcoholism and Alcohol Abuse (1995). *The physician's guide to helping patients with alcohol problems* (pp. 95–3769). Washington, DC: NIH publication.

Orth-Gomer, K., Wamala, S. P., Horsten, M., Schenck-Gustafsson, K., Schneiderman, N., & Mittleman, M. A. (2000). Marital stress worsens prognosis in women with coronary heart disease: The Stockholm Female Coronary Risk Study. *Journal of the American Medical Association*, *284*(23), 3008–3014.

Plichta, S. (1992). The effects of woman abuse on health care utilization and health status: A literature review. *Women's Health Issues*, *2*(3), 154–163.

Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

Resnick, H., Acierno, R., & Kilpatrick, D. (1997). Health impact of interpersonal violence. 2: Medical and mental health outcomes. *Behavioral Medicine*, *23*, 65–78.

Roberts, R. O., Bergstralh, E. J., Schmidt, L., & Jacobsen, S. J. (1996). Comparison of self-reported and medical record health care utilization measures. *Journal of Clinical Epidemiology*, *49*(9), 989–995.

Rosenberg, H. J., Rosenberg, S. D., Wolford, G. L., II, Manganiello, P. D., Brunette, M. F., & Boynton, R. A. (2000). The relationship between trauma, PTSD, and medical utilization in three high-risk medical populations. *International Journal of Psychiatry and Medicine*, *30*(3), 247–259.

Samet, J., Horton, N., Dukes, K., Meli, S., Tripp, T., LaBelle, C., & Freedberg, K. (2002). A randomized controlled trial (RCT) of a multidimensional intervention to enhance adherence to antiretriviral therapy in HIV-infected patients with a history of alcohol problems. Paper presented at the XIV International AIDS Meeting, Barcelona, Spain.

Samet, J., Phillips, S., Horton, N., Traphagen, E., & Freedberg, K. A. (2004). Detecting alcohol problems in HIV-infected patients: Use of the CAGE questionnaire. *AIDS Research and Human Retroviruses*, *20*, 151–156.

Samet, J. H., Libman, H., LaBelle, C., Steger, K., Lewis, R., Craven, D. E., & Freedberg, K. A. (1995). A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. *Archives of Internal Medicine*, *155*(15), 1629–1633.

Schelling, G., Stoll, C., Haller, M., Briegel, J., Manert, W., Hummel, T., Lenhart, A., Heyduck, M., Polasek, J., Meier, M., Preuss, U., Bullinger, M., Schuffel, W., & Peter, K. (1998). Health-related quality of life and posttraumatic stress disorder in survivors of the acute respiratory distress syndrome. *Critical Care Medicine*, *26*(4), 651–659.

Schnurr, P. P., Friedman, M. J., Sengupta, A., Jankowski, M. K., & Holmes, T. (2000). PTSD and utilization of medical treatment services among male Vietnam veterans. *Journal of Nervous and Mental Diseases*, *188*(8), 496–504.

Tjaden, P., & Thoennes, N. (2000). *Prevalence, incidence and consequences of violence against women: Findings from the National Violence Against Women Survey*: National Institute for Justice and Centers for Disease Control and Prevention.

Vlahov, D., Wientge, D., Moore, J., Flynn, C., Schuman, P., Schoenbaum, E., & Zierler, S. (1998). Violence among women with or at risk for HIV infection. *AIDS and Behavior*, *2*(1), 53–60.

Windle, M. (1994). Substance use, risky behaviors, and victimization among a US national adolescent sample. *Addiction*, *89*, 175–182.

Zierler, S., Cunningham, W. E., Andersen, R., Shapiro, M. F., Nakazono, T., Morton, S., Crystal, S., Stein, M., Turner, B., St Clair, P., & Bozzette, S. A. (2000). Violence victimization after HIV infection in a US probability sample of adult patients in primary care *American Journal of Public Health*, *90*(2), 208–215. (Published erratum appears in *American Journal of Public Health*, 2000, *90*(3), 447.

Zierler, S., Feingold, L., Laufer, D., Velentgas, P., Kantrowitz-Gordon, I., & Mayer, K. (1991). Adult survivors of childhood sexual abuse and subsequent risk of HIV infection. *American Journal of Public Health*, *81*(5), 572–575.

CLINICAL PRACTICE

HIV Postexposure Prophylaxis in Sexual Assault: Current Practice and Patient Adherence to Treatment Recommendations in a Large Urban Teaching Hospital

Judith A. Linden, MD, Paula Oldeg, MD, Supriya D. Mehta, MHS, PhD, Kerry K. McCabe, MD, Colleen LaBelle, RNP

Abstract

Background: Although rare, HIV transmission is one of the most feared consequences of sexual assault. While availability of medications to prevent HIV transmission (HIV nonoccupational postexposure prophylaxis [HIV nPEP]) is increasing, little is known about emergency department (ED) prescribing practices and patient adherence to treatment recommendations. **Objectives:** To determine factors associated with offering, following up with, and adhering to treatment when HIV nPEP is initiated for sexual assault victims. **Methods:** This was a retrospective chart review of female patients presenting with complaint of sexual assault to an urban ED from October 1, 1999, to September 30, 2002. HIV nPEP medications and/or follow-up were provided without charge. Chi-square analysis identified factors associated with being offered or referred for nPEP and follow-up. Variables significant at the $p < 0.10$ level were entered into logistic regression analysis. **Results:** Two hundred twenty-nine charts were reviewed. The final sample size was 181. Mean age was 29.1 years; median time from assault to presentation was 10.1 hours; 51.5% of the assailants were known to the victims. HIV nPEP was offered to 89 (49%) patients, and 11 patients were referred to an HIV nurse. Eighty-five (85%) patients accepted, 38 of these 85 (45%) followed up, and 18 of the 85 (21%) completed treatment. In multivariate analysis, three variables were statistically significantly associated with increased likelihood of referral or being offered HIV nPEP: unknown assailant, having insurance, and younger age. Treatment was completed by 15 of 82 (18%) of ED-initiated patients, versus three of three (100%) referred for initiation. The authors were unable to identify factors associated with completing treatment. **Conclusions:** HIV nPEP was offered to less than half of sexual assault patients, and few completed treatment. Further studies are needed to evaluate and improve appropriateness of HIV nPEP administration and follow-up.
Key words: sexual assault treatment; HIV prophylaxis; emergency department. ACADEMIC EMERGENCY MEDICINE 2005; 12:640–646.

HIV postexposure prophylaxis (HIV PEP) is a form of secondary prevention of HIV transmission, and is often divided into occupational (i.e., exposure while on the job) and nonoccupational. Nonoccupational exposures include sexual assault (SA), human bites with skin break, unintended needlesticks (outside of the job), unsafe sexual practices, and injection drug use. HIV PEP has become the standard of care for occupational exposures, with clear recommendations supported by the literature.[1–3] The indications for nonoccupational postexposure prophylaxis (nPEP), however, are not well defined. The Centers for Disease Control and Prevention (CDC) recently came out with clear recommendations for treatment after exposure to known HIV-positive body fluids within 72 hours, but these recommendations leave treatment of those with sexual exposure to an unknown source to the discretion of the provider, "taking into account the risks and benefits on a case by case basis."[4] They do, however, acknowledge that certain aspects of an SA may put the victim at higher risk for HIV transmission, and that the HIV status of the assailant is most often unknown.[4] There is some evidence that HIV PEP can be beneficial when given early in nonoccupational exposures to HIV.[5] Possible exposure during SA is the least controversial use of HIV nPEP (since the exposure is isolated and unlikely to recur) and accounts for a large percentage of patients seeking nPEP.[6]

The calculated risks of acquiring HIV through sexual contact are similar to those in occupational exposure, ranging from 0.3% to 0.5% per contact (anal intercourse) to 0.1% (penile–vaginal intercourse).[7–9] There have also been isolated reports of HIV transmission by the penile–oral route.[10] People with genital and mucous membrane injury or inflammation are at increased risk of contracting HIV.[9,11,12] Bodily trauma, possible exposure to other sexually transmitted diseases (STDs),

From Boston University School of Medicine, Boston Medical Center (JAL, SDM, KKM), Boston Medical Center Emergency Medicine Residency (PO), and Boston Medical Center (CL), Boston, MA.
Received November 29, 2004; revision received January 26, 2005; accepted January 31, 2005.
Address for correspondence and reprints: Judith A. Linden, MD, Department of Emergency Medicine, Boston University School of Medicine, Dowling 1 South, 818 Harrison Avenue, Boston, MA 02118. Fax: 617-414-7759; e-mail: jlinden@bu.edu.
doi:10.1197/j.aem.2005.01.015

and presence of multiple assailants may increase the risk of HIV transmission in SA. Furthermore, this risk is a real concern to SA survivors, with 25% to 40% expressing serious concerns.[13,14]

Few states have released recommendations for HIV nPEP in the general population.[15–17] Although many emergency departments (EDs) and health centers are offering HIV nPEP to SA survivors, there are few published studies evaluating patient adherence to treatment recommendations.[6,18–21] There are no studies examining factors associated with physician offering of HIV nPEP in the ED. The purpose of this study was to determine factors associated with offering HIV nPEP to SA patients presenting to an urban ED, and subsequent follow-up and completion of HIV nPEP.

METHODS

Study Design. We conducted a retrospective chart review of female patients aged 18 years or older presenting to the ED with a complaint of SA between October 1, 1999, and September 30, 2002. The institutional review board of Boston Medical Center approved the study.

Study Setting and Population. The study was conducted in the ED of Boston Medical Center, an urban teaching hospital with an annual census of 100,000. Eligible patients were aged 18 years or older, presenting with a chief complaint of SA. Exclusion criteria were: male gender, examination not performed in the ED, known HIV-positive status, already entered into the study, attempted assault without genital penetration, digital/object or protected assault, patient left before evaluation, patient left before examination, patient left before treatment was completed, incomplete history due to intoxication, or altered mental status. Patients were identified via ED visit log by diagnosis, and via a secondary ED psychiatry log. Follow-up data were obtained from an HIV clinic nurse to determine compliance with clinic visits and completion of medication regimen.

Study Protocol. The ED has access to a sexual assault nurse examiner (SANE) program funded through the Department of Public Health. Standard ED treatment for SA victims included evaluation by an emergency physician with or without a SANE. Treatment included a screening medical examination (performed by an emergency medicine [EM] resident), history and physical examination, collection of evidence utilizing a state-supplied evidence kit performed by a SANE or an EM resident (if the patient consented), serologic tests for hepatitis, and urine pregnancy testing. Emotional support was offered by an ED psychiatric nurse, a rape crisis advocate, the SANE (when present), and the emergency physician. Individual emergency physicians were responsible for prescribing STD and pregnancy prophylaxis when indicated. Our department's guidelines recommended that nPEP be offered to any SA patient with penetration and body fluid exposure. There was a written protocol for initiating nPEP, which was adapted from the occupational PEP protocol. Efforts to educate staff about nPEP and our protocol included a grand rounds on HIV nPEP, and ongoing input from SANE consultants. The decision to provide nPEP was made by the treating physician in consultation with the SANE. The HIV nPEP medications and follow-up were provided free of charge. Three-day nPEP starter packs were distributed in the ED, with follow-up arranged in one to three days. Protocol included follow-up with a designated nurse in the HIV clinic on the next business day, initiated by the patient.

Measures. Study variables included age, mode of arrival to the ED, insurance status, time elapsed between assault and ED presentation, identity and number of assailants (acquaintance or stranger), type of assault, and presence and source of bleeding in the patient or assailant (as reported by the patient). Most of these variables were abstracted from the standardized report form in the Massachusetts Sexual Assault Evidence Collection Kit.

All study variables were explicitly defined. All data abstractors met periodically to review and discuss uniform handling of the data that were missing, ambiguous, or conflicting. If there was no documentation of discussion about HIV nPEP in the nurses' notes, the physicians' notes, or the SANE notes, then the patient was classified as "not offered" HIV nPEP. A 20% random sample of the final data set was selected for an interrater reliability check of data abstraction.

Data Analysis. Data were entered into a Microsoft Access 2000 database (Microsoft Corp., Redmond, WA), and analyzed using Stata SE 7.0 for Windows (StataCorp LP, College Station, TX). Chi-square and Student's t-test were used to identify factors associated with the offering of, follow-up with, and completion of HIV nPEP in the ED. All variables statistically significant at the $p < 0.10$ level from exploratory and univariate logistic regression analysis were entered into multivariate logistic regression analysis. A p-value of <0.10 was chosen for the exploratory analysis due to the small sample size and limited power. We chose a less conservative threshold for significance because this was not a study of intervention effect, but an attempt to identify clinically significant factors associated with being offered or referred for HIV nPEP. Outcomes for analysis were: 1) whether a patient was offered/given PEP in the ED, 2) whether the patient followed up for at least one visit at the HIV clinic, and 3) whether

he or she completed the one-month recommended course of therapy.

RESULTS

Two hundred twenty-nine charts were identified; 188 were eligible for inclusion, and seven charts were unavailable for review, leaving a final sample size of 181 (96%). Reasons for exclusion were as follows: male gender (7); previously documented HIV-positive status (3); left before evaluation, examination, or treatment was completed (14); patient history incomplete due to intoxication or altered mental status (6); digital, protected, or attempted assault without actual penetration (4); already entered into the study (5); and examination not performed in the ED (2).

Patients reported being assaulted by one assailant in 150 (83%) cases, and two or more assailants in 21 (12%) cases. Three patients (1.7%) reported an assault by an assailant with known or high risk for HIV. Vaginal assault was the most common type of assault (53%), followed by vaginal and oral (15%); vaginal, anal, and oral (6%); and oral (5%). Fifteen percent of the assaults were of undocumented or unknown type. Median time from assault to ED presentation was 10.1 hours (range: 0–241, binomially obtained 95% CI = 7.1 to 12.9). Characteristics of the patients are presented in Table 1.

For the 20% of charts that were randomly selected for interrater reliability, kappa values were 0.68 for whether nPEP was offered/given, $\kappa = 0.92$ for insurance status, $\kappa = 0.61$ for blood exposure, $\kappa = 0.87$ for type of assault, and $\kappa = 0.78$ for type of assailant ($n = 41$ each measure).

The mean age of the subjects was 29.1 years (median 27.6, range 18–82). Nearly half (48%) of the patients had Medicaid as their primary insurance, 17% were self-pay (no insurance), 28% had private insurance, and 16% had Free Care coverage by the Massachusetts uncompensated care pool (funds distributed to hospitals for the care of patients who are not eligible for Medicare or Medicaid but are financially eligible for assistance from state funds).

The HIV nPEP treatment was offered to 89 of 181 (49%) patients, of whom seven of 89 (8%) declined, and 82 of 89 (92%) accepted (Figure 1). Of the 82 patients who accepted treatment offered in the ED, 78 of 82 (95%) were prescribed a two-drug regiment of zidovudine and lamivudine (common name 3 TC; Combivir is the combination of the two), and four of 82 (5%) were prescribed a three-drug regimen of Combivir and nelfinavir. Of the three patients whose care was started by the HIV nurse, two (66%) were started on the two-drug regimen, and one (33%) was started on the three-drug regimen.

The HIV nNPEP treatment was not offered to 57 of 181 patients (31%). Additionally, 14 of 181 (8%) were not given nPEP because it was not indicated, 11 of 181 (6%) were referred to the HIV clinic to discuss possible treatment, and ten of 181 (6%) had unclear charts (i.e., contradictions in physician notes, nursing notes, or discharge instructions). HIV nPEP was not indicated in 14 patients for the following reasons: time of presentation was outside the suggested treatment window (>72 hours) ($n = 10$), no body fluid exposure ($n = 1$), and assaulted by current significant other ($n = 3$).

Of the 82 patients who began nPEP in the ED, 35 of 82 (43%) followed up for at least one visit, but only 15 of 82 (18%) completed the full one-month course of treatment. Of the 11 patients referred to the HIV clinic nurse, only six actually went. Half (three of six) of the patients who were referred to the HIV clinic nurse to discuss treatment were started on nPEP, and three of three (100%) completed treatment, although one stopped nelfinavir secondary to side effects but continued the Comibivir. There was no documented seroconversion during the follow-up period (Figure 1).

When evaluating factors associated with being offered or referred nPEP in the ED, patients with genital bleeding and younger age were more likely to be offered or referred (Table 2). Patients with a known

TABLE 1. Patient and Assault Characteristics

Patient Characteristics	*n*	%
Age		
≤21 years	60	36.5
22–32 years	55	30.4
≥33 years	66	33.2
Mode of arrival		
EMS	83	45.9
Other	98	54.1
Insurance status		
Self-pay (uninsured)	30	16.6
Insured (private, Medicare, Medicaid)	143	79.0
Unknown	8	4.4
Time to ED presentation		
<24 hours	132	72.9
24–72 hours	25	13.8
>72 hours	24	13.3
Number of assailants		
1	150	82.9
2	10	5.5
>2	11	6.1
Unknown	10	5.5
Type of assault		
Vaginal	97	53.6
Oral	10	5.5
Vaginal and oral	27	14.9
Vaginal and anal	7	3.9
Anal and oral	1	0.6
Vaginal, anal, and oral	11	6.1
Unknown	28	15.5
SANE involved?		
Yes	127	70.2
No	54	29.8

SANE = sexual assault nurse examiner.

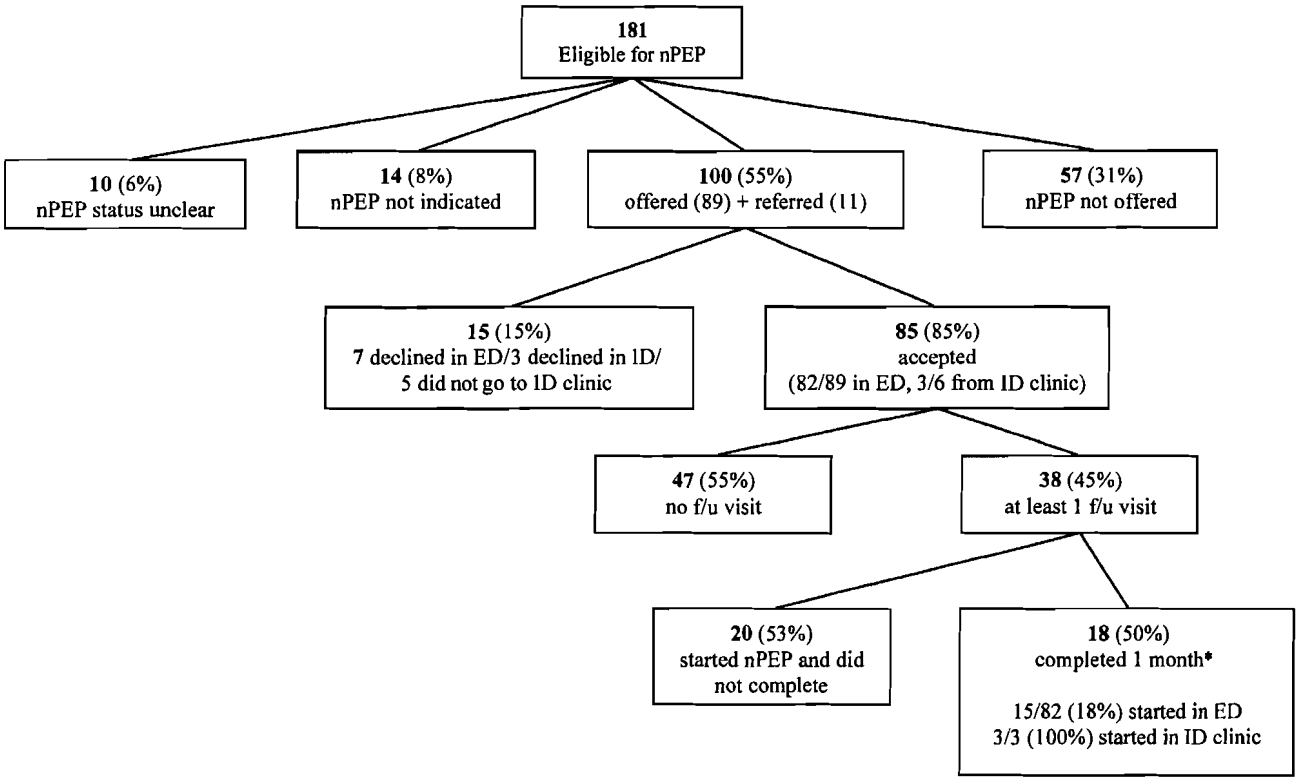


Figure 1. Follow-up (f/u) and completion status of patients started on HIV nonoccupational postexposure prophylaxis (nPEP) in the emergency department (ED) and in the infectious disease (ID) clinic. *Eighteen of 85 (21%) of the total who accepted nPEP completed one month.

assailant and no insurance were less likely to be offered nPEP. Time to presentation, type of assault, number of assailants, high-risk exposure, mode of arrival to the ED, and SANE involvement were not associated with whether patients were offered or referred for nPEP. In multivariate analysis (Table 2), three variables were associated with offering nPEP: stranger assault, any type of insurance, and age less than 33 years. Patients assaulted by an acquaintance and uninsured patients were less than half as likely to be offered nPEP (odds ratio [OR] = 0.41, 95% confidence interval [95% CI] = 0.19 to 0.88). Younger patients and those with genital bleeding were more likely to be offered nPEP.

TABLE 2. Factors Associated with Being Offered or Referred for HIV Nonoccupational Postexposure Prophylaxis (nPEP) in the ED

	n (%) nPEP	Chi-square p-value	OR (95% CI)	AOR (95% CI)
Insurance status*				
Any insurance	86/123 (70%)			
No insurance	13/28 (46%)	0.018	0.37 (0.16, 0.86)	0.26 (0.10, 0.66)
Age†				
≥33 years	29/57 (51%)			
22–32 years	33/49 (67%)		1.99 (0.90, 4.39)	2.49 (1.05, 5.95)
≤21 years	38/51 (75%)	0.032	2.82 (1.25, 6.38)	3.30 (1.32, 8.82)
Acquaintance‡				
No	51/70 (73%)			
Yes	47/80 (59%)	0.070	0.53 (0.27, 1.06)	0.41 (0.19, 0.88)
Genital bleeding‡				
No	86/141 (61%)			
Yes	13/15 (87%)	0.050	4.16 (0.90, 19.1)	3.05 (0.62, 15.1)

*Reference category was "any insurance."
†Reference category was age "≥33 years."
‡Reference category was "no."
OR = odds ratio from univariate logistic regression; AOR = adjusted odds ratio from multivariate logistic regression.

Only 18 patients completed nPEP; we were unable to assess factors associated with completion due to small sample size.

DISCUSSION

To the best of our knowledge, ours is one of the largest studies evaluating HIV nPEP given to SA victims from the ED.[20,21] Other large studies included injection drug exposures, subsets of SA victims (pediatric and adolescent), and all blood/body fluid exposures.[18,22] While the practice of offering SA victims prophylaxis for pregnancy and STDs is standard of care, offering HIV nPEP is becoming increasingly common, given SA victims' concerns regarding acquiring HIV from the rape. Studies evaluating nPEP in men at high risk for HIV exposure and given medications to begin immediately upon exposure have shown a compliance rate of greater than 90%, and efficacy in decreasing seroconversion.[5] Kahn et al. studied completion rates in volunteers with sexual exposure or injection drug use, and found a 78% completion rate, with 12% returning for a repeat exposure.[22] Sexual assault victims differ in that the serostatus of the assailant is often unknown, therefore making estimates of risk of exposure and subsequent treatment recommendations more difficult. One study of incarcerated prisoners charged with a sexual offense in Rhode Island suggests a higher rate of HIV positivity than in the general incarcerated population (1% vs. 0.3%), although this may vary by region, and the prevalence is still low.[23]

Sexual assault victims are often in crisis, and may have many safety concerns not related to HIV, making comprehension of and decision making in a very complicated topic more difficult. Recent CDC recommendations acknowledge that HIV can be transmitted through SA (albeit rarely), and that nPEP treatment has become widely encouraged. The CDC also recommends that SA survivors will benefit from support services to increase the likelihood of completing the one-month therapy.[4]

Our data confirm findings of previous studies reporting a low likelihood of follow-up and completion of nPEP treatment in victims of SA, with follow-up rates ranging from 38% to 53%, and completion rates ranging from 11% to 63%.[18–20,24] In fact, the completion rate of health care workers who begin nPEP (a highly motivated and knowledgeable group of patients) is in the vicinity of 30%, although they are more likely to be given a three-drug regimen that carries with it a higher likelihood of adverse reactions.[25] Although our overall completion rate was 21%, the completion rate of those patients started on nPEP from the HIV clinic was 100%, whereas only 18% of those patients who were started from the ED completed the 30-day regimen. We are unable to tell whether this is because the group of patients who went to the HIV clinic were more motivated, were more organized, or perceived a higher risk, or whether the delayed discussion of the pros and cons of nPEP was more beneficial. All victims received the support of the HIV nurse, whether they were started in the ED or in the infectious disease clinic.

We were able to identify factors that were associated with a lower likelihood of being offered nPEP in our ED. One of these factors (acquaintance rape) is not surprising, as the HIV status of the assailant is more likely to be known. There may be a bias on the part of the provider that there is ongoing exposure. In fact, ongoing exposure may have been a rational reason on the part of the provider not to provide nPEP. Two other factors, lack of insurance and older age, are more disturbing. At our hospital, the medications given in nonoccupational exposures can cost $600.00 (two-drug regimen) to $1,200.00 (three-drug regimen) (Epocrates Rx, version 6.13, 2004, San Mateo, CA), and follow-up visits and laboratory tests an additional $1,200.00 (Boston Medical Center laboratory finances department), possibly posing a significant barrier to uninsured/underinsured patients, or those not wanting to use their insurance (for instance, college students covered by their parents' insurance). Massachusetts has an HIV drug assistance program (HDAP) (funded through the state Department of Public Health) that covers nPEP medications for uninsured patients. A recent survey, however, suggests that many ED practitioners are not aware of this program.[26] Older age and not being offered nPEP may possibly be due to an age bias, although our study population was skewed toward younger age. The reasons behind the decision-making process cannot be known from this retrospective review, and warrant further study. Standardized protocols and guidelines may clarify treatment recommendations and allow more uniform access to ED-initiated nPEP in SA victims. Unfortunately, this is difficult, given the paucity of clear outcome data from large-scale studies in this population. The new CDC recommendations suggest incorporating a more selective approach for offering HIV nPEP.[4] Guidelines that include consideration of exposure risk category and HIV risk status of source may assist providers.[15–17,27]

LIMITATIONS

This study is limited by the weaknesses inherent in retrospective reviews. However, abstraction bias was limited by using formal inclusion criteria, prospective variable definition, abstractor training, standardized abstraction forms, abstractor monitoring, and assessment of interrater reliability.[28]

The HIV nPEP treatment may have been discussed with and offered to the patient but not documented. The decision to treat, although made in conjunction with an HIV nurse, with the guidance of Boston

Medical Center ED guidelines, was ultimately made by the treating physician. Although the SANE program and our ED had treatment guidelines, the standard of care for nonoccupational HIV exposures is in flux. The most recent CDC recommendations for HIV nPEP treatment in SA victims acknowledge that several aspects of SA may put victims at increased risk for HIV exposure, but that reported seroconversions remain extremely rare. The CDC report stops short of recommending nPEP, but acknowledges that the use of nPEP in SA survivors has been widely encouraged in the United States and elsewhere by many experts.[4,29,30]

Our results may not be generalizable to other settings. Our study did not include male SA victims (due to small numbers), who are often higher-risk exposures. Finally, our sample size was not large enough to allow us to identify factors associated with increased likelihood of follow-up and completion of PEP, limiting our ability to identify and define a population of better candidates for nPEP.

CONCLUSIONS

Nonoccupational HIV postexposure prophylaxis was offered to less than half of the sexual assault survivors in the ED, according to our chart review. Of those who opted to begin prophylaxis, completion rate was low; less than half followed up for one visit, and only 18% of those who began treatment in the ED completed the treatment course. Patients who accepted nPEP through the HIV clinic more often completed the one-month course; however, these patients were started after the initial ED visit, and may have been more able to focus on the discussion about risks/benefits of nPEP and make a more rational decision. HIV nPEP was more likely to be offered to those patients with insurance, younger age, and genital bleeding.

Postexposure prophylaxis for nonoccupational exposures to HIV, including SA, is an evolving topic, with recent data helping to inform practice. Emergency departments are well suited to offer treatment to prevent pregnancy, HIV disease, and other STDs for SA survivors, given the familiarity with protocols for PEP in health care workers, and the time-sensitive nature of the intervention. Although treatment in high-risk populations is feasible, the population of sexually assaulted patients in the ED represents a particular challenge, as follow-up and completion rates appear to be poor.[5] More information clarifying the recommendations for treatment and methods for increasing the likelihood of completion of treatment is needed. It is important to ensure that patients are offered treatment without bias with respect to insurance status or age. Prospective studies using standardized treatment protocols, taking into account risk of exposure and regional HIV seroprevalence rate, with 24-hour expert counseling by persons knowledgeable in the pros and cons of HIV PEP, and interviews exploring reasons why patients stop taking the PEP medications (e.g., side effects, low risk of transmission, unaffordability of medications) will help establish more definitive protocols for prescription of nPEP.

The authors gratefully acknowledge the contributions of Joshua Frank and Joyia Richman of Boston University School of Medicine for their chart review and data acquisition. The authors also acknowledge the contributions of Dr. Gail D'Onofrio and Dr. Andrew Ulrich for their manuscript review and editorial comments.

References

1. Cardo DM, Culver DH, Ciesielski CA, et al. A case-control study of HIV seroconversion in health care workers after percutaneous exposure. N Engl J Med. 1997; 337:1485–90.
2. Gerberding JL. Prophylaxis for occupational exposure to HIV. Ann Intern Med. 1996; 125:497–501.
3. Centers for Disease Control and Prevention. Updated U.S. Public Health Service guidelines for the management of occupational exposures to HBV, HCV, HIV and recommendations for postexposure prophylaxis. MMWR. 2001; 50(rr11):1–52.
4. Centers for Disease Control and Prevention. Antiretroviral postexposure prophylaxis after sexual, injection-drug use, or other nonoccupational exposure to HIV in the United States. MMWR. 2005; 50(rr02):1–20.
5. Schechter M, Lago RF, Ismerio R, et al. Behavioral impact, acceptability, and HIV incidence among homosexual men with access to postexposure chemoprophylaxis for HIV. J Acquir Immun Defic Syndr. 2004; 35:519–25.
6. Merchant RC, Becker BM, Mayer KH, Fuerch J, Schreck B. Emergency department blood and body fluid exposure evaluations and HIV postexposure prophylaxis usage. Acad Emerg Med. 2003; 10:1345–53.
7. Downs AM, De Vicenzi I. Probability of heterosexual transmission of HIV: relationship to the number of unprotected sexual contacts. J Acquir Immun Defic Syndr Hum Retrovirol. 1996; 11:388–95.
8. DeGruttola V, Seage GR, Mayer KH, Horsburgh CR Jr. Infectiousness of HIV between male homosexual partners. J Clin Epidemiol. 1989; 42:849–56.
9. Gray RH, Wawer MJ, Brookmeyer R, et al. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1 discordant couples in Rakai, Uganda. Lancet. 2001; 357:1149–53.
10. Lifson AR, O'Malley PM, Hessol NA, Buchbinder SP, Cannon L, Rutherford GW. HIV seroconversion in two homosexual men after receptive oral intercourse with ejaculation: implications for counseling concerning safe sexual practices. Am J Public Health. 1991; 80:1509–11.
11. Nicolosi A, Leite MLC, Musicco M, Arici C, Gavazzeni G, Lazzarin A. The efficiency of male-to-female sexual transmission of the human immunodeficiency virus: a study of 730 stable couples. Epidemiology. 1994; 5:570–5.
12. Seidlin M, Vogler M, Lee E, Lee YS, Dubin N. Heterosexual transmission of HIV in a cohort of couples in New York City. AIDS. 1993; 7:1247–54.
13. Kilpatrick DG, Edmunds CN, Seymour AK. Rape in America—A Report to the Nation. Crime Victim Research and Treatment Center. Charleston, SC: Medical University of South Carolina, 1992.
14. Baker TC, Burgess AW, Brickman E, Davis RC. Rape victims' concern about possible exposure to HIV infection. J Interpers Viol. 1990; 5(1):49–60.

15. Koh HK, DeMaria A, Flatley McGuire J. Clinical Advisory—HIV Prophylaxis for Non-Occupational Exposures. The Commonwealth of Massachusetts. Available at: http://www.mass.gov/dph/aids/guidelines/ca_exposure_nonwork.htm, published Oct 2000. Accessed Nov 21, 2004.
16. Mayer KH, Merchant RC, Browning CA. Nonoccupational Human Immunodeficiency Virus Postexposure Prophylaxis Guidelines for Rhode Island Healthcare Practitioners. Available at: http://www.health.state.ri.us/disease/NPEPFinalDraftJuly26.pdf, published Aug 2002. Accessed Nov 21, 2004.
17. New York State Guidelines. Available at: http://www.hivguidelines.org/public_html/npep/npep.htm, published Jun 2004. Accessed Nov 21, 2004.
18. Babl FE, Cooper ER, Damon B, Louie T, Kharasch S, Harris J. HIV postexposure prophylaxis for children and adolescents. Am J Emerg Med. 2000; 18:282–7.
19. Klemens S, Weaver AG, Wong C, Sable N. HIV prophylaxis following sexual assault: experience with 30 survivors [abstract]. J Clin Infect Dis. 2001; 33:1214.
20. Wiebe ER, Connor SE, McGregor M, Ducceschi S. Offering HIV prophylaxis to people who have been sexually assaulted: 16 months' experience in a sexual assault service. Can Med Assoc J. 2000; 162:641–5.
21. Myles J, Hirozzawa A, Katz M, Kimmerling R, Bamberger J. Post-exposure prophylaxis for HIV after sexual assault [letter]. JAMA. 2000; 284:1516–8.
22. Kahn JO, Martin JN, Roland ME, Bamberger JD, et al. Feasibility of postexposure prophylaxis (PEP) against human immunodeficiency virus infection after sexual or injection drug use exposure: The San Francisco PEP study. J Infect Dis. 2001; 183:707–14.
23. Rich JD, Merchant RC, Salas C, Marcussen P, Grundy M. HIV seroprevalence of adult males incarcerated for a sexual offense in Rhode Island, 1994–1999 [letter]. JAMA. 2002; 288: 164–5.
24. Limb S, Kawsar M, Forster GE. HIV post-exposure prophylaxis after sexual assault: the experience of a sexual assault service in London. Int J STD AIDS. 2002; 13:602–5.
25. Tokars JL, Marcus R, Culver DH, et al. Surveillance of HIV infection and zidovudine use among health care workers after occupational exposure to HIV-infected blood: the CDC Cooperative Needlestick Surveillance Group. Ann Intern Med. 1993; 118:913–9.
26. Grayson J, Lawrence P, Kunches L, et al. Post-exposure prophylaxis (nPEP) protocols and barriers to nPEP utilization in Massachusetts EDs [abstract]. Ann Emerg Med. 2004; 44:S125.
27. Havens PL, and the Committee on Pediatric AIDS. Clinical report: postexposure prophylaxis in children and adolescents for nonoccupational exposure to human immunodeficiency virus. Pediatrics. 2003; 111:1475–89.
28. Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta D, Steiner J. Medical chart reviews in emergency medicine research: where are the methods? [abstract] Acad Emerg Med. 1994; 1:A62.
29. Katz MH, Geberding JL. The care of persons with recent sexual exposure to HIV. Ann Intern Med. 1998; 128:306–12.
30. Gostin LO, Lazzarini Z, Alexander D, Brandt AM, Mayer KH, Silverman DC. HIV testing, counseling and prophylaxis after sexual assault. JAMA. 1994; 271:1436–44.

The Assessment of Trauma History in Women With Co-occurring Substance Abuse and Mental Disorders and a History of Interpersonal Violence

Gregory J. McHugo, PhD
Yael Caspi, ScD
Nina Kammerer, PhD
Ruta Mazelis, BS
Elizabeth W. Jackson, PhD
Lisa Russell, PhD
Colleen Clark, PhD
Jane Liebschutz, MD
Rachel Kimerling, PhD

Abstract

The Women, Co-occurring Disorders, and Violence Study (WCDVS) was a large (N = 2729) multisite study of the effectiveness of integrated and trauma-informed services for women with substance use and mental health disorders and a history of interpersonal violence (physical or sexual abuse). Study participants' exposure to lifetime and current traumatic events was assessed at baseline and follow-up via in-person interviews. This article describes the choice of the Life Stressor Checklist-Revised (LSC-R) to assess trauma history to meet the WCDVS's research aims and to respond to consumer input. Quantitative data address the breadth and prevalence of potentially traumatic events in the past and current lives of study participants, the formation and properties of summary measures, and test-retest reliability. Qualitative data address tolerance of the instrument by interviewers and respondents and the generalizability of quantitative findings about trauma prevalence. Finally, recommendations are offered for improvements to the WCDVS version of the LSC-R for use in future research.

Address correspondence to Gregory J. McHugo, PhD, Research Associate Professor, NH-Dartmouth Psychiatric Research Center, 2 Whipple Place, Suite 202, Lebanon, NH 03766. E-mail: gregory.mchugo@dartmouth.edu.

Yael Caspi, ScD, is a senior researcher in the Department of Psychiatry, Rambam Medical Center, Haifa, Israel.

Nina Kammerer, PhD, is a lecturer in the Department of Anthropology at Brandeis University, Waltham, Mass.

Ruta Mazelis, BS, is a private consultant from Carollton, Ohio, affiliated with Sidran Institute, Carrollton, Ohio.

Elizabeth Jackson, PhD, is a research associate in the Sheps Health Services Research Center, UNC-Chapel Hill, Chapel Hill, NC.

Lisa Russell, PhD, is an associate director of research, ETR Associates, Scotts Valley, Calif.

Colleen Clark, PhD, is a research assistant professor in the University of South Florida/FMHI, Tampa, Fla.

Jane Liebschutz, MD, is an assistant professor of medicine, Boston Medical Center, Boston, Mass.

Rachel Kimerling, PhD, is a research clinical psychologist in the National Center for PTSD, Palo Alto VA Health Care System (352–117MPD), Menlo Park, Calif.

Journal of Behavioral Health Services & Research, 2005, 32(2), 113–127.

The Women, Co-occurring Disorders, and Violence Study (WCDVS) was a large federally funded multisite study ($N = 2729$) of the effectiveness of comprehensive, integrated, trauma-informed, and consumer-involved services for women with substance use and mental health disorders and a history of interpersonal violence.* The aims of this article are to contribute to the understanding of measuring potentially traumatic events and to describe the kinds and frequencies of such events in the lives of women with co-occurring disorders and histories of physical or sexual abuse. An existing trauma assessment instrument, the Life-Stressors Checklist-Revised (LSC-R), was adapted for the WCDVS. Data are presented on the test-retest reliability of the WCDVS version of the LSC-R, on study participants' tolerance of the instrument, and on the breadth and scope of potentially traumatic experiences in the past and current lives of women in the study sample. The development, psychometric properties, and use of summary measures are described, and recommendations are presented for the improvement of the WCDVS version of the LSC-R for use in future research. A literature review focused on issues and challenges in the measurement of stressful and violent events sets the stage by describing the context of the WCDVS's selection of the LSC-R as the instrument with which to assess past and current exposure to potentially traumatic events.

Measuring Traumatic Events

Some of the questions faced by WCDVS investigators in their choice of a measure of traumatic events were: Should the focus be on stressful life events or be limited to the formal assessment of traumatic events? What events should be included? Are self-reports of such events reliable (temporally stable)? How deeply should interviewers probe for details of the events? What is sufficient for research purposes, as opposed to clinical purposes? What meaningful and analytically useful summary variables can be computed from the basic information? What instrument is most appropriate to the target population?

Differences between stressful life events and traumatic life events

The *Diagnostic and Statistical Manual's* (*DSM*) definition of trauma has changed over time, as understanding of the scope and impact of trauma has grown. The initial definition was of an event that would be markedly distressing to almost anyone and outside the range of usual human experience. More recently, traumatic events have been expanded to include experiencing, witnessing, or being confronted with events that involve actual or threatened death or serious injury, or a threat to the physical integrity of oneself or others. The latest definition also added the necessary condition of an emotional response of fear, helplessness, or horror at the time of the event. Whatever the definition, traumatic events are considered to be the cause of the development of subsequent post-traumatic reactions, and the current diagnostic procedure for post-traumatic stress disorder (PTSD) requires a formal assessment of trauma, based on the definition in *DSM-IV* (Criterion A: threat of death or serious injury and emotional response of fear, helplessness, or horror at the time of the precipitating event).

The framers of the original criteria for the PTSD diagnosis had in mind events such as war, torture, rape, and natural and man-made disasters. Traumatic events were considered different from

*This study was funded under Guidance for Applicants (GFA) No. TI 00-003, titled Cooperative Agreement to Study Women with Alcohol, Drug Abuse and Mental Health (ADM) Disorders who have Histories of Violence: Phase II from the Department of Health and Human Services, Public Health Service, Substance Abuse and Mental Health Services Administration's 3 centers: Center for Substance Abuse Treatment, Center for Mental Health Services, and Center for Substance Abuse Prevention (March 2000).

the painful and stressful events that constitute the normal vicissitudes of life, such as divorce, loss, serious illness, and financial misfortune. Stress research has been primarily concerned with the scientific exploration of a causal relationship between life stress and illness. Early studies focused on the characteristics of events that were likely to be stressful and to lead to changes in psychiatric or physical health.[1] A parallel view of stressful life events emerged through the lens of bereavement, loss, and adaptation. Events such as divorce or death of a loved one were understood as rendering obsolete the individual's assumptive world and requiring a psychosocial transition that could generate internal turmoil, denial, and depression.[2,3] This distinction between stressful life events and traumatic events assumes that most individuals can cope with ordinary stress, whereas the adaptive capacities of most people are likely to be overwhelmed when confronted by a traumatic event.

Self-report

Self-report is often the only possible source of information on past or current experience of stressful events. The subjectivity of self-report of traumatic life events is often considered a hindrance to valid measurement. The recent discourse about "false memories" manifests the multifaceted problems involved in assessing the relative accuracy of self-reports, especially when they involve the report of past sexual or physical abuse.[4] Nonetheless, recent studies examining the use of self-report measures of trauma in psychiatric populations have concluded that they were reliable and valid.[5,6]

Which traumatic events to measure?

The WCDVS was designed with the understanding that trauma is often at the core of co-occurring problems of substance use and mental illness. Traumatic events may be brief, single incidents or prolonged and repeated. They may have occurred at any point in the lifespan, and, once begun, they increase vulnerability to additional trauma. In making a choice of trauma history assessment, WCDVS investigators wished to include life events that may not necessarily meet *DSM-IV* criteria for trauma but which were highly stressful nonetheless. The view adopted was that trauma is a complex, multifaceted experience that can result in adverse outcomes well beyond those of diagnosable PTSD. Thus, a measure was sought that was comprehensive and appropriate to the experiences of women.

Gender issues in the assessment of trauma history

Three general factors affect the sensitivity of trauma history measures with respect to gender: (1) the extent to which trauma exposure is queried in language that respondents understand easily and are willing to endorse, (2) the extent to which specific characteristics of traumatic events are measured, and (3) the inclusiveness of events or experiences examined. The importance of using language that respondents are willing to endorse became apparent when women with sexual experiences that met the legal definition of rape did not label their experiences as such and did not endorse items such as "Have you ever been raped?"[7,8] Characteristics such as age at the time of the event, severity, and chronicity are especially important with respect to gender, because they define the parameters of exposure that explain several gender differences in PTSD prevalence and comorbid symptoms. Gender-sensitive measures also include content relevant to the stressors and traumatic experiences that are specific to, or more common for, women or men.

Instrument choice

The Life Stressor Checklist-Revised (LSC-R) was chosen for the WCDVS because it is specifically tailored to the trauma exposure and stressful life experiences of women and has established validity.[9] The LSC-R is a 30-item instrument that includes stressors relevant to the lives of women who

do not usually meet *DSM* criteria for a traumatic event but may be relevant to understanding the context of trauma exposure, such as prolonged and unwanted separation from children or caregiving for someone ill or disabled. It also includes unique assessments for abortion and miscarriage, and differentiates forced sexual touching (sexual assault) from forced penetrative sex (rape). The LSC-R uses behaviorally specific language, and it includes probes to assess age at the time of the first event, chronicity, and *DSM-IV* criteria for life threat and emotional response. Information regarding the relationship to the perpetrator (someone known well versus not known well) is imbedded in the item wording when relevant. The LSC-R has demonstrated good criterion-related validity for PTSD in diverse populations of women and in several languages.[10–12]

Methods

Description of the WCDVS

The primary aim of the WCDVS was to evaluate the effectiveness of interventions that provide comprehensive, integrated, trauma-informed, and consumer-involved services to women who have co-occurring substance use and mental health disorders and a history of interpersonal violence.[13] The WCDVS used a quasi-experimental, longitudinal design with a common, standardized interview at baseline and 4 follow-ups: at 6 and 12 months to obtain descriptive and outcome data, and at 3 and 9 months to obtain interim services utilization data. The WCDVS was conducted at 9 sites across the country, where women in intervention services were compared with women in services as usual. Although the sites had common core service elements to provide integrated trauma, mental health, and substance abuse treatments, they differed by portal (substance abuse, mental health, or community-based) and modality (outpatient, inpatient, or mixed). Women aged 18 or older were eligible for the WCDVS if they had a substance use disorder *and* a mental health disorder, had accessed the formal treatment system on at least 2 occasions, and had a history of interpersonal abuse.

The eligibility criteria were very broad, even though they included both a substance use disorder and a mental health disorder. Only one of these disorders had to be current at the time of study entry; the other could be current or within the past 5 years. In terms of substance use disorders, a woman at study entry might have been just out of detoxification services or in recovery for nearly 5 years. In terms of mental health disorders, a woman at study entry might have been seeking services for untreated psychological symptoms or she might have been symptomatically stable. WCDVS women could have any major mental health disorder, including not only PTSD and depression but also anxiety, bipolar, multiple personality, obsessive-compulsive, personality, and schizophrenia spectrum disorders.

Modifications to the LSC-R

A workgroup of researchers, clinicians, and consumer-survivor-recovering women (CSRs) modified the LSC-R for the WCDVS to tailor it specifically to the study population. The original authors of the LSC and LSC-R had already addressed content validity,[9] but the workgroup felt that additional modifications were needed for the target population and the aims of the WCDVS. CSRs stressed the importance of gender specificity, cultural sensitivity, and caution regarding questions about abuse. They also expressed concern about potential distress to respondents as well as to interviewers and contributed to the common cross-site interviewer training and to the design of supports for interviewers. For example, CSRs endorsed separating death of a child from death of others close to the respondent and including a question about abuse by service providers. Their recommendations also led to the development of probes for frequency and age of onset of abuse that provided sufficient data for research purposes while maintaining a clinically sensitive approach and not prying too deeply into the details of specific stressful events.

The modified LSC-R, henceforth referred to as the WCDVS version of the LSC-R, contained 31 specific items and a final, open-ended item that asked about "any other stressful events." Eighteen items in the WCDVS version were identical to those in the original LSC-R, 6 items resulted from either merging or splitting LSC-R items, and 7 items were new (Table 1). Each item inquired about lifetime exposure to a stressful event (yes or no; "has this ever happened"), and, if positively endorsed, about current exposure (yes or no; "has this happened in the past 6 months"). There were no further probes for the first 17 items. The next 14 items, which focused on interpersonal abuse and neglect, included follow-up probes concerning frequency and age of onset. The frequency probe, "How often has this happened," had 3 response options: once, a few times, and a lot. The age of onset probe, "How old were you when this (first) happened," had 5 response options: 0–5 years, 6–10 years, 11–13 years, 14–17 years, and 18 years or older, which were also anchored by school-based developmental periods (preschool, elementary school, junior high or middle school, high school, beyond high school). As a result of these modifications and unlike the original LSC-R, the WCDVS version of the LSC-R does not assess trauma formally (ie, using *DSM-IV* Criterion A). Instead, it provides a comprehensive and sensitive assessment of stressful life events, many of which are presumed to lead not just to post-traumatic reactions but also to other Axis I or II mental health disorders such as depression, generalized anxiety, dissociative identity disorder, and substance use disorders.

The WCDVS version of the LSC-R was embedded in a comprehensive interview that included measures of personal and family characteristics, substance abuse and mental health treatment history, patterns of substance use, general mental health symptoms, PTSD symptoms, service utilization, and perceptions of care. Common training in standardized research interviewing and study-specific procedures was provided for interviewers from all study sites.

Retest sample

A subset of women completed the cross-site baseline interview on 2 separate occasions, close in time, to enable analysis of test-retest reliability (temporal stability). The retest sample numbered 186, which was approximately 20 women from each study site. For those women who completed the baseline interview in 1 session ($n = 174$), the retest interview occurred an average of 7 days later (SD = 4.2; Range = 2–35 days). For the remaining 12 women, who required 2 sessions for their baseline interview, the time between the second session and the retest interview was also 7 days (SD = 4.8; Range = 1–20 days).

Survey of sites concerning LSC-R tolerance

A concern shared by CSRs, clinicians, and researchers was the potential of the trauma history assessment to elicit distress and to trigger unwanted emotional reactions. To assess how well participants in the WCDVS tolerated the trauma history assessment, we analyzed the responses from the study sites to 4 questions addressing (1) research interviewers' opinions of women's reactions, both positive and negative; (2) whether interviewers noticed any patterning in women's reactions; (3) what it was like for interviewers to administer the WCDVS version of the LSC-R; and (4) whether there were any adverse events associated with it.

Summary variables

A WCDVS workgroup discussed ways in which information from the WCDVS version of the LSC-R could be used to create summary variables for use as measures of (1) individual differences in *lifetime* trauma history, for description and as covariates in statistical analyses, and (2) individual differences in *current* exposure to traumatic events, as covariates in statistical analyses and as secondary outcome measures.

Table 1

Item response frequency and test-retest reliability for the WCDVS version of the Life Stressor Checklist-Revised (LSC-R)*

Stressful life events	Ever experienced			Experienced within past 6 months		
	Frequency "yes" (%) ($N = 2729$)	κ ($n = 186$)	Absolute agreement, % ($n = 186$)	Frequency "yes" (%) ($N = 2729$)	κ ($n = 186$)	Absolute agreement, % ($n = 186$)
1. Been in a *serious* disaster?†	804 (29.5)	0.75	89.7	54 (2.0)	0.39	98.4
2. Had a serious accident or an accident-related injury?†	1449 (53.1)	0.60	79.6	116 (4.3)	0.32	97.9
3. Close family member ever sent to jail?†	1817 (67.0)	0.68	84.8	541 (20.0)	0.57	89.1
4. Been sent to jail or juvenile detention?†	1909 (70.0)	0.85	92.5	714 (26.2)	0.85	94.6
5. Ever put in foster care or put up for adoption?†	566 (20.8)	0.84	94.6	N/A	N/A	N/A
6. Did parents separate or divorce while you were living with them?†	1613 (59.4)	0.80	90.3	N/A	N/A	N/A
7. Been separated or divorced?†	1431 (52.6)	0.85	92.4	316 (11.6)	0.76	95.1
8. Been homeless?†	1959 (71.8)	0.82	92.5	996 (36.6)	0.47	89.8
9. Have serious money problems?†	2374 (87.1)	0.67	94.1	1772 (65.2)	0.63	86.3
10. Have a very serious physical or mental physical mental illness?†	1657 (60.7)	0.57	78.5	939 (34.5)	0.60	81.7
11. Had an abortion?‡	1516 (55.6)	0.95	97.3	65 (2.4)	0.56	98.4
12. Had a miscarriage?‡	1209 (44.4)	0.97	98.4	82 (3.0)	0.59	97.8
13. Been separated from child(ren) against your will?†	1646 (60.9)	0.74	86.9	820 (30.4)	0.76	91.2
14. Baby or child of yours have a severe physical or mental handicap?†	374 (13.9)	0.64	92.8	141 (5.2)	0.61	96.7
15. Death of child?§	228 (9.7)	0.84	97.5	13 (0.6)	N/A	100.0
16. Been responsible for taking care of someone close to you (*other than your child*) who had severe physical or mental handicap?†	1089 (39.9)	0.67	84.4	284 (10.4)	0.73	95.7
17. Anyone close to you (*other than your child*) ever died?‡	2340 (85.8)	0.59	88.7	489 (18.0)	0.70	91.4
18. Before age 18, see physical violence between family members?†	2046 (75.1)	0.64	88.2	N/A	N/A	N/A
19. Been emotionally abused or neglected?†	2290 (84.0)	0.56	87.1	879 (32.4)	0.56	82.3

20. Been physically neglected?†	882 (32.4)	0.66	85.5	162 (6.0)	0.42	92.5
21. Been physically abused—for example, hit, choked, burned, or beaten—or severely punished—for example, locked up, shut in a closet, tied up, or chained—by someone you knew well such as a parent, sibling, boyfriend, or girlfriend?‡	2316 (84.9)	0.52	86.0	486 (17.9)	0.61	89.7
22. Been robbed, mugged, or physically, *not sexually*, attacked by a stranger or someone you did not know well?†	1297 (47.6)	0.64	82.8	211 (7.7)	0.34	88.7
23. Seen a robbery, mugging, or attack taking place?†	1388 (51.1)	0.72	86.0	256 (9.4)	0.38	91.4
24. Been stalked or had anyone threaten to kill or seriously harm you?§	1946 (71.4)	0.57	81.7	457 (16.8)	0.64	92.4
25. Been strip searched, forcibly restrained, or held against will by a provider of MH or SA services?§	568 (20.8)	0.79	93.6	174 (6.4)	0.64	94.6
26. Discriminated against because of race, ethnic group, gender, sexual orientation, or religion?§	912 (33.5)	0.62	83.9	369 (13.6)	0.63	93.6
27. Been the victim of a hate crime? Have violence directed at you because of your race, ethnic group, gender, sexual orientation, or religion?§	409 (15.0)	0.71	93.0	97 (3.6)	0.48	96.8
28. Been bothered or harassed by sexual remarks, jokes, inappropriate touching or demands for sexual favors by someone at work or school?†	1143 (41.9)	0.65	82.8	248 (9.1)	0.66	94.6
29. Touched or made to touch someone else in a sexual way because you felt forced in some way or threatened by harm to yourself or someone else?‡	1827 (67.1)	0.56	80.1	243 (9.0)	0.57	92.9
30. Have sex because you felt forced in some way or threatened by harm to yourself or someone else?‡	1975 (72.6)	0.63	85.0	275 (10.2)	0.75	95.7
31. Have unwanted sex in exchange for money, drugs, or other material goods such as shelter or clothing?§	1545 (56.7)	0.87	93.5	480 (17.6)	0.71	90.2

*WCDVS indicates the Women, Co-occurring Disorders, and Violence Study; N/A, not applicable.
†Relationship to original LSC-R: Same as, or similar to, original.
‡Relationship to original LSC-R: Split or merged from original.
§Relationship to original LSC-R: New for WCDVS version.

Five summary variables resulted: *Lifetime exposure to stressful events* (LESE) uses information from all 31 specific items to indicate the breadth of exposure to stressful events in each woman's lifetime. A score on the LESE is equal to the total number of items endorsed (Range = 0–31); *Lifetime frequency of interpersonal abuse* (LFIA) uses follow-up probe information from 9 items (Table 1: items 18, 19, 20, 21, 24, 25, 29, 30, 31) to quantify the frequency of interpersonal abuse in each woman's lifetime. Respondents can answer "never" (0), "once" (1), "a few times" (2), or "a lot" (3) to each probe about the frequency of the abuse. Three of the 9 items (21, 29, 30) ask about frequency of abuse in both childhood and adulthood, and thus 12 responses are summed (Range = 0–36); *Frequency of childhood abuse* (FCA) is formed by summing the responses to 3 frequency probes pertaining to childhood physical and sexual abuse (Table 1: items 21, 29, 30); *Current exposure to interpersonal abuse* (CEIA) consists of the summed responses ("no" = 0; "yes" = 1) to 8 items pertaining to recent exposure to interpersonal abuse (Table 1: items 19, 20, 21, 24, 25, 29, 30, 31); and *Current exposure to other stressors* (CEOS) sums the responses to the 20 recent-exposure items that are not included in the CEIA scale (Range = 0–20). In addition, the workgroup developed rules for combining item and probe responses to produce 4 dichotomous variables that indicated the presence or absence of childhood physical abuse, childhood sexual abuse, adulthood physical abuse, and adulthood sexual abuse.

These summary variables assume that a high frequency of trauma exposure and a wider exposure are indications of increased severity, but they intentionally refrain from assigning weights to different types of experiences. Although several factors other than frequency have been proposed to relate to the severity of trauma exposure, for example, age of onset and relationship to the perpetrator, the field is still far from generating an accepted severity algorithm. The purpose here is not to codify these summary variables, but rather to suggest possible ways to combine single-item data into meaningful composites and to examine their psychometric properties. Despite serious limitations of summary measures when comparing individual women, there is a clear need for aggregate trauma history variables for statistical analyses of group differences.

Qualitative analysis of the final open-ended item

The Trauma Measures Qualitative Workgroup analyzed the responses to the final item of the WCDVS version of the LSC-R ("Are there any other upsetting or stressful events we did not include that you would like to mention? What was the event?").* Members of the workgroup represented CSR, clinical, and research perspectives and included 2 WCDVS interviewers.

The workgroup combined 2 different methodologies. Initially, analysis of responses was guided solely by grounded theory, according to which themes arise from the data rather than being imposed on it by the analyst.[14] As workgroup members identified themes in the baseline responses, they

---

*The authors acknowledge their debt to all the members of the Trauma Qualitative Workgroup, which was responsible for analyzing study participants' responses to the final, open-ended question of the WCDVS version of the LSC-R. In alphabetical order, members, followed by their WCDVS role, site or affiliation, and location, are: Paula Bjelajac (CSR and clinician, Prototypes, Culver City, Calif), Catherine Coughlan (interviewer, ALLIES, Stockton, Calif), Elizabeth Jackson (researcher, WCDVS Coordinating Center, University of North Carolina, Chapel Hill, NC), Nina Kammerer (researcher, WELL Project, Cambridge, Mass), Ruta Mazelis (CSR consultant, WCDVS Coordinating Center, Policy Research Associates Inc, Delmar, NY), Debra Kram-Fernandez (researcher, Portal Project, New York, NY), Lisa Miller (researcher, Boston Consortium for Families, Boston, Mass), Susan Mockus (CSR consultant, TAMAR Project, Baltimore, Md), Terri Nadlicki (researcher, WCDVS Coordinating Center, University of North Carolina, Chapel Hill, NC), Andrea Savage (researcher, Portal Project, New York, NY), Debra Wagler (interviewer, ALLIES, Stockton, Calif), and Wendy Vogel (researcher, WCDVS Coordinating Center, Policy Research Associates Inc, Delmar, NY). Nina Kammerer drafted the workgroup's contributions to this article and revised them on the basis of feedback from workgroup members. She thanks MeeLee Kim, Carol Prost, and Dominique Simon for illuminating conversations and editorial suggestions. Limitation on the number of authors prevented officially including all members of the workgroup as authors.

recognized that some responses described events that were covered in one of the preceding specific items. In addition, they recognized that many of the themes identified were closely related to the events covered in those items. Therefore, a second analytic strategy was employed, which derived themes or codes from categories in the specific items. Combining this top-down approach with the bottom-up approach of grounded theory resulted in a tripartite coding structure: Coding I, themes that are the same as those in the 31 specific items; Coding II, themes that are similar to those in the specific items; and Coding III, themes that are new and different from those in the specific items.

Results

Characteristics of the WCDVS study participants

The WCDVS baseline sample ($N = 2729$) represents the largest study sample of dually diagnosed treatment-seeking women who have completed a common trauma history assessment. A description of the characteristics of these women sheds light on the generalizability of the WCDVS findings on the past prevalence and 6-month incidence of stressful events in women's lives, as well as on tolerance of the WCDVS version of the LSC-R.

The WCDVS participants were, on average, 36 years old, ranging from 18 to 76. Fifty percent had completed high school, and another 24% had a Graduation Equivalency Diploma (GED). Women self-identified as white/Caucasian (54%), black/African American (29%), and Hispanic/Latina (18%). More than one third (38%) were married or partnered, while almost one third (30%) had never married. These women had, on average, $682 during the past 30 days, with a range from $0 to $30,000; 12.6% were employed currently; and 71.8% had been homeless sometime in their life. Most of the women had had children (86.7%), but only 31.7% were living with a child younger than 18 years (See McHugo et al[13] and Becker et al[15] for fuller descriptions). In addition, study participants lived in urban, rural, and suburban settings in 6 states and the District of Columbia.

Reactions to the WCDVS version of the LSC-R

When asked about respondents' reactions to the trauma history assessment, interviewers reported that most women either had little reaction or found it a positive experience. Women stated that the questions gave them perspective on how much they had endured and survived, or how fortunate they were to have avoided some of the events asked about. A number of women said that they appreciated being asked directly about stressful experiences in their lives and that they had told interviewers things that they had never disclosed to anyone else. For the few women who appeared distressed during the administration of the WCDVS version of the LSC-R, it was expressed as nervousness or tenseness, particularly at the baseline interview, or as tearfulness or weeping. Interviewers noted that women's reactions appeared to be related to their stages of recovery (eg, length of sobriety) and therapeutic experiences (eg, how often they had described their trauma experiences in clinical settings). Those few women who struggled with the trauma assessment tended to be more fragile before the interview than women who did not, and women who had experienced interpersonal violence during the preceding 6 months showed greater emotion. No interviewer reported that any woman refused the trauma history assessment. Only 1 site reported adverse events related to the WCDVS version of the LSC-R to their institutional review board (IRB); 3 such events were reported; none had long-term negative consequences. Another site referred 2 women for clinical services as a result of minor distress expressed during the trauma assessment. Thus, out of roughly 6000 baseline and follow-up interviews, there were only a handful of cases where special considerations were necessary.

Interviewers reported that they appreciated that women were not asked to tell their stories or probed for details, and that the items progressed from less intrusive to more intrusive events. Some interviewers said that it was difficult to hear about the traumatic experiences of the women they

interviewed, but as they became more experienced, they felt that the interviews became easier to administer. Although avoiding the details of traumatic events is protective of respondents and interviewers, a few interviewers felt uncomfortable asking women about such events without giving them an opportunity to tell their stories.

Lifetime and current experiences of stressful events

Table 1 presents the frequency statistics for the 31 specific items of the WCDVS version of the LSC-R. The lifetime frequencies from the baseline interview ($N = 2729$) range from 9.7% for the death of a child to 87.1% for money problems. Other low-frequency ($\leq$20%) lifetime events were having a handicapped child and being the victim of a hate crime. Other high-frequency ($\geq$80%) lifetime events were loss of a loved one, emotional abuse, and physical abuse. Some of these exposures occurred in the past 6 months. The current frequencies range from 0.6% for death of a child to 65.2% for money problems. Other low-frequency ($\leq$5%) current events were experiencing a disaster, being in a serious accident, having an abortion or a miscarriage, and being the victim of a hate crime. Other high-frequency ($\geq$30%) current events were homelessness, having a serious physical or mental illness, becoming separated against her will from a child, and emotional abuse.

Test-retest reliability of the WCDVS version of the LSC-R

Table 1 also presents the test-retest reliability statistics for the specific items, based on the retest sample ($n = 186$). The percent absolute agreement is presented to indicate obtained levels of agreement, and κ, which reflects the chance-corrected level of agreement, is presented as the index of reliability. All of the lifetime items exceed published criteria for acceptable reliability based on κ ($\geq$0.40), and many show high levels of agreement between test and retest.[16] In general, test-retest reliability based on κ was lower for the current items, although the percent absolute agreement was seldom lower than 90%, a discrepancy that is likely due to low base rates.[17] Kappa for 4 items was below 0.40, but most were moderate to high, indicating good test-retest reliability for reports of recent stressful events.

Open-ended questions are not designed for test-retest reliability, and no standard method of calculating their reliability exists. Yet there was interest in examining the correspondence between test and retest responses to the final item. In the test-retest sample of 186 women, 134 (72%) did not answer this item at either administration. Of the 52 women who answered it at one or both administrations, the pattern of responses was 20 at both, 16 at test only, and 16 at retest only. Of the 20 who answered at both administrations, 9 gave responses with identical content both times, 7 gave responses that overlapped partially, and 4 gave responses that were completely different.

Summary variables

The distributions of the lifetime summary variables (LESE, LFIA, FCA) had good properties, with means centered within the range and good variability (Table 2). The intraclass correlation coefficient was computed as an index of test-retest reliability for continuous measures.[18] The lifetime summary variables showed high test-retest reliability. On average, women in the WCDVS reported exposure to about half of the 31 stressful events in their lifetimes (LESE mean = 16.32, with a range from 2 to 30). The scales for current exposure (CEIA and CEOS) had lower test-retest reliability, although it was still acceptable. On average, the women reported about 1 exposure to interpersonal abuse and about 3 exposures to other stressors during the past 6 months. Table 3 presents the frequencies and the test-retest statistics for the 4 indicator variables. The frequencies indicate very high rates in all 4 abuse categories among the women in the WCDVS study group, and test-retest reliability (κ) is moderate to high for these 4 indicator variables.

Table 2

Descriptive and test-retest (intraclass correlation coefficient) statistics for the summary variables from the WCDVS version of the Life Stressor Checklist-Revised*

Summary variable	Mean	SD	Range	Intraclass correlation coefficient
Lifetime exposure to stressful events (LESE)	16.32	4.57	2–30	0.86
Lifetime frequency of interpersonal abuse (LFIA)	15.86	7.39	0–36	0.88
Frequency of childhood abuse (FCA)	3.72	2.99	0–9	0.86
Current exposure to interpersonal abuse (CEIA)	1.16	1.54	0–8	0.77
Current exposure to other stressors (CEOS)	3.14	1.99	0–16	0.77

*WCDVS indicates the Women, Co-occurring Disorders, and Violence Study.

Final item: Coding I and II

The item-by-item specifications for the WCDVS comprehensive interview instructed interviewers to record the response to the open-ended item verbatim and to record the event even if it fits into an earlier question. At baseline, 583 (21.36%) of 2729 women described "other" events.

There are several reasons why women may have reported events appropriate to the specific items in response to the final, open-ended item. One reason is not understanding the close-ended item that covered their response. Some instances of differential understanding may be due to the structure of 14 items, which contain a parenthetical clarification that was read only if a woman asked about the question's meaning. For instance, the parenthetical to the item about ever having a serious physical or mental illness includes "tried to kill yourself" as an example, but some women's responses to final item mention their suicidality or suicide attempts. The final item also offered some women an opportunity to tell their "story" by naming or recounting traumatic events that they had experienced. Other women may have described an event because they felt that its emotional weight, or some nuance of its meaning or nature, was not fully captured by any of the preceding questions. Some of the responses to the final item suggest that women were conveying the complexity or multifaceted nature of the event or series of events, which are qualities not captured in the close-ended questions.

Many responses to the final item concerned experiences closely related to those covered by the specific items, but they included additional dimensions. Examples include sexual harassment outside the context of work or school, and harassment that was not experienced as sexual in content. Two questions for which additional dimensions were frequent were death of a child ($n = 8$; 1.4% of the 583 responses) and death of someone close to the respondent other than her child ($n = 73$; 12.5%

Table 3

Descriptive and test-retest statistics for the indicator variables based on the WCDVS version of the Life Stressor Checklist-Revised*

Indicator variable	Frequency "yes" (% yes)	κ	% absolute agreement
Childhood sexual abuse	1688 (62.0)	0.76	88.7
Childhood physical abuse	1696 (62.2)	0.67	84.4
Adulthood sexual abuse	1637 (60.3)	0.69	85.5
Adulthood physical abuse	2195 (85.4)	0.51	86.3

*WCDVS indicates the Women, Co-occurring Disorders, and Violence Study.

of the 583 responses). These additional dimensions included death by suicide, murder, accident, or disaster; death that the respondent was present at or witnessed; multiple losses; and death followed by the respondent finding or seeing the body.

Many responses to the final item were similar to the two specific items concerning sexual abuse (Table 1: items 29 and 30). Because some women may not attach the word "rape" or "incest" to their experiences, neither word was included in these specific items, but without the inclusion of these words, some women found that these items did not capture their experiences. In the words of one woman when asked about other stressful events, "I consider my rape and kidnapping something you didn't specifically get to." A number of women mentioned incest or molestation by a relative in response to the final item. The wording of the specific question about forced penetrative sex (Table 1: item 30; "Did you ever have sex...?") may have implied active participation on the respondent's part that she did not feel occurred. An interviewer noted that a respondent identified a rape in the final item but did not include it in response to the specific item, because "that question was about having sex when you did not want to, not rape." These events may also have been reported in the final item because respondents did not consider them to have involved either the force or threat that is explicit in the specific item, or they did not consider forced penetration to be "sex."

Final item: Coding III

Using grounded theory, the Trauma Measures Qualitative Workgroup identified a number of themes that are either not among or not closely related to the events covered in the specific items. These themes include (1) a number of events or experiences of the respondent (eg, risky sexual behavior, self-harm), (2) experiences of the respondent's children (eg, respondent abused her children, someone else abused her children, respondent abandoned her children), (3) other violence perpetrated by the respondent (eg, respondent killed someone), (4) issues related to the respondent's natal family (eg, mental illness of a parent), (5) the respondent being falsely accused or not believed, and (6) the respondent keeping secrets or fearing negative consequences of telling the truth. A stressful personal experience that 18 of 583 (3%) reported in answer to the final question was being kidnapped or abducted.

Discussion

The findings indicate that the WCDVS version of the LSC-R was well received by respondents and interviewers, asks about stressful life events that are common to women in the target population, and has good test-retest reliability (temporal stability). The relevance of the items to the experiences of the women can be inferred from the high rates of endorsement of single items and membership in all 4 abuse categories. The test-retest reliabilities for single items were similar to those found in other studies with similar populations.[5,6] The summary variables were also temporally stable (high test-retest reliability) and had good measurement properties. By combining quantitative and qualitative information, the WCDVS findings document not only the prevalence of these stressful events but also the complexity of the lives of women with co-occurring disorders. Most responses to the final item concerned events covered in specific items or captured additional dimensions of an event covered in a specific item. If anything, therefore, the responses to the final, open-ended question suggest that some specific items are undercounted rather than overcounted.

All women in the WCDVS reported, when being screened for eligibility, that they had accessed services for substance use or mental health problems on at least 2 previous occasions. The duration of each previous contact could have been as short as a single office visit. All women in the study also endorsed an item on the common eligibility screening instrument that asked if she had ever been physically or sexually abused. This does not mean that the women in the study had previously been assessed for trauma or had previously received counseling in any format for their trauma. Information from site interviewers and clinicians and from responses to the open-ended question

in the WCDVS version of the LSC-R indicated that some of the women had never revealed their traumatic experiences prior to their participation in the WCDVS. Even more had never received treatment that addressed their trauma prior to the WCDVS.

Thus, women in the WCDVS represent a large, demographically and geographically diverse sample with varying degrees of current substance use and mental health symptoms and varying frequencies of past and current stressful life experiences. In addition, they had received and were receiving varying amounts of services, including treatment for trauma, in a variety of settings. This suggests that findings concerning the lifetime prevalence and the past 6 months' incidence of stressful events are broadly generalizable to treatment-seeking women aged 18 and older who have co-occurring substance use and mental health disorders and a history of interpersonal violence. Despite convenience, clustered sampling in WCDVS, the broad eligibility criteria, the large sample size, and the site diversity support the wide applicability of the WCDVS version of the LSC-R for the assessment of trauma among women accessing public-sector services for substance abuse and mental health treatment and rehabilitation.

Overall, WCDVS data on the LSC-R support the conclusion of Newman and colleagues that women "expressed few adverse reactions to the inquiry, and in many cases, derived benefit from participation in the study."[19] The WCDVS version of the LSC-R was well tolerated by participants and liked by interviewers. There were many reports of positive reactions to it, and there were very few adverse reactions. Diversity among the WCDVS study participants in terms of severity of previous and current trauma, whether or not they had ever been assessed for trauma or previously revealed their stressful life experiences ("told their story"), and intensity of both prior and current treatment suggests that the WCDVS version of the LSC-R would be well tolerated by most women with trauma histories. Because researchers and IRBs are obligated to follow the principals of nonmalfeasance and beneficence, it is important to examine tolerance of trauma history assessments such as the WCDVS version of the LSC-R. The data gathered through the survey of interviewers indicate the value of multiple strategies for collecting information on study participants' tolerance of trauma history assessment. The WCDVS participant interviews would have benefited from the inclusion of self-report items on reactions to the trauma assessment.

Six recommendations for the improvement of the WCDVS version of the LSC-R emerged from the Trauma Measures Qualitative Workgroup's analysis of baseline responses to the final item. *First*, parenthetical examples and clarifications contained in specific items should be either dropped or included directly in the questions. To prevent respondents from taking examples as exhaustive rather than illustrative, they should be preceded by phrasing such as, "Some examples of the kinds of events this question covers are . . ." *Second*, based on the frequency of open-ended responses that noted added dimensions of the death of a child or of someone else close to the respondent, these dimensions should be added as specific probes following an affirmative answer to either question. The *third* recommendation concerns one of the specific questions about physical abuse (Table 1: item 21). Some instances of physical abuse in which the perpetrator was not well known to the respondent were missed because physical abuse by a stranger was included with being robbed or mugged. Separating physical abuse by a stranger from a robbery or mugging would elicit a broader range of reported events.

The *fourth* recommendation is that the scope of several specific questions should be expanded. The specific item about the separation or divorce of parents pertains only to legally married parents. Similarly, another specific item asks whether the respondent has "ever been separated or divorced?" Some women who answered either of these items in the affirmative may have been counting partnerships that were not legal, whereas other women may have saved such instances for the final item. The specific item about serious money problems includes the example of not having enough money to pay the rent, but it does not include not having enough money to pay the mortgage. The specific item about being "strip searched, forcibly restrained, or held against your will by a provider of mental health or substance abuse services" could include representatives of any sort of service or

official agency, including corrections officers and members of the clergy. Women's responses to the final item also indicated other reasons for discrimination and other contexts for sexual harassment.

The *fifth* recommendation is to add a specific question about kidnapping and abduction, whether by a stranger or someone known well. *Finally*, the high frequency of responses to the final item that were coded as sexual abuse indicates that these 2 specific items require revision (Table 1: items 28 and 29). Although not using the terms "molestation," "childhood sexual abuse," "rape," and "incest" in specific items may prevent undercounting among women who do not apply them to their own experiences, it may lead to undercounting among women who do. As one woman said in response to the final, open-ended item at the 12-month follow-up, "I think rape should be a specific question . . ."

On the basis of these recommendations and other experiences during the WCDVS, a workgroup is revising the WCDVS version of the LSC-R further to increase the clarity and scope of the items, and, consequently, to increase reliability and validity (content validity and construct validity). Copies of the revised WCDVS version of the LSC-R will be available from the original author.*

Implications for Behavioral Health

The WCDVS version of the LSC-R provides researchers and service providers with a trauma history assessment that is especially appropriate for women who have substance use and mental health problems and who have a history of interpersonal violence and abuse. This version of the LSC-R is appropriate for research where a trauma assessment is needed that is comprehensive and has good psychometric properties but is not distressing to respondents. The summary variables used in the WCDVS, or variations on them, may be especially useful for research where the specific details of trauma history are less important than global measures of exposure and severity. The WCDVS version of the LSC-R is also appropriate as an initial assessment of trauma history in clinical settings, where in-depth follow-up assessments would be used to determine post-traumatic diagnoses or to prepare for trauma-specific treatments. In addition, the findings from this study can be used to reassure IRBs and other stakeholders concerning the minimal risks associated with the assessment of trauma history. More broadly, the WCDVS highlighted the value of involving consumers in developing and tailoring assessments to special populations. It also showed that trauma history assessment can be tolerated well, and even regarded positively, despite the vulnerabilities of the target population. As service systems and single agencies move toward trauma-informed and trauma-specific services, the need to assess both past and recent trauma history will increase accordingly. The WCDVS version of the LSC-R offers a safe and reliable assessment of trauma exposure and life stressors, which will enable improved treatment and rehabilitation strategies and lead to better outcomes for women with co-occurring disorders.

References

1. Mueller D, Edwards D, Yarvis R. Stressful life events and psychiatric symptomatology: change or undesirability? *Journal of Health and Social Behavior*. 1977;18:307–317.
2. Parkes C. Psychosocial transitions: a field for study. *Social Science and Medicine*. 1971;5:101–115.
3. Parkes C. What becomes of redundant world models? A contribution to the study of adaptation to change. *British Journal of Medical Psychology*. 1975;48:131–137.

---

*Copies of the Life Stressor Checklist-Revised (LSC-R) and the WCDVS version of the LSC-R are available from Rachel Kimerling, PhD, National Center for PTSD, Palo Alto VA Health Care System (352–117MPD), Menlo Park, CA 94025. E-mail: Rachel.Kimerling@med.va.gov. Members of the Trauma Qualitative Workgroup are currently collaborating with Dr Kimerling on a revised WCDVS version on the basis of the findings of this study. When completed, this revised WCDVS version of the LSC-R will also be available from Dr Kimerling.

4. Briere J. Psychological assessment of child abuse effects in adults. In: Wilson JP, Keane TM, eds. *Assessing Psychological Trauma and PTSD*. New York: The Guilford Press; 1997:43–68.
5. Goodman LA, Thompson KM, Weinfurt K, et al. Reliability of reports of violent victimization and posttraumatic stress disorder among men and women with serious mental illness. *Journal of Traumatic Stress*. 1999;12:587–599.
6. Mueser KT, Salyers MP, Rosenberg SD, et al. Psychometric evaluation of trauma and posttraumatic stress disorder assessments in persons with severe mental illness. *Psychological Assessment*. 2001;13:110–117.
7. Kilpatrick DG, Saunders BE, Amick-McMullan A, et al. Victim and crime factors associated with the development of crime-related post-traumatic stress disorder. *Behavior Therapy*. 1989;20:199–214.
8. Koss MP. The hidden rape victim: personality, attitudinal, and situational characteristics. *Psychology of Women Quarterly*. 1985;9:193–212.
9. Wolfe J, Kimerling R. Gender issues in the assessment of posttraumatic stress disorder. In: Wilson JP, Keane TM, eds. *Assessing Psychological Trauma and PTSD*. New York: Guilford; 1997:192–238.
10. Brown PJ, Stout RL, Mueller T. Substance use disorder and posttraumatic stress disorder comorbidity: addiction and psychiatric treatment rates. *Psychology of Addictive Behaviors*. 1999;13:115–122.
11. Gavrilovic J, Lecic-Tosevski D, Knezevic G, et al. Predictors of posttraumatic stress in civilians 1 year after air attacks: a study of Yugoslavian students. *Journal of Nervous and Mental Disease*. 2002;190:257–262.
12. Kimerling R, Calhoun KS, Forehand R, et al. Traumatic stress in HIV-infected women. *AIDS Education and Prevention*. 1999;11:321–330.
13. McHugo GJ, Kammerer N, Jackson EW, et al. Women, Co-Occurring Disorders, and Violence Study: evaluation design and study population. *Journal of Substance Abuse Treatment*. 2005;28:91–107.
14. Strauss A, Corbin J. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, Calif: Sage; 1990.
15. Becker MA, Noether CD, Larson MJ, et al. Characteristics of women engaged in treatment for trauma and co-occurring disorders: findings from a national multi-site study. *Journal of Community Psychology*. In press.
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data models. *Biometrics*. 1977;33:159–174.
17. Grove WM, Andreason NC, McDonald-Scott P, et al. Reliability studies of psychiatric diagnosis. *Archives of General Psychiatry*. 1981;38:408–413.
18. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996;1:30–46.
19. Newman E, Walker EA, Gefland A. Assessing the ethical costs and benefits of trauma-focused research. *General Hospital Psychiatry*. 1999;21:187–196.

American Journal of Epidemiology Advance Access published July 27, 2005

American Journal of Epidemiology

Vol. 162, No. 5
Printed in U.S.A.
DOI: 10.1093/aje/kwi217

Temporal Trends in the Incidence of Intermittent Claudication from 1950 to 1999

Joanne M. Murabito[1,2], Jane C. Evans[1], Ralph B. D'Agostino, Sr.[1,3], Peter W. F. Wilson[4], and William B. Kannel[1]

[1] Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, MA.
[2] Section of General Internal Medicine, School of Medicine, Boston University, Boston, MA.
[3] Statistics and Consulting Unit, Boston University, Boston, MA.
[4] Departments of Endocrinology, Diabetes, and Medical Genetics, Medical University of South Carolina, Charleston, SC.

Received for publication January 13, 2005; accepted for publication April 7, 2005.

Declines in coronary disease and stroke mortality have occurred, but it remains unclear whether intermittent claudication (IC) incidence and mortality rates have changed. The authors sought to examine long-term trends for IC in the community. Cases of IC among Framingham Study participants aged ≥40 years were classified according to date of onset from the 1950s to the 1990s. IC was defined as the presence of exertional calf discomfort that was relieved with rest. Age- and sex-adjusted incidence rate ratios were estimated using log-linear Poisson regression, and 10-year survival was calculated using the Kaplan-Meier method. IC occurred in 668 participants (286 women). The age- and sex-adjusted incidence rate of IC fell from 282 per 100,000 person-years during the period 1950–1969 to 225 per 100,000 person-years in the 1990s. The decline in IC incidence across time periods was significant (p for trend = 0.01), with an initial increase in the 1970s being followed by declines of 16% in the 1980s and 18% in the 1990s. Approximately 40% of persons with IC died within 10 years of diagnosis, with no significant change occurring during the study period. IC incidence has declined since 1950, but mortality has remained high and unchanged. Factors contributing to the declining incidence of IC need clarification.

cardiovascular diseases; intermittent claudication; mortality

Abbreviations: CVD, cardiovascular disease; PAD, peripheral arterial disease.

Intermittent claudication is a symptomatic expression of peripheral arterial disease (PAD), a highly prevalent atherosclerotic condition affecting more than 5 million persons in the United States (1). Claudication is associated with increased risks of mortality, nonfatal cardiovascular diseases (CVDs) (myocardial infarction, congestive heart failure, and cerebrovascular disease), and impaired lower extremity function (2–6). All major CVD risk factors, including smoking, diabetes, hypertension, and an elevated cholesterol level, are associated with increased risk of intermittent claudication (7–9). It is unclear whether or not the incidence of intermittent claudication is declining in the US population in response to changes in smoking behavior and increased awareness and treatment of hypertension and elevated blood cholesterol (10–12).

Recent national data have shown a relatively low awareness of PAD among primary care physicians and less intensive treatments for modifiable risk factors in persons with the disease as compared with persons with other manifestations of CVD (13). Thus, given that secondary prevention efforts are utilized less in PAD patients, it is uncertain whether survival following the onset of intermittent claudication has improved in parallel with improvements in survival following the onset of other CVDs such as myocardial infarction and congestive heart failure (14, 15). Although antiplatelet drug therapies effectively reduce fatal ischemic events in patients with PAD (16), these therapies, as well as other secondary prevention efforts, appear to be underutilized in patients with the disease (13).

We examined temporal trends in intermittent claudication incidence and subsequent survival among participants in the Framingham Heart Study during the time interval from 1950 through 1999. Since its inception, the Framingham Heart

Correspondence to Dr. Joanne M. Murabito, Framingham Heart Study, 73 Mount Wayte Avenue, Suite 2, Framingham, MA 01702-5827 (e-mail: Murabito@bu.edu).

Study has used standardized criteria in the ascertainment of intermittent claudication, and the study cohorts have been monitored continuously with respect to vital status and the occurrence of CVD endpoints.

MATERIALS AND METHODS

Study sample

The Framingham Heart Study was established in 1948, enrolling 5,209 men and women aged 28–62 years in a prospective epidemiologic cohort study. Members of the original cohort have undergone follow-up examinations every 2 years. In 1971, 5,124 offspring of the original cohort members and offspring spouses ranging in age from 5 years to 70 years were enrolled in the Framingham Offspring Study. Follow-up examinations are conducted approximately every 4 years. Study design and entry criteria have been reported elsewhere (17–20). Written informed consent was obtained from all study participants, and the institutional review board of Boston Medical Center approved the examination content for both the original cohort and offspring examinations.

Ascertainment of intermittent claudication

At each original cohort and offspring cohort examination, intermittent claudication was assessed using a standardized physician-administered questionnaire (8) that asked about the presence of exertional calf discomfort related to rapidity of walking or uphill walking and whether symptoms were relieved with rest. A second physician independently interviewed all participants suspected to have claudication. A review panel comprising three Framingham Heart Study investigators examined all available evidence (including hospital records and personal physician office records) and made the final diagnostic determination of the presence of intermittent claudication. The diagnosis was adjudicated solely on the basis of medical history. Confirmation with ankle-brachial blood pressure testing was not available. If the exact date of onset of claudication was unavailable, the midpoint between the date of the last symptom-free Heart Study examination and the date of the Heart Study examination at which claudication was identified was assigned. Persons with intermittent claudication at study entry were excluded. Lower extremity bypass, lower extremity angioplasty, and lower extremity amputation for ischemia were not considered, since these data were not available at all Heart Study examinations and thus could not be compared across all decades.

Measurement of risk factors and identification of CVDs

Each Heart Study examination included a standardized medical history interview, a physical examination, laboratory testing, and electrocardiography. Blood pressure was measured twice by the examining physician with the participant in the seated position. Two blood pressure measurements were averaged to determine the presence of hypertension. Hypertension was defined as a systolic blood pressure of $\geq$140 mmHg, a diastolic blood pressure of $\geq$90 mmHg, or the use of antihypertensive medication. Height and weight were obtained using standard protocols, and body mass index was defined as weight in kilograms divided by height in meters squared. Participants were asked whether they had smoked cigarettes regularly during the year preceding each examination, and if the answer was yes, the number of cigarettes smoked per day was recorded. Diabetes was defined as a fasting blood glucose level of $\geq$126 mg/dl, a nonfasting blood glucose level of $\geq$200 mg/dl, or the use of insulin or oral hypoglycemic medication. Elevated blood cholesterol was defined as a total cholesterol level of $\geq$240 mg/dl or the use of cholesterol-lowering medication.

New CVD events were identified at each examination on the basis of a medical history interview, physical examination findings, electrocardiography, and review of outside medical records (hospital records, personal physicians' records, and death certificates). A review panel of three physician investigators (or a panel of study neurologists for cerebrovascular outcomes) examined all available evidence and made the final determination of events using established Framingham Heart Study criteria. CVD outcomes included angina, coronary insufficiency, myocardial infarction, congestive heart failure, intermittent claudication, stroke and transient ischemic attack, and death from cardiovascular causes.

Statistical analysis

Cases of intermittent claudication among participants aged 40 years or older were grouped by date of onset: 1950–1969 (191 cases), 1970–1979 (191 cases), 1980–1989 (148 cases), and 1990–1999 (138 cases). Because crude trends were similar in men and women and there were no significant age-sex interactions, we pooled men and women for all analyses. Age- and sex-adjusted incidence rates of intermittent claudication and 95 percent confidence intervals for each time period were calculated using the direct method. Log-linear Poisson regression (PROC GENMOD in SAS) was used to estimate sex- and age-adjusted rate ratios for intermittent claudication incidence during the 1970s, 1980s, and 1990s, with the 1950–1969 time period being used as the reference period (21). A p value less than 0.05 was considered statistically significant.

We conducted prespecified subgroup analyses to examine whether the time period effects for intermittent claudication incidence differed by 1) smoking status (ever smoking vs. never smoking) or 2) prevalent CVD at diagnosis of intermittent claudication (yes/no). We chose to examine smoking status given the importance of this risk factor in the risk of intermittent claudication. Furthermore, prior work demonstrated a decline in the prevalence and incidence of claudication in association with a decline in smoking among Icelandic men between 1968 and 1986 (22). Stratification by baseline CVD allowed us to explore possible differences in intermittent claudication incidence related to primary versus secondary prevention efforts over time (a decline in claudication incidence among persons free of CVD would suggest primary prevention effects, whereas declines in claudication incidence in persons with CVD would suggest secondary prevention effects). Follow-up was restricted to

TABLE 1. Age- and sex-adjusted data on cardiovascular disease risk factors at the time of diagnosis with intermittent claudication, by time period, Framingham Heart Study, 1950–1999

Risk factor	Time period				*p* for trend
	1950–1969 (*n* = 191)	1970–1979 (*n* = 191)	1980–1989 (*n* = 148)	1990–1999 (*n* = 138)	
Mean age (years)	61 (8)*	64 (9)	66 (10)	69 (10)	
Male sex (%)	62	53	55	57	
Cigarette smoking (%)					
Current smoker	46	35	37	37	<0.001
Former smoker	22	31	33	38	
Never smoker	32	34	30	25	
Pack-years of smoking (ever smokers)	24	29	37	47	<0.001
Diabetes (%)	2	10	16	25	<0.001
Body mass index† >30 (%)	10	14	11	26	0.002
High cholesterol level (≥240 mg/dl) (%)	47	40	34	33	0.004
Cholesterol treatment (%)	3	3	4	9	0.012
Hypertension (%)	41	59	53	62	0.004
Cardiovascular disease (%)	27	41	61	49	<0.001

* Numbers in parentheses, standard deviation.
† Weight (kg)/height $(m)^2$.

the 10-year period following the onset of claudication. Kaplan-Meier analysis was used to compare age group-adjusted (<60, 60–69, or ≥70 years at the time of intermittent claudication diagnosis) and sex-adjusted 10-year survival over the four time periods. The log-rank test was used to test for homogeneity across strata.

Risk factor profiles over the time periods of study for persons with incident intermittent claudication were compared using data from the last examination attended prior to the diagnosis. Adjustment for sex and 5-year age group was performed using the direct method. Time trends in risk factors in the full sample of men and women at risk of intermittent claudication were also examined; for each participant, risk factor values for all examinations attended within each time period were averaged. Linear regression, adjusting for 5-year age group and sex, was used to obtain least-squares mean values for each risk factor according to time period.

RESULTS

From 1950 through 1999, intermittent claudication occurred in 668 participants (43 percent women). The mean age at diagnosis of claudication was 61 years in the 1950–1969 time period; it rose to 69 years in the 1990–1999 time period. Age- and sex-adjusted risk factors for CVD at the time of diagnosis with intermittent claudication for each time period are shown in table 1. The majority of persons diagnosed with intermittent claudication were current or former cigarette smokers, and pack-years of smoking increased with each successive time period. Of note, from 1950 to 1969, 32 percent of cases with intermittent claudication had never smoked, but in the 1990–1999 time period, only 25 percent of intermittent claudication cases were never smokers. Among claudication cases, the prevalences of diabetes, obesity (body mass index >30), use of cholesterol-lowering medication, and hypertension increased across time periods. Age- and sex-adjusted prevalent CVD was present in approximately 40–60 percent of cases in the later time periods, as compared with only 27 percent of cases in the referent period.

The age- and sex-adjusted incidence rate of intermittent claudication in the 1950–1969 referent period was 282 per 100,000 person-years. In the 1970s, the rate rose to 345 per 100,000 person-years; it then fell in the 1980s and 1990s to 243 per 100,000 person-years and 225 per 100,000 person-years, respectively (figure 1). The results of age- and sex-adjusted log-linear Poisson regression analyses are shown in table 2. Overall, the incidence of intermittent claudication fell by 16 percent in the 1980s and by 18 percent in the 1990s relative to the 1950–1969 period (for trend across time periods, $p = 0.01$). Never smokers experienced a steeper decline in intermittent claudication incidence across the time periods than ever smokers (42 percent vs. 20 percent, respectively). In the subgroup of subjects without prevalent CVD, the incidence of intermittent claudication was 20 percent lower in the 1990–1999 time period than in 1950–1969 (p for trend = 0.035), whereas in subjects with prevalent CVD, no significant decline in the incidence of claudication was observed (p for trend = 0.31).

Temporal trends in risk factors from 1950 to 1999 among subjects at risk of developing intermittent claudication are shown in table 3. The age- and sex-adjusted prevalences of

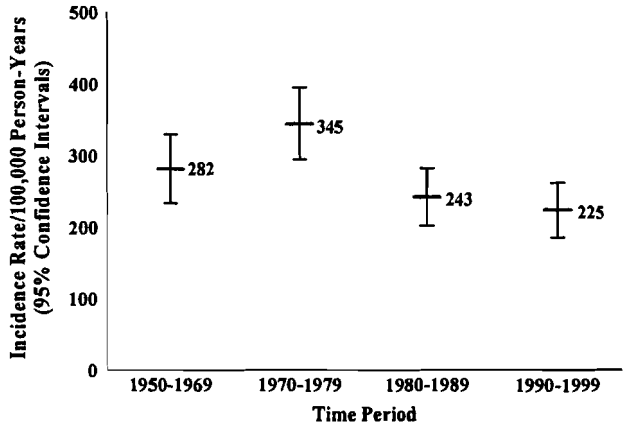


FIGURE 1. Temporal trends in the age- and sex-adjusted incidence rate (per 100,000 person-years) of intermittent claudication, Framingham Heart Study, 1950–1999. Bars, 95% confidence interval.

diabetes, obesity, and antihypertensive treatment increased, whereas the prevalences of current smoking and high blood cholesterol declined in the study population over the four time periods. Of note, the prevalence of diabetes increased dramatically among persons with prevalent CVD, from 7 percent in the 1950–1969 time period to 22 percent in the 1990–1999 time period. The prevalence of CVD in the full sample was 12 percent in the first three time periods and 11 percent in the 1990–1999 time period.

Survival following the onset of intermittent claudication remained unchanged over the course of the study (figure 2) ($p = 0.27$). By 10 years following the onset of intermittent claudication, nearly 40 percent of claudicants in all time periods had died. In subgroup analyses, no significant differences in mortality after claudication onset across time periods were observed in never or ever smokers or in claudicants with or without prevalent CVD at the time of intermittent claudication diagnosis.

DISCUSSION

The age- and sex-adjusted incidence of intermittent claudication in our population-based sample declined by 16–18 percent in the 1980s and 1990s in comparison with the period 1950–1969. The decline in intermittent claudication incidence was steeper among never smokers, with rates in the 1990s being over 40 percent lower than rates in 1950–1969. Smokers may have experienced less of a decline in intermittent claudication incidence than nonsmokers, because some of the smoking-related adverse effects on the arterial wall may be cumulative and irreversible (23). By the time smokers enter middle age, the thickness of the peripheral arterial wall is already greater than that of nonsmokers (24). In our sample, the incidence of intermittent claudication also declined significantly among persons without prevalent CVD, suggesting that primary prevention efforts aimed at modifying risk factors were successful. Among persons at risk of intermittent claudication who were free of CVD, rates of smoking and high cholesterol decreased whereas blood pressure treatment rates increased during the latter two decades of our study (table 3). In the Reykjavik Study, declines in intermittent claudication among men from 1968 to 1986 have been linked to declines in smoking and serum cholesterol (22).

Despite reports that persons with CVD were more likely to receive treatment for risk factors (13), we did not find a significant decline in incidence of claudication among those with prevalent CVD. The reasons for the lack of improvement in claudication incidence among persons with CVD over the calendar decades studied are unclear. Although rates of smoking and high cholesterol decreased and blood pressure treatment increased greatly, prevalences of diabetes and obesity among persons with prevalent CVD progressively increased across calendar time. It is possible that emerging medical and surgical therapies for CVD over time resulted in improved overall function among persons with CVD. The improvement in function may have allowed participants to walk a distance great enough to develop symptoms of claudication, whereas in the past persons with CVD may have been too functionally limited by symptoms

TABLE 2. Age- and sex-adjusted rate ratios for incidence of intermittent claudication, by decade, Framingham Heart Study, 1950–1999

	Time period							
	1950–1969 (referent)	1970–1979		1980–1989		1990–1999		*p* for trend
		RR*	95% CI*	RR	95% CI	RR	95% CI	
Overall	1.0	1.16	0.95, 1.43	0.84	0.67, 1.04	0.82	0.65, 1.03	0.01
Smoking status								
Ever smoker	1.0	1.20	0.94, 1.53	0.81	0.63, 1.05	0.80	0.61, 1.04	0.01
Never smoker	1.0	0.73	0.47, 1.13	0.60	0.37, 0.96	0.58	0.35, 0.94	0.02
CVD* status								
No CVD	1.0	1.43	1.10, 1.87	0.94	0.70, 1.26	0.80	0.59, 1.09	0.035
Prevalent CVD	1.0	0.83	0.60, 1.16	0.70	0.50, 0.99	0.88	0.60, 1.26	0.31

* RR, rate ratio; CI, confidence interval; CVD, cardiovascular disease.

TABLE 3. Age- and sex-adjusted prevalence (%) of cardiovascular disease risk factors in persons at risk of intermittent claudication, Framingham Heart Study, 1950–1999*

Risk factor	Time period			
	1950–1969	1970–1979	1980–1989	1990–1999
Full sample				
Smoking	42	30	23	16
Diabetes	5	8	9	11
Body mass index† >30	13	15	18	25
Blood pressure treatment	15	21	31	29
High cholesterol level (≥240 mg/dl)	48	32	28	17
Persons with prevalent CVD‡				
Smoking	38	30	27	21
Diabetes	7	14	16	22
Body mass index >30	15	18	20	31
Blood pressure treatment	17	26	55	54
High cholesterol level (≥240 mg/dl)	46	33	31	16
Persons free of CVD				
Smoking	42	31	23	16
Diabetes	4	7	8	10
Body mass index >30	13	15	18	25
Blood pressure treatment	14	20	28	27
High cholesterol level (≥240 mg/dl)	49	32	28	17

* p for trend < 0.001 for all risk factors in the full sample and in the samples stratified by cardiovascular disease status.
† Weight (kg)/height $(m)^2$.
‡ CVD, cardiovascular disease.

such as chest pain or shortness of breath to experience claudication symptoms. Furthermore, survival following the onset of coronary heart disease and cerebrovascular disease has improved, which may have permitted participants to develop claudication. The increasing prevalence of diabetes among persons with prevalent CVD, along with improvements in CVD treatment and survival over the time periods studied, may have resulted in the lack of significant improvement in incidence of intermittent claudication among persons with prevalent CVD in our study.

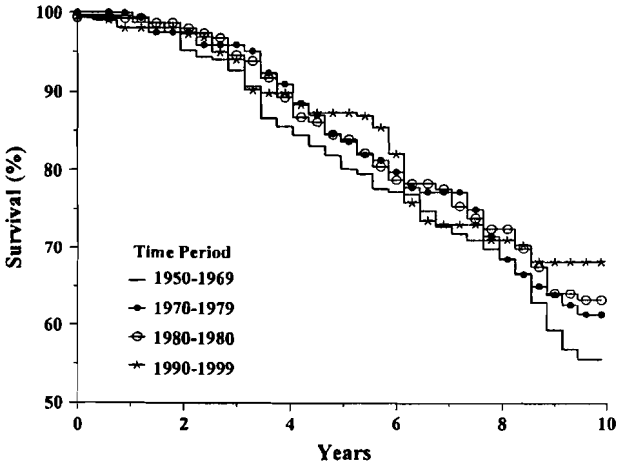


FIGURE 2. Ten-year survival following the onset of intermittent claudication, by time period, Framingham Heart Study, 1950–1999. Shown are the results of Kaplan-Meier survival analysis adjusted for sex and age group at diagnosis of intermittent claudication.

No temporal improvement in survival after the onset of claudication was observed. Approximately 40 percent of claudicants died within 10 years of onset in all time periods. The lack of improvement in survival following the diagnosis of intermittent claudication contrasts with the significant decline observed in rates of sudden and nonsudden cardiac death (25), as well as observed declines in mortality following the onset of other CVDs, including myocardial infarction, congestive heart failure, and stroke in the community (14, 15, 26, 27). Heart disease and stroke have been the focus of national prevention programs, with education efforts being directed at both physicians and the general public, whereas PAD has not yet emerged as a target for a national public health campaign.

Recent national data demonstrate that PAD is underdiagnosed in the primary care setting; with the exception of smoking, risk factors are less intensively treated in persons with PAD than in persons with other types of CVD (13). Survey data from primary care physicians, cardiologists, and vascular surgeons suggest that deficiencies in physician knowledge and attitude play an important role in the lower

intensity of risk factor modification efforts in PAD patients (28). It appears that PAD patients do not fully appreciate the increased risk of cardiovascular events associated with their disease and the benefits of risk factor reduction (29). Misperceptions on the part of PAD patients may contribute to the lower rates of risk factor control. Only a small number of randomized controlled trials of CVD risk reduction have been carried out in persons with PAD. Nevertheless, data support the beneficial effects of antiplatelet therapies and angiotensin-converting enzyme inhibitor therapies on cardiovascular outcomes in PAD patients (16, 30, 31). Moreover, in the Scandinavian Simvastatin Survival Study, lipid-lowering therapy with simvastatin in coronary heart disease patients resulted in a 38 percent reduction in new or worsening intermittent claudication (32), and a review of randomized trials of lipid-lowering therapies in PAD patients demonstrated a marked but nonsignificant reduction in mortality (odds ratio = 0.2, 95 percent confidence interval: 0.03, 1.17) associated with treatment (33). More recently, results of the Heart Protection Study demonstrated that simvastatin significantly reduced the risk of major vascular events by approximately 25 percent among persons with PAD (34). Furthermore, in a subgroup analysis of persons with vascular disease, including PAD, simvastatin substantially reduced the risk of stroke (35). Intensive blood pressure control in persons with both diabetes and PAD has been shown to be associated with a marked reduction in CVD events (36). These data suggest a beneficial role for modification of high blood cholesterol levels and high blood pressure in persons with PAD. Despite existing evidence demonstrating a clear benefit, antiplatelet therapies have been prescribed less frequently in PAD patients than in patients with other CVDs (13). Barriers to effective secondary prevention of CVD probably explain, in large part, the lack of improvement in survival among claudicants in recent decades, given that the majority of deaths among persons with the disease are due to coronary and cerebrovascular causes (2–4, 7).

Smoking is the single most important risk factor for PAD (7, 37, 38), and it is associated with disease progression and risk of limb amputation. Although, in our study, the decline in incidence of intermittent claudication was seen in both ever smokers and never smokers, the magnitude of the decline was greater in persons who had never smoked. The declining trend in intermittent claudication in the later decades is not surprising, given the concurrent decline in cigarette smoking in the population at risk. In the 1990s, the prevalence of current smoking in our study sample was 17 percent, less than half that observed in the 1950–1969 time period. Smoking cessation in the Quebec Cardiovascular Study was associated with a decline in risk of intermittent claudication such that men who stopped smoking for 1 year had a risk of claudication similar to that of nonsmokers (37). In the Reykjavik Study, a prospective study of Icelandic men, Ingolfsson et al. (22) also reported a marked decline in incidence and prevalence of claudication from 1968 to 1986 in association with a decline in smoking rates and a lowering of cholesterol levels in Iceland. However, other studies have demonstrated a persistent elevation in risk of PAD among former smokers (39). In addition to smoking, all other major CVD risk factors have been shown to be associated with risk of claudication (7). Thus, national efforts aimed at detection, treatment, and control of high blood pressure and high blood cholesterol have probably contributed to the observed decline in the incidence of intermittent claudication in our study. Of concern is recent national survey data demonstrating an increasing prevalence of multiple risk factors among adult men and women in the US population (40), which in turn may lead to a reversal of gains made in decreasing CVD incidence.

Our study had a number of potential limitations. First, the Framingham Study sample is not nationally representative and is primarily Caucasian; results may not be generalizable to other racial or ethnic groups. Studies of racially diverse samples have observed a higher prevalence of PAD among non-Hispanic Blacks than among Whites (1, 41, 42). Second, we studied intermittent claudication diagnosed only by classical medical history and did not have confirmatory testing. Thus, misclassification is possible, but diagnostic criteria remained the same over the decades studied. Any misclassification would therefore be expected to be random and result in a bias toward the null. Perhaps more importantly, most persons with PAD are asymptomatic (41, 43–45); some have atypical leg discomfort (46, 47), and others stop walking to prevent symptom onset. Our study was only able to observe incidence and mortality trends for the fraction of persons with PAD who develop classical symptoms. Finally, there have only been a few randomized controlled treatment trials in persons with PAD from which to establish the benefit of antiplatelet and risk factor reduction therapies (16, 31–36). Prior to these trials, physicians needed to extrapolate treatment benefits from observations in patients with CVD. We would not have been able to observe any recent improvements in mortality following the onset of claudication that may have occurred in response to the latest trial results.

In conclusion, the incidence of intermittent claudication has declined in the general population over the past 50 years, but mortality following the onset of claudication has not improved; approximately 40 percent of persons with intermittent claudication died within 10 years in all time periods studied. Improvements in primary and secondary prevention to modify CVD risk factors and increasing utilization of effective therapies in persons with PAD are needed.

ACKNOWLEDGMENTS

This work was supported by the National Heart, Lung, and Blood Institute (contract N01-HC-25195).

Conflict of interest: none declared.

REFERENCES

1. Selvin E, Erlinger TP. Prevalence of and risk factors for peripheral arterial disease in the United States: results from

the National Health and Nutrition Examination Survey, 1999–2000. Circulation 2004;110:738–43.
2. Smith GD, Shipley MJ, Rose G. Intermittent claudication, heart disease risk factors, and mortality: The Whitehall Study. Circulation 1990;82:1925–31.
3. Bainton D, Sweetnam P, Baker I, et al. Peripheral vascular disease: consequence for survival and association with risk factors in the Speedwell prospective heart disease study. Br Heart J 1994;72:128–32.
4. Bowlin SJ, Medalie JH, Flocke SA, et al. Intermittent claudication in 8343 men and 21-year specific mortality follow-up. Ann Epidemiol 1997;7:180–7.
5. Kannel WB. The demographics of claudication and the aging of the American population. Vasc Med 1996;1:60–4.
6. McDermott MM, Liu K, Greenland P, et al. Functional decline in peripheral arterial disease: associations with the ankle brachial index and leg symptoms. JAMA 2004;292:453–61.
7. Kannel WB, McGee DL. Update on some epidemiologic features of intermittent claudication: The Framingham Study. J Am Geriatr Soc 1985;33:13–18.
8. Murabito JM, D'Agostino RB, Silbershatz H, et al. Intermittent claudication: a risk profile from the Framingham Heart Study. Circulation 1997;96:44–9.
9. Bowlin SJ, Medalie JH, Flocke SA, et al. Epidemiology of intermittent claudication in middle-aged men. Am J Epidemiol 1994;140:418–30.
10. Chobanian AV, Bakris GL, Black HR, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 2003;289:2560–72.
11. 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;106:3143–421.
12. From the Centers of Disease Control and Prevention. State-specific trends in high blood cholesterol awareness among persons screened—United States, 1991–1999. JAMA 2001; 286:1706–7.
13. Hirsch AT, Criqui MH, Treat-Jacobson D, et al. Peripheral arterial disease detection, awareness, and treatment in primary care. JAMA 2001;286:1317–24.
14. Guidry UC, Evans JC, Larson MG, et al. Temporal trends in event rates after Q-wave myocardial infarction: The Framingham Heart Study. Circulation 1999;100:2054–9.
15. Levy D, Kenchaiah S, Larson MG, et al. Long-term trends in the incidence of and survival with heart failure. N Engl J Med 2002;347:1397–402.
16. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. BMJ 2002;324:71–86.
17. Feinleib M, Kannel WB, Garrison RJ, et al. The Framingham Offspring Study: design and preliminary data. Prev Med 1975;4:518–25.
18. Kannel WB, Feinleib M, McNamara PM, et al. An investigation of coronary heart disease in families: The Framingham Offspring Study. Am J Epidemiol 1979;110:281–90.
19. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: The Framingham Study. Am J Public Health 1951;41:279–81.
20. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: The Framingham Study. Ann N Y Acad Sci 1963;107:539–56.
21. SAS/STAT software: changes and enhancements through release 6.12 PROC GENMOD. Cary, NC: SAS Institute, Inc, 1997:247–38.
22. Ingolfsson IO, Sigurdsson G, Sigvaldason H, et al. A marked decline in the prevalence and incidence of intermittent claudication in Icelandic men 1968–1986: a strong relationship to smoking and serum cholesterol—The Reykjavik Study. J Clin Epidemiol 1994;47:1237–43.
23. Howard G, Wagenknecht LE, Burke GL, et al. Cigarette smoking and progression of atherosclerosis: The Atherosclerosis Risk in Communities (ARIC) Study. JAMA 1998;279: 119–24.
24. van den Berkmortel FW, Smilde TJ, Wollersheim H, et al. Intima-media thickness of peripheral arteries in asymptomatic cigarette smokers. Atherosclerosis 2000;150:397–401.
25. Fox CS, Evans JC, Larson MG, et al. Temporal trends in coronary heart disease mortality and sudden cardiac death from 1950 to 1999: The Framingham Heart Study. Circulation 2004;110:522–7.
26. McGovern PG, Shafar E, Sprafka JM, et al. The role of stroke attack rate and case fatality in the decline of stroke mortality: The Minnesota Heart Survey. Ann Epidemiol 1993; 3:483–7.
27. Wolf PA, D'Agostino RB, O'Neal MA, et al. Secular trends in stroke incidence and mortality: The Framingham Study. Stroke 1992;23:1551–5.
28. McDermott MM, Hahn EA, Greenland P, et al. Atherosclerotic risk factor reduction in peripheral arterial disease: results of a national physician survey. J Gen Intern Med 2002;17: 895–904.
29. McDermott MM, Mandapat AL, Moates A, et al. Knowledge and attitudes regarding cardiovascular disease risk and prevention in patients with coronary or peripheral arterial disease. Arch Intern Med 2003;163:2157–62.
30. Yusuf S, Sleight P, Pogue J, et al. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 2000;342: 145–53.
31. Ostergren J, Sleight P, Dagenais G, et al. Impact of ramipril in patients with evidence of clinical or subclinical peripheral arterial disease. Eur Heart J 2004;25:17–24.
32. Pedersen TR, Kjekshus J, Pyorala K, et al. Effect of simvastatin on ischemic signs and symptoms in the Scandinavian Simvastatin Survival Study (4S). Am J Cardiol 1998;81: 333–5.
33. Leng GC, Price JF, Jepson RG. Lipid-lowering for lower limb atherosclerosis. Cochrane Database Syst Rev 1998;(4): CD000123. (Electronic article). (http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD000123/frame.html).
34. 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;360:7–22.
35. Collins R, Armitage J, Parish S, et al. Effects of cholesterol-lowering with simvastatin on stroke and other major vascular events in 20536 people with cerebrovascular disease or other high-risk conditions. Lancet 2004;363:757–67.
36. Mehler PS, Coll JR, Estacio R, et al. Intensive blood pressure control reduces the risk of cardiovascular events in patients with peripheral arterial disease and type 2 diabetes. Circulation 2003;107:753–6.
37. Dagenais GR, Maurice S, Robitaille NM, et al. Intermittent claudication in Quebec men from 1974–1986: The Quebec Cardiovascular Study. Clin Invest Med 1991;14:93–100.
38. Fowkes FG, Housley E, Riemersma RA, et al. Smoking, lipids, glucose intolerance, and blood pressure as risk factors for peripheral atherosclerosis compared with ischemic heart

disease in the Edinburgh Artery Study. Am J Epidemiol 1992; 135:331–40.
39. Fowler B, Jamrozik K, Norman P, et al. Prevalence of peripheral arterial disease: persistence of excess risk in former smokers. Aust N Z J Public Health 2002;26:219–24.
40. Greenlund KJ, Zheng ZJ, Keenan NL, et al. Trends in self-reported multiple cardiovascular disease risk factors among adults in the United States, 1991–1999. Arch Intern Med 2004;164:181–8.
41. Newman AB, Sutton-Tyrrell K, Kuller LH. Lower-extremity arterial disease in older hypertensive adults. Arterioscler Thromb 1993;13:555–62.
42. Collins TC, Petersen NJ, Suarez-Almazor M, et al. The prevalence of peripheral arterial disease in a racially diverse population. Arch Intern Med 2003;163:1469–74.
43. Murabito JM, Evans JC, Nieto K, et al. Prevalence and clinical correlates of peripheral arterial disease in the Framingham Offspring Study. Am Heart J 2002;143:961–5.
44. Newman AB, Sutton-Tyrrell K, Rutan GH, et al. Lower extremity arterial disease in elderly subjects with systolic hypertension. J Clin Epidemiol 1991;44:15–20.
45. Vogt MT, Cauley JA, Kuller LH, et al. Prevalence and correlates of lower extremity arterial disease in elderly women. Am J Epidemiol 1993;137:559–68.
46. McDermott MM, Greenland P, Liu K, et al. Leg symptoms in peripheral arterial disease: associated clinical characteristics and functional impairment. JAMA 2001;286:1599–606.
47. Newman AB, Naydeck BL, Sutton-Tyrrell K, et al. The role of comorbidity in the assessment of intermittent claudication in older adults. J Clin Epidemiol 2001;54:294–300.

Sibling Cardiovascular Disease as a Risk Factor for Cardiovascular Disease in Middle-aged Adults

Joanne M. Murabito, MD, ScM
Michael J. Pencina, PhD
Byung-Ho Nam, PhD
Ralph B. D'Agostino, Sr, PhD
Thomas J. Wang, MD
Donald Lloyd-Jones, MD, ScM
Peter W. F. Wilson, MD
Christopher J. O'Donnell, MD, MPH

Context While parental cardiovascular disease (CVD) doubles the risk for CVD in offspring, the extent of increased risk associated with sibling CVD is unclear.

Objective To determine, using validated events, whether sibling CVD predicts outcome in middle-aged adults independent of other risk factors.

Design, Setting, and Participants The Framingham Offspring Study, an inception cohort of the Framingham Heart Study, a prospective population-based cohort study initiated in 1948 with the offspring cohort initiated in 1971. Participants (n = 2475) were members of the offspring cohort aged 30 years or older, free of CVD, and with at least 1 sibling in the study; all were followed up for 8 years.

Main Outcome Measures Association of sibling CVD with 8-year personal risk for CVD using pooled logistic regression. A secondary analysis restricted to offspring with both parents in the study assessed the joint impact of parental and sibling CVD occurrence.

Results Among 973 person-examinations in the sibling CVD group (mean age, 57 years) and 4506 person-examinations in the no sibling CVD group (mean age, 47 years), 329 CVD events occurred during follow-up. Baseline risk factors were more prevalent in the sibling CVD group compared with the no sibling CVD group. Sibling CVD was associated with a significantly increased risk for incident CVD (age- and sex-adjusted odds ratio [OR], 1.55; 95% confidence interval [CI], 1.19-2.03). Adjustment for risk factors did not substantially attenuate the risk (adjusted OR, 1.45; 95% CI, 1.10-1.91). In the analysis restricted to persons with both parents in the study, in models adjusting for both sibling and parental CVD, the multivariable-adjusted OR for sibling CVD (1.99; 95% CI, 1.32-3.00) exceeded that for parental CVD (1.45; 95% CI, 1.02-2.05).

Conclusion Using validated events, sibling CVD conferred increased risk of future CVD events above and beyond established risk factors and parental CVD in middle-aged adults.

JAMA. 2005;294:3117-3123 www.jama.com

CARDIOVASCULAR DISEASE (CVD) in a first-degree relative confers increased risk for CVD,[1] but whether familial CVD is truly an independent risk factor remains controversial. Parental CVD doubles the risk of CVD in adult offspring.[2] A seminal report by Marenberg et al[3] established increased risk for death from coronary heart disease in twins. Risk associated with CVD in siblings in multiplex families is less certain because published estimates are largely derived from case-control studies that generally lack sibling CVD event validation.[4-6] Furthermore, estimates regarding magnitude of risk associated with a history of sibling CVD vary greatly. Some studies have reported CVD risk similar to that conferred by a history of parental CVD[5,7]; others have described much greater CVD risk in relation to sibling history than in relation to parental history.[6]

Accurate information about familial CVD will have increasing importance in prevention and treatment of CVD in the post-genome era.[8] Recent national survey data document that adults believe family history information is important to their health, but few have systematically collected this information from relatives.[9] Recall bias, especially for premature CVD in the family, may re-

Author Affiliations: National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Mass (Drs Murabito, Pencina, Nam, D'Agostino, Wang, and O'Donnell); Section of General Internal Medicine, School of Medicine (Dr Murabito) and Statistics and Consulting Unit (Drs Pencina, Nam, and D'Agostino), Boston University, Boston, Mass; Cardiology Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston (Dr Wang); Department of Preventive Medicine and Bluhm Cardiovascular Institute, Division of Cardiology, Feinberg School of Medicine, Northwestern University, Chicago, Ill (Dr Lloyd-Jones); Departments of Endocrinology, Diabetes, and Medical Genetics, Medical University of South Carolina, Charleston (Dr Wilson); and National Heart, Lung, and Blood Institute, Bethesda, Md (Dr O'Donnell).
Corresponding Author: Joanne M. Murabito, MD, ScM, Framingham Heart Study, 73 Mt Wayte Ave, Suite 2, Framingham, MA 01702 (murabito@bu.edu).

duce the usefulness of reported family history information.[10] It is also possible that risk estimates for sibling history in case-control studies are exaggerated due to bias related to differences in family size and age at disease onset.[11] Risk estimates for sibling history may also be inflated due to confounding such as that caused by the sharing of higher risk factor levels by siblings of persons with CVD.[12,13] Associations between reported sibling history and subclinical CVD are attenuated when adjusted for CVD risk factors.[14,15]

We sought to determine whether the occurrence of a validated sibling CVD event independently and prospectively predicted CVD events in a cohort of middle-aged adults. We further sought to examine the impact of sibling CVD over and above that of parental CVD.

METHODS

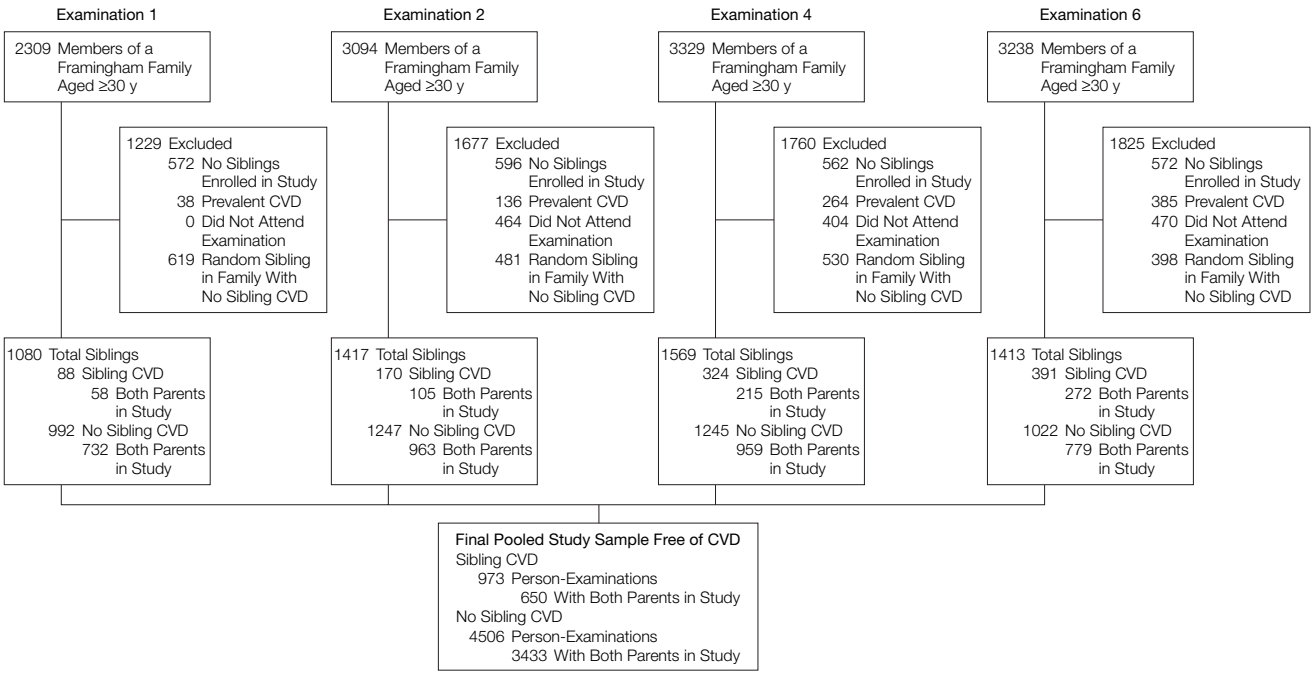
Study Sample and Definitions

In 1971, 5124 participants (offspring of the original Framingham Heart Study cohort and offspring spouses) aged 5 to 70 years were enrolled in the Framingham Offspring Study, a prospective epidemiologic cohort study.[16,17] The offspring cohort included 3498 participants who were members of an identified Framingham Study family. Participants have undergone follow-up examinations approximately every 4 years since study inception. Study design and entry criteria for both the offspring cohort and the original cohort of the Framingham Heart Study have been previously reported.[16,18] All participants provided written informed consent at each examination attended, and all study protocols were reviewed by the institutional review board at Boston Medical Center.

Data from 4 offspring examinations, each with 8 years of follow-up, were pooled: offspring cohort examinations 1 (1971 to 1975), 2 (1979 to 1982), 4 (1987 to 1990), and 6 (1995 to September 2, 1998). Follow-up for the final examination cohort ended in December 2004. Since the first and second examinations were about 8 years apart, to ascertain comparable lengths of follow-up after each of the 4 examinations, we chose to examine the 8-year occurrence of CVD events. All offspring participants who were members of a Framingham Study family and were aged 30 years and older at any of the 4 examinations were eligible for inclusion in our study if they had at least 1 sibling enrolled in the Framingham Offspring Study and if they were free of CVD at the time of examination (thus, the sibling contributing the positive CVD occurrence was excluded) (FIGURE). In addition, we randomly excluded 1 sibling from families with no sibling CVD to provide comparable structure between families with and without sibling CVD. Our final study sample included 2475 unique participants (1188 men) contributing 973 person-examinations from participants in a family with sibling CVD and 4506 person-examinations from participants in a family without sibling CVD (Figure).

Sibling CVD was defined as the occurrence of a validated sibling CVD event prior to an examination. Cardiovascular disease events included an-

Figure. Framingham Offspring Study Examination Cycles 1, 2, 4, and 6 (Each With 8 Years' Follow-up), Pooled



CVD indicates cardiovascular disease.

gina pectoris, coronary insufficiency, myocardial infarction, stroke or transient ischemic attack, intermittent claudication, coronary heart disease death, and CVD death. All CVD events (both sibling events and incident events occurring in participants in the study sample) were adjudicated by a panel of 3 senior investigators (or a panel of study neurologists for cerebrovascular disease events) who were unaware of sibling CVD status, using standardized criteria previously reported.[19]

Parental occurrence of premature CVD was available in a subsample of offspring participants with both parents enrolled in the original Framingham cohort (Figure). Parental events were adjudicated using the same protocol and standardized criteria. Parental premature CVD was defined as the occurrence of a validated parental event prior to an offspring examination and before age 55 years in fathers or age 65 years in mothers. These age cut points were derived from existing guidelines regarding family history of premature CVD.[20]

Risk factors were measured at each examination. Height and weight were obtained by trained technicians, and body mass index was calculated as weight in kilograms divided by the square of height in meters. Blood pressure was measured twice at rest by the examining physician, and the mean of the 2 blood pressure readings was used. Hypertension was defined as systolic blood pressure of 140 mm Hg or greater, diastolic blood pressure of 90 mm Hg or greater, or use of antihypertensive medication. Current smoking was defined as smoking 1 or more cigarettes per day in the year preceding examination. Blood was obtained in the fasting state, and the ratio of total cholesterol to high-density lipoprotein cholesterol was calculated. The presence of diabetes was defined by a fasting glucose level of 126 mg/dL (7.0 mmol/L) or greater or use of insulin or oral hypoglycemic agents.

Statistical Analysis

Pooled logistic regression analyses weighted for sibship size were used to examine the risks of incident CVD events associated with the occurrence of sibling CVD. The method of pooling person-examinations allows for time-dependent covariance of risk factors and sibling CVD events and has been shown to provide estimates of effect similar to those provided by time-dependent Cox analyses.[21] A weight for family size was used to diminish bias related to differences in the number of siblings across families. For all logistic regression analyses, the reference group consisted of participants with no sibling CVD prior to the examination. Odds ratios (ORs) and 8-year event rates with corresponding 95% confidence intervals (CIs) were calculated in unadjusted, age- and sex-adjusted, and multivariable-adjusted models. These analyses were repeated examining sibling premature CVD (defined as a CVD event in a brother before age 55 years or in a sister before age 65 years). Covariates in the multivariable model included age, sex, systolic blood pressure, use of antihypertensive medication, ratio of total cholesterol to high-density lipoprotein cholesterol, body mass index, diabetes mellitus, and current cigarette smoking. Significant interactions with sibling CVD were noted for age, systolic blood pressure, and hypertension treatment. Therefore, we repeated the analyses stratifying by age group above and below the median age ($\leq$48 years, $>$48 years) and hypertension status (yes or no).

In secondary analyses limited to participants with both parents in the study, logistic regression analyses were repeated with the multivariable-adjusted model including parental premature CVD in addition to the other covariates. Finally, to examine whether there was a dose effect for occurrence of CVD in any first-degree relative, the multivariable analysis was repeated, this time entering predictor variables for parental premature CVD only, sibling CVD only, and both parental premature CVD and sibling CVD.

To assess the added usefulness of the sibling CVD status in predicting future CVD events for each participant, we calculated a risk score according to the risk guidelines from the National Cholesterol Education Program Adult Treatment Panel III.[20] Participants were then stratified into 10-year risk categories for coronary heart disease (low = less than 10% risk; intermediate = 10% to 19% risk; and high = 20% risk or greater). These guidelines also place persons with known diabetes into the high-risk group. We used sex- and age-adjusted logistic regression analyses to compare 8-year CVD event rates between individuals with and without sibling CVD within each risk strata defined above, as well as across strata of individual CVD risk factors. All statistical analyses were performed using SAS version 8.0 (SAS Institute Inc, Cary, NC). *P* values were 2-sided; $P<.05$ was used to determine statistical significance.

RESULTS

Most sibling events occurred prematurely, with a mean age at onset of 48.2 (SD, 4.79), 48.7 (SD, 7.36), 50.7 (SD, 8.27), and 53.6 (SD, 9.94) years at the 4 pooled examinations, respectively. Compared with the group with no sibling CVD, those with sibling CVD were older and had higher prevalence of all traditional risk factors except current smoking (TABLE 1). Of the 329 incident CVD events during follow-up, there were 11 coronary deaths, 8 other CVD deaths, 99 nonfatal cases of myocardial infarction or coronary insufficiency, 106 cases of angina pectoris, 59 strokes or transient ischemic attacks, and 46 cases of intermittent claudication. There were 223 events in the group with no sibling CVD over 34 110 person-years of follow-up, yielding a crude rate of 6.54 events per 1000 person-years; there were 106 events in the sibling CVD group over 6943 person-years of follow-up, yielding a crude rate of 15.27 events per 1000 person-years.

The 8-year CVD event rates and ORs for CVD events associated with the occurrence of sibling CVD are shown in TABLE 2. Sibling CVD was associated with a significantly increased risk for

Table 1. Baseline Characteristics by Occurrence of Sibling Cardiovascular Disease (CVD)—The Framingham Offspring Study*

Characteristic	Full Sample (5479 Person-Examinations)	Sibling CVD (973 Person-Examinations)	No Sibling CVD (4506 Person-Examinations)
Demographic Data			
Age, mean (SD), y	48.57 (10.63)	56.65 (10.18)	46.83 (9.89)
Men, No. (%)	2558 (46.7)	445 (45.7)	2113 (46.9)
Clinical Data			
Blood pressure, mean (SD), mm Hg			
Systolic	124.6 (17.67)	132.1 (18.38)	123.0 (17.09)
Diastolic	78.43 (10.19)	79.39 (9.95)	78.23 (10.23)
Lipids, mean (SD), mg/dL			
Total cholesterol	204.5 (39.29)	212.3 (41.27)	202.9 (38.65)
HDL-C	50.08 (14.91)	48.80 (15.08)	50.36 (14.86)
Total cholesterol–HDL-C ratio	4.45 (1.73)	4.73 (1.71)	4.39 (1.73)
Body mass index, mean (SD)†	26.69 (4.85)	27.63 (4.97)	26.49 (4.80)
Medical History			
Hypertension, No. (%)‡	725 (13.3)	245 (25.2)	480 (10.7)
Current cigarette smoker, No. (%)‡	1631 (29.8)	254 (26.1)	1377 (30.6)
Diabetes, No. (%)‡	295 (5.4)	93 (9.6)	202 (4.5)
Parental premature CVD, No. (%)§	1184 (21.6)	238 (24.5)	946 (21.0)

Abbreviation: HDL-C, high-density lipoprotein cholesterol.
SI conversion factor: To convert total cholesterol and HDL-C to mmol/L, multiply mg/dL values by 0.0259.
*Data are based on 5479 person-examinations from 1188 men and 1287 women.
†Calculated as weight in kilograms divided by the square of height in meters.
‡See "Methods" section for definition.
§Restricted to offspring with both parents in the study (650 person-examinations in the sibling CVD group, 3433 person-examinations in the no sibling CVD group). Parental premature CVD defined as an event before age 55 years in fathers and before age 65 years in mothers.

Table 2. Eight-Year Risk for Cardiovascular Disease (CVD) in Middle-aged Adults, by Occurrence of Sibling Cardiovascular Disease—The Framingham Offspring Study

Model	8-Year CVD Risk, % (95% CI)		Risk for CVD, OR (95% CI)
	Sibling CVD	No Sibling CVD	
Full sample*			
Unadjusted	10.91 (9.04-13.10)	5.00 (4.40-5.67)	2.33 (1.82-2.98)
Age- and sex-adjusted	8.06 (6.29-10.29)	5.40 (4.45-6.55)	1.55 (1.19-2.03)
Multivariable-adjusted†	7.62 (5.40-10.57)	5.53 (4.03-7.48)	1.45 (1.10-1.91)
Sample restricted to participants with both parents in the FHS‡			
Unadjusted	10.75 (8.54-13.46)	4.67 (4.01-5.42)	2.46 (1.82-3.32)
Age- and sex-adjusted	7.88 (5.82-10.60)	5.03 (3.99-6.33)	1.64 (1.18-2.26)
Multivariable-adjusted§	7.51 (4.86-11.29)	5.11 (3.41-7.49)	1.56 (1.11-2.18)

Abbreviations: CI, confidence interval; FHS, Framingham Heart Study; OR, odds ratio.
*n = 5479 person-examinations.
†Adjusted for age, sex, systolic blood pressure, hypertension treatment, total cholesterol–high-density lipoprotein cholesterol ratio, body mass index, diabetes, and current smoking. Complete covariate data available for 5338 person-examinations.
‡n = 4083 person-examinations.
§Adjusted for factors listed above and for parental occurrence of premature CVD. Complete covariate data available for 3970 person-examinations.

incident CVD (age- and sex-adjusted OR, 1.55; 95% CI, 1.19-2.03); this association persisted even after adjustment for risk factors (multivariable-adjusted OR, 1.45; 95% CI, 1.10-1.91). The attributable risk percentage for sibling CVD was 27.4%; this represents the proportion of the 8-year CVD risk among those in the sibling CVD group that theoretically could be prevented if members of the group had not had a sibling with CVD. Because there were significant interactions in the multivariable model, we examined the model stratified by age (≤48 years vs >48 years) and hypertension status (yes or no). The impact of sibling CVD was stronger in the younger age group (multivariable-adjusted OR for sibling CVD vs no sibling CVD in the younger group: 2.22; 95% CI, 1.22-4.02; in the older group: 1.33; 95% CI, 0.98-1.80) and in participants free of hypertension (multivariable-adjusted OR in those without hypertension: 1.98; 95% CI, 1.31-2.99; in those with hypertension: 1.19; 95% CI, 0.82-1.72), but the differences were not statistically significant. In a secondary analysis taking into account the age at onset of sibling CVD, only premature onset of sibling CVD was significantly related to CVD incidence; multivariable ORs were 1.58 (95% CI, 1.18-2.12) for sibling premature CVD and 1.04 (95% CI, 0.61-1.77) for sibling nonpremature CVD. However, of the 973 person-examinations in the sibling CVD group, only 180 person-examinations came from families with nonpremature onset of sibling CVD.

We examined the impact of parental CVD in the secondary analysis restricted to participants with both parents in the study. The occurrence of sibling CVD remained a significant predictor of incident CVD, even after adjusting for all risk factors and for parental premature CVD (OR, 1.56; 95%

CI, 1.11-2.18) (Table 2). Sibling CVD was associated with a stronger risk for incident CVD than was parental premature CVD; multivariable-adjusted ORs were 1.45 (95% CI, 1.02-2.05) for parental premature CVD alone, 1.99 (95% CI, 1.32-3.00) for sibling CVD alone, and 1.53 (95% CI, 0.93-2.51) for both parental premature CVD and sibling CVD. Of note, the lower multivariable-adjusted risks among participants with both parental and sibling CVD might be explained by the particularly high prevalence of risk factors in this group (prevalence of diabetes and treatment for hypertension was 12.2% and 31.5%, respectively, in those with both parental and sibling CVD, compared with 5.8% and 22.3%, respectively, in those with sibling CVD only and 5.8% and 14.1%, respectively, in those with parental premature CVD only).

When we stratified participants by elevated levels of individual risk factors and estimates of overall CVD risk categories, sibling CVD information added substantially to discrimination of observed 8-year event rates (TABLE 3). Of note, when we stratified by age, sibling CVD was associated with a significantly higher event rate in those aged 30 to 59 years but not in those aged 60 years and older. Sibling CVD was associated with increased risk in persons with adverse levels of most risk factors and in persons in the intermediate and high Adult Treatment Panel III risk categories. However, sibling CVD was not associated with significantly increased risk among persons with known diabetes or hypertension who already had substantially higher event rates.

COMMENT

Using a prospective design and validated sibling CVD events, we found that sibling CVD was associated with a significantly increased risk for incident CVD events in middle-aged adults. The OR remained statistically significantly elevated in age- and sex-adjusted and multivariable-adjusted models, suggesting that age and traditional risk factors explain part but not all of the increased risk associated with sibling CVD. Furthermore, in analyses restricted to participants with both parents in the Framingham Study, the presence of sibling CVD conferred an increased CVD risk independent of parental premature CVD, and sibling CVD may be more strongly associated with incident events than is parental premature CVD. Our findings provide strong evidence that sibling CVD is an important risk factor for incident CVD and represents a useful marker of familial vulnerability to CVD events.

Results from a previously reported, population-based, case-control study conducted to estimate risk of coronary heart disease associated with various definitions of a family history clearly established the value of going beyond the simple yes or no response to questions about presence of disease in a first-degree relative.[4] This work demonstrated that distinguishing between an affected parent and an affected sibling was important, particularly for younger ages of disease onset. Moreover, having a sibling with coronary heart disease was associated with significantly increased risk, even in families with a parent already affected at a young age. The importance of a sibling history is consistent with twin studies in which coronary heart disease

Table 3. Eight-Year Cardiovascular Disease (CVD) Event Rates per 1000 Persons, by Presence or Absence of Sibling CVD, Stratified by Individual Risk Factor Levels—The Framingham Offspring Study

Risk Factor	Person-Examinations, No.	8-Year Event Rate/ 1000 Persons (95% CI)		*P* Value*
		Sibling CVD	No Sibling CVD	
Overall	5479	80 (63-102)	54 (44-65)	.001
Age, y				
30-59	4561	78 (57-106)	44 (35-55)	<.001
≥60	918	124 (86-176)	107 (75-151)	.44
Lipids, mg/dL				
Total cholesterol				
<200	2529	40 (25-66)	35 (24-50)	.51
200-239	1815	98 (65-144)	58 (42-79)	.009
≥240	1044	150 (102-217)	95 (68-132)	.02
HDL-C				
≥40	4026	56 (40-78)	42 (33-55)	.10
<40	1349	148 (105-207)	89 (66-120)	.004
Hypertension†				
Yes	1595	126 (92-171)	110 (85-141)	.37
No	3876	62 (43-90)	32 (23-43)	<.001
Smoking status†				
Nonsmoker	3839	59 (43-80)	47 (37-61)	.16
Current smoker	1631	134 (92-193)	70 (52-95)	<.001
Diabetes†				
Yes	295	231 (136-368)	139 (82-227)	.06
No	5184	70 (53-92)	49 (40-61)	.009
Body mass index‡				
<25	2215	61 (38-97)	33 (23-49)	.01
25-29.9	2155	79 (54-114)	66 (50-88)	.35
≥30	1084	124 (80-187)	72 (50-104)	.01
ATP III risk group				
<10%	4356	44 (30-65)	36 (27-47)	.27
10%-19%	652	170 (110-254)	106 (72-154)	.03
≥20% or diabetes	471	237 (157-342)	155 (105-223)	.04

Abbreviations: ATP III, Adult Treatment Panel III; CI, confidence interval; HDL-C, high-density lipoprotein cholesterol.
SI conversion factor: To convert total cholesterol and HDL-C to mmol/L, multiply mg/dL values by 0.0259.
*Represent the age- and sex-adjusted difference in 8-year event rates per 1000 persons between the sibling CVD group and the group with no sibling CVD.
†See "Methods" section for definition.
‡Calculated as weight in kilograms divided by the square of height in meters.

death in a young twin was especially associated with increased risk of death in the other twin.[3,22] Our study adds to these important previous studies, because we examined nontwin siblings and were able to confirm that sibling CVD increased risk for incident events principally if onset of sibling CVD was premature and if the sibling at risk was a young adult.

Some case-control studies have reported that the risk conferred by a history of sibling CVD is similar to that conferred by a history of parental CVD,[5,23] but others have found that sibling CVD confers a greater risk than parental CVD.[6] A recent study that examined the association between family history and subclinical measures of coronary atherosclerosis found a much stronger association between the presence and extent of coronary artery calcification on electron beam tomography and sibling history compared with parental history. In the analysis considering both together, the ORs for coronary calcification associated with sibling history and parental history were 2.3 and 1.3, respectively.[14] When we conducted an analysis comparing parental premature CVD with sibling CVD, we also found stronger associations for sibling than for parental CVD. The magnitude of risk for incident CVD associated with sibling CVD and parental premature CVD in our study was strikingly similar (ORs, 1.99 and 1.45, respectively) to that in the prior report. In the prior report, recall bias cannot be excluded as an explanatory factor contributing to the stronger ORs for sibling history compared with parental history; however, our data are not susceptible to this limitation, because sibling and parental events were based on medical records rather than on self-report.

Concern has been raised that risk associated with familial CVD history can be largely explained by familial aggregation of traditional risk factors. Consistent with other investigations of sibling history,[12,13,24-27] we found that participants with sibling CVD had higher prevalence of risk factors compared with participants with no sibling CVD. This finding was especially striking in participants with both sibling CVD and parental premature CVD, who had high prevalence of diabetes and use of medication for hypertension. Nevertheless, the risk associated with sibling CVD remained significant in multivariable models. This finding suggests that a significant proportion of risk is explained by factors other than traditional risk factors, in turn suggesting that other genetic risk factors may influence susceptibility to CVD. Among the possible factors for increased risks conferred by sibling CVD are shared early environmental exposures (in utero or early childhood) and a shared genetic background.

Strengths and Limitations

Several strengths and limitations of our study merit comment. Our prospective study design allowed the examination of incident CVD events, and all events within the family were adjudicated by a panel of senior investigators using the same standardized criteria. Thus, sibling events were validated by review of medical records rather than reliance on self-reports. Risk factors were directly measured for all participants and updated over time, independent of sibling CVD. Thus, measures of lipid levels and diabetes were ascertained directly and not obtained by self-report, which is more susceptible to misclassification. The use of reported risk factors in multivariable analyses may not fully account for shared risk factors within families and in turn results in an overestimation of the OR associated with sibling CVD. The original and offspring cohorts of the Framingham Heart Study are primarily white, potentially limiting the extent to which our findings can be generalized to other groups. Further, we did not have risk factor or event information on siblings who declined enrollment in the Framingham Study. It is possible that siblings with early-onset CVD died prior to enrollment or declined enrollment due to poor health status; this would be expected to bias our results toward an underestimation of the risk associated with sibling CVD.

Clinical and Research Implications

Using validated events and a prospective design, our study substantially extends the knowledge base regarding the importance of sibling CVD. We observed that sibling CVD confers increased risks of CVD events above and beyond traditional risk factors and parental premature CVD. Thus, sibling CVD should be considered as important as parental premature CVD in the assessment of risk. Further investigation is needed to better understand why sibling history may be a stronger predictor for CVD than parental history, including exploration of the contribution of an early shared environment to increased sibling risk. Moreover, investigation of whether to incorporate sibling CVD as well as parental CVD into existing risk prediction and prevention algorithms is warranted.

Author Contributions: Drs Murabito, Nam, Pencina, and O'Donnell had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Murabito, Pencina, Nam, D'Agostino, Wang, Lloyd-Jones, O'Donnell.
Acquisition of data: Murabito, D'Agostino, Wilson, O'Donnell.
Analysis and interpretation of data: Murabito, Pencina, Nam, D'Agostino, Wang, Lloyd-Jones, Wilson, O'Donnell.
Drafting of the manuscript: Murabito, O'Donnell.
Critical revision of the manuscript for important intellectual content: Murabito, Pencina, Nam, D'Agostino, Wang, Lloyd-Jones, Wilson, O'Donnell.
Statistical analysis: Pencina, Nam, D'Agostino, Lloyd-Jones, O'Donnell.
Obtained funding: D'Agostino.
Administrative, technical, or material support: Murabito, Wilson, O'Donnell.
Study supervision: Murabito, O'Donnell.
Financial Disclosures: None reported.
Funding/Support: This work was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (contract N01-HC-25195).
Role of the Sponsor: The National Heart, Lung, and Blood Institute approved the manuscript for publication but had no direct involvement in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation of the manuscript.

REFERENCES

1. Williams RR, Hunt SC, Heiss G, et al. Usefulness of cardiovascular family history data for population-based preventive medicine and medical research (the Health Family Tree Study and the NHLBI Family Heart Study). *Am J Cardiol*. 2001;87:129-135.
2. Lloyd-Jones DM, Nam BH, D'Agostino RB Sr, et al.

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.

3. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med*. 1994;330:1041-1046.

4. Silberberg JS, Wlodarczyk J, Fryer J, Robertson R, Hensley MJ. Risk associated with various definitions of family history of coronary heart disease: the Newcastle Family History Study II. *Am J Epidemiol*. 1998;147:1133-1139.

5. Ciruzzi M, Schargrodsky H, Rozlosnik J, et al; Argentine FRICAS (Factores de Riesgo Coronario en America del Sur) Investigators. Frequency of family history of acute myocardial infarction in patients with acute myocardial infarction. *Am J Cardiol*. 1997;80:122-127.

6. Roncaglioni MC, Santoro L, D'Avanzo B, et al; GISSI-EFRIM Investigators. Role of family history in patients with myocardial infarction: an Italian case-control study. *Circulation*. 1992;85:2065-2072.

7. Bertuzzi M, Negri E, Tavani A, La Vecchia C. Family history of ischemic heart disease and risk of acute myocardial infarction. *Prev Med*. 2003;37:183-187.

8. Guttmacher AE, Collins FS, Carmona RH. The family history—more important than ever. *N Engl J Med*. 2004;351:2333-2336.

9. Centers for Disease Control and Prevention (CDC). Awareness of family health history as a risk factor for disease—United States, 2004. *MMWR Morb Mortal Wkly Rep*. 2004;53:1044-1047.

10. Murabito JM, Nam BH, D'Agostino RB Sr, Lloyd-Jones DM, O'Donnell CJ, Wilson PW. Accuracy of offspring reports of parental cardiovascular disease history: the Framingham Offspring Study. *Ann Intern Med*. 2004;140:434-440.

11. Khoury MJ, Flanders WD. Bias in using family history as a risk factor in case-control studies of disease. *Epidemiology*. 1995;6:511-519.

12. Becker DM, Becker LC, Pearson TA, Fintel DJ, Levine DM, Kwiterovich PO. Risk factors in siblings of people with premature coronary heart disease. *J Am Coll Cardiol*. 1988;12:1273-1280.

13. Brenn T, Njolstad I. Coronary heart disease risk factors in subjects whose brothers, sisters or husbands developed premature myocardial infarction during 12 years of follow-up: the Finnmark Study (1977-1989). *J Cardiovasc Risk*. 1998;5:325-330.

14. Nasir K, Michos ED, Rumberger JA, et al. Coronary artery calcification and family history of premature coronary heart disease: sibling history is more strongly associated than parental history. *Circulation*. 2004;110:2150-2156.

15. Pankow JS, Heiss G, Evans GW, et al. Familial aggregation and genome-wide linkage analysis of carotid artery plaque: the NHLBI family heart study. *Hum Hered*. 2004;57:80-89.

16. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol*. 1979;110:281-290.

17. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study: design and preliminary data. *Prev Med*. 1975;4:518-525.

18. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci*. 1963;107:539-556.

19. Abbott RD, McGee DL. *The Framingham Study: An Epidemiologic Investigation of Cardiovascular Disease, Section 37: The Probability of Developing Certain Cardiovascular Diseases in Eight Years at Specified Values of Some Characteristics*. Bethseda, Md: National Heart, Lung, and Blood Institute; 1987.

20. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). 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;106:3143-3421.

21. D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Stat Med*. 1990;9:1501-1515.

22. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, de Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med*. 2002;252:247-254.

23. Friedlander Y, Siscovick DS, Weinmann S, et al. Family history as a risk factor for primary cardiac arrest. *Circulation*. 1998;97:155-160.

24. Allen JK, Young DR, Blumenthal RS, et al. Prevalence of hypercholesterolemia among siblings of persons with premature coronary heart disease: application of the Second Adult Treatment Panel guidelines. *Arch Intern Med*. 1996;156:1654-1660.

25. Rissanen AM. Familial occurrence of coronary heart disease: effect of age at diagnosis. *Am J Cardiol*. 1979;44:60-66.

26. Barrett-Connor E, Khaw K. Family history of heart attack as an independent predictor of death due to cardiovascular disease. *Circulation*. 1984;69:1065-1069.

27. Becker DM, Yook RM, Moy TF, Blumenthal RS, Becker LC. Markedly high prevalence of coronary risk factors in apparently healthy African-American and white siblings of persons with premature coronary heart disease. *Am J Cardiol*. 1998;82:1046-1051.

ELSEVIER

UROLOGIC COMPLICATIONS OF SEXUAL TRAUMA AMONG MALE SURVIVORS OF TORTURE

MARIE NORREDAM, SONDRA CROSBY, RICARDO MUNARRIZ, LIN PIWOWARCZYK, and MICHAEL GRODIN

ABSTRACT

Objectives. To describe the urologic and sexual complications of male survivors of sexual torture, including prevalence, sequelae, diagnosis, and treatment.

Methods. Through chart reviews, we identified all male survivors of torture who had been treated for physical and/or psychological symptoms due to sexual trauma at the Boston Center for Refugee Health and Human Rights at Boston Medical Center between January 1, 2001 and January 1, 2002. Of the 72 men seen, 20 (28%) were survivors of sexual trauma. Our study focused on genital trauma leading to urologic and/or sexual dysfunction. Therefore, all cases of male genital trauma that had been referred to the urology department (3 of 20) were selected for this review.

Results. The patients presented with chronic genital and erectile pain, lower urinary tract symptoms, and sexual dysfunction. The diagnostic workup included history, physical examination, and ultrasonography. Treatment included steroid injections for chronic pain and oral erectogenic agents for sexual dysfunction.

Conclusions. The apparent prevalence and severity of the physical and mental sequelae to sexual trauma make it an important area for screening when treating survivors of torture. Our study is the first of its kind to document urologic complications of sexual torture in a foreign-born U.S. cohort of tortured men, including prevalence, diagnosis, and treatment. The proposed use of steroid injections in the clinical treatment of these patients has not been previously reported. UROLOGY **65:** 28–32, 2005. © 2005 Elsevier Inc.

Thousands of asylum seekers and refugees enter Western countries every year. Between 1991 and 2000, about 1 million people applied for asylum in the United States.[1] Among those seeking asylum are torture survivors. In 1999, 400,000 survivors of torture were estimated to reside in the United States.[2] Although not necessarily self-proclaimed, survivors of torture and refugee trauma are consequently bound to appear in the offices of health professionals. A U.S. study of the prevalence of torture survivors among a random sample of foreign-born patients in primary care in a metropolitan area showed that 25% had suffered from torture.[3] It is of concern that health professionals often fail to ask about torture or are not trained to recognize the physical and psychological symptoms.[2] Additionally, survivors rarely volunteer their history of persecution because of shame or out of fear because the clinical setting is reminiscent of prior experiences of torture.[4] Men are especially ashamed to discuss sexual trauma because of fear of stigmatization and the shame of a perceived loss of masculinity.[5]

The United Nations Convention against Torture and Other Cruel, Inhuman and Degrading Treatment or Punishment defines torture as "an act by which severe pain or suffering, whether physical or mental, is intentionally inflicted on a person for such purposes as obtaining from him or a third person information or confession. . . ."[6] Male sexual trauma, which is a form of torture, can be characterized according to the methods used[7,8]:

1. Direct genital trauma: hitting, kicking, or applying electric shocks to genitals and/or anus, object inserted into urethral meatus and/or anus, cigarette burns to penis

From the Department of Health Law, Bioethics, and Human Rights, Boston University School of Public Health; Boston Center for Refugee Health and Human Rights, Section of General Internal Medicine, and Department of Urology, Boston Medical Center, Boston, Massachusetts; and Department of Health Services Research, Institute of Public Health, University of Copenhagen, Copenhagen, Denmark

Reprint requests: Marie Norredam, M.D., Department of Health Services Research, Institute of Public Health, University of Copenhagen, Blegdamsvej 3B, Copenhagen North 2200, Denmark. E-mail: m.norredam@pubhealth.ku.dk

Submitted: May 20, 2004, accepted (with revisions): August 11, 2004

0090-4295/05/$30.00
doi:10.1016/j.urology.2004.08.006

TABLE I. ***Most common methods of sexual torture and physical sequelae among 20 survivors of torture treated at Boston Center for Refugee Health and Human Rights at Boston Medical Center between January 1, 2001 and January 1, 2002***

Physical Trauma	Acute Physical Sequelae	Chronic Physical Sequelae
Forced nakedness	Groin/genital pain	Groin/genital pain
Anal rape	Fecal incontinence	Erectile dysfunction
Genital beatings with fists, sticks, or other instruments	Rectal bleeding	Urethral stricture
Electric shocks to genitals	Rectal pain	Fecal incontinence
Forced fellatio		
Insertion of toothpicks into the penis		

2. Nonconsensual sexual acts: pawing, anal rape, forced masturbation, and forced fellatio
3. Mental assaults: forced nakedness, sexual humiliation, and threats

On the basis of our work with survivors of torture and refugee trauma, we report 3 cases of genital trauma and their physical sequelae, focusing on the urologic complications. Our aim was to sensitize medical professionals to the male genital trauma that occurs in the context of sexual torture by discussing the methods of torture and the physical sequelae, diagnosis, and treatment strategies.

MATERIAL AND METHODS

Through chart reviews, we identified all male torture survivors who had been treated for physical or psychological symptoms due to sexual trauma at the Boston Center for Refugee Health and Human Rights at the Boston Medical Center. Between January 1, 2001 and January 1, 2002, 20 (28%) of the 72 men seen were identified as survivors of sexual trauma. The patients were all between 31 and 50 years of age and refugees from African countries in conflict. They had all been imprisoned or captured and had suffered torture, including severe genital beatings. Table I lists the most common methods of sexual torture, as well as the acute and chronic physical sequelae, among the 20 survivors of torture. Most patients had experienced multiple sexual assaults of the same or different kinds. We focus on the genital trauma leading to urologic and/or sexual dysfunction. Therefore, all cases of male genital trauma that had been referred to the urology department (3 of 20) were selected for this review. The cases are presented and serve to illustrate the different aspects of the symptoms, diagnosis, and treatment of male genital trauma. All patients have been kept anonymous. The Institutional Review Board at Boston Medical Center approved the study.

RESULTS

Case 1

Patient 1 was a 50-year-old man who had fled from a country in Central Africa after being persecuted by the military for rebel activity and imprisoned on 7 occasions between 1996 and 2000. During imprisonment, he was repeatedly beaten with fists and sticks all over his body, including the genital area. On one occasion, he was suspended naked from the ceiling by his arms with his legs bent, his feet secured to the floor wide apart, and his genitalia tied to the floor for 12 hours. He was unable to stand upright without painfully stretching his genitalia. He presented with complaints of erectile dysfunction and chronic incapacitating genital pain. On physical examination, he had diminished tunica elasticity and compliance, consistent with Peyronie's disease. Penile stretching elicited exquisite proximal dorsal penile pain radiating to the pubic bone and suspensory ligament. After he provided informed consent, two subcutaneous steroid injections (triamcinolone 50 mg) on the dorsal aspect of the penis, fundiform ligament, and pubic tubercles were performed within a 4-month period. After the first steroid injection, the patient experienced immediate and significant improvement of his incapacitating genital pain. After the second injection, the pain resolved completely. At last follow-up, his erectile dysfunction was being managed with oral erectogenic agents (sildenafil) with excellent results.

Case 2

Patient 2 was a 31-year-old man who had also been repetitively beaten in the genitals by fists and sticks while imprisoned in an East African country for political activities. He also presented with complaints of erectile dysfunction and severe and chronic genital pain. He had no orgasmic or ejaculatory problems, but his erections were reduced in rigidity compared with previously. The physical examination revealed exquisite pubic tubercle, cord, and suspensory ligament tenderness. Penile duplex Doppler ultrasonography after intracavernosal injection of vasoactive agents revealed decreased cavernosal systolic velocities and normal end-diastolic velocities. These findings were consistent with pure (no corporeal occlusive dysfunction) cavernosal artery insufficiency, most likely secondary to blunt perineal trauma. No penile plaques or tunica thickening were observed. After he provided informed consent, two steroid injections (triamcinolone 50 mg) were given into the cord, fundiform, and pubic tubercle, with complete resolution of his genital pain. At last follow-up, he was taking oral erectogenic agents (silde-

nafil) for the management of his erectile dysfunction, with excellent results.

CASE 3

Patient 3 was a 39-year-old man who had fled from a West African country after being captured by rebels and forced to do hard labor. During captivity, he was severely beaten in the genital area with fists and sticks on multiple occasions. He presented with complaints of erectile dysfunction and lower urinary tract symptoms (International Prostate Symptom Score of 23) characterized by decreased force of stream, incomplete emptying, and urinary frequency. He achieved an approximately 50% erection, which had poor spontaneity and sustaining capabilities. His past medical history was only remarkable for a urethral stricture managed endoscopically several years previously. His physical examination was unremarkable, but urethroscopy revealed a tight bulbar urethral stricture. Penile duplex Doppler ultrasonography after intracavernosal injection of vasoactive agents revealed a peak systolic velocity of 31 and 15 cm/s for the right and left cavernosal artery, respectively, with normal end-diastolic velocities. These findings were consistent with pure (no corporeal occlusive dysfunction) cavernosal artery insufficiency, most likely secondary to blunt perineal trauma because of torture. He underwent internal (endoscopic) urethrotomy of a long bulbar stricture. However, the stricture recurred within 6 months, requiring a second internal urethrotomy with excellent results. The patient also began taking sildenafil, with excellent results. A detailed vascular evaluation was obtained to assess the feasibility of penile revascularization from the dorsal to the cavernosal artery in an attempt to reestablish normal erectile function.

COMMENT

Little has been written about the sexual trauma of men within the published medical reports on torture. Attention has so far mainly been devoted to female survivors of sexual trauma. The myth that men are only aggressors and not victims seems prevalent.[9] The legal definition of rape in many countries and in some states in the United States dismisses that men may be victims.[10] The prevalence of male sexual trauma is uncertain. Estimates, however, have shown that 5000 to 8000 men were raped in the former Yugoslavia and that thousands of men and boys were raped during the Iraqi invasion of Kuwait.[11] Male sexual trauma mostly takes place during detention and is perpetrated by guards, interrogators, or other prisoners. Agger[12] found that 52% of male political prisoners who sought help after torture had been sexually abused. Peel *et al.*[13] specifically studied male sexual trauma among Tamil refugees, who had been imprisoned, and found a prevalence of 21%. In a different study, Peel[8] studied 607 men from 45 countries, of whom 25% had been sexually assaulted. Of these, 21% had been raped, 47% had had assaults to the genitals, 27% had had electric shocks to the genitals, and/or 21% had had an object inserted in the anus or urethral meatus. These prevalence figures may be underestimates because of the taboo and stigmatization to discussing male sexual trauma inherent in most cultural norms. A study of British male victims of sexual assaults in general showed that 79% of raped men sought no help for a mean time of 16 years after the assault.[14]

The sequelae of sexual torture include both psychological[15–17] and physical manifestations.[15,18,19] In this report, we focused on the physical sequelae. When reviewing published reports, we found few studies describing the physical sequelae of sexual trauma.[15,18,19] Blunt genital trauma from other causes has, however, been documented. Erectile dysfunction has resulted from innocent falls or blows to the crotch.[20,21] Avid bikers can develop erectile dysfunction and groin pain owing to prolonged pressure on the pudendal nerves[22] and arteries.[23,24] Earlier studies and our work have suggested that blunt genital trauma may roughly result in four clinical problems that are not mutually exclusive:

1. Chronic genital pain, including erectile pain
2. Peyronie's disease characterized by penile pain, penile curvature, and erectile dysfunction
3. Lower urinary tract symptoms
4. Erectile dysfunction

These clinical problems arise from different pathologic mechanisms. Chronic genital pain may result from injury to the inguinal cord, suspensory and/or fundiform ligaments, or tunica albuginea, leading to Peyronie's disease.[25] Peyronie's disease is characterized by decreased tunica elasticity and compliance secondary to fibrosis and plaque formation. Lower urinary tract symptoms may be due to urethral trauma leading to stricture formation. Apart from Peyronie's disease, erectile dysfunction may result from compression injuries of the pudendal arteries as they enter the perineum through Alcock's canal and/or corporal damage leading to corporeal veno-occlusive dysfunction (venous leak). Depending on the symptoms, the diagnosis of male genital trauma is based on the physical examination and penile duplex Doppler ultrasound findings. Not yet described in published reports, we propose that chronic genital pain due to Peyronie's disease, pubitis, or suspensory ligament injury can be successfully treated with steroid injections. Penile curvature can be managed medi-

cally (verapamil, interferon, colchicine, vitamin E)[25] or surgically (Nesbit plication versus tunica incision/excision and grafting).[25] Urethral strictures can be generally treated. The treatments for erectile dysfunction include oral erectogenic agents (sildenafil, vardenafil, tadalafil),[26] penile revascularization procedures, intracavernosal therapy, and penile implants.[27–29] Erectile dysfunction is most common in survivors who have been exposed to sexual trauma, but it may also arise after nonsexual torture.[30]

Sexual trauma may also lead to psychological disturbances, including post-traumatic stress disorder and major depressive disorder. These disorders include symptoms such as insomnia, nightmares, irritability, avoidance behavior, and depression,[15–17] according to the DSM-IV. Post-traumatic stress disorder and/or major depressive disorder occur in most patients who have been tortured,[15,17] and the recognition and treatment of the underlying psychological disorder is essential to therapeutic success.

Torture survivors may be exposed to human immunodeficiency virus through contaminated blood, anal rape, or unsterile instruments of torture, or when seeking medical care for wounds that occurred during torture. Thus, all male rape victims and torture survivors with male genital trauma should be screened for human immunodeficiency virus.

CONCLUSIONS

The results of this study and others have indicated that male sexual trauma is an important issue in refugee and asylum seeker populations. Still, the subject of male sexual trauma has rarely been addressed in published reports and is little understood. This is partly a result of the silence of the victims, as well as lack of awareness by, or discomfort among, medical professionals. The apparent prevalence and the severity of the physical and mental sequelae to sexual trauma, however, make it an important area for screening when treating survivors of torture and refugee trauma. It is, therefore, crucial that health professionals working with asylum seeker and refugee populations are aware of the problem and trained to recognize the physical and psychological symptoms and make appropriate referrals to knowledgeable providers or treatment centers (National Consortium of Torture Treatment Programs). This includes referral to a urologist in the case of urologic complications. Also, urologists need to be aware of the diagnostic tools and possible treatment strategies. In this context, our study is the first of its kind to document the urologic complications occurring after sexual torture in a U.S. cohort of tortured men, including the prevalence, diagnosis, and treatment. The proposed use of steroid injections in the clinical treatment of these patients has not been previously reported.

REFERENCES

1. Available at: http://www.refugees.org/world/statistics Accessed on April 15, 2004.
2. Piwowarczyk L, Moreno A, and Grodin M: Health care of torture survivors. JAMA 284: 539–541, 2000.
3. Crosby S, Norredam M, Piwowarczyk L, *et al:* Prevalence of torture survivors among foreign-born patients at an urban ambulatory care practice. J General Internal Medicine Suppl 19: 198, 2004 (abstract).
4. Moreno A, Piwowarczyk L, and Grodin M: Health care of torture survivors: human rights violations and refugee health. JAMA 285: 1215, 2001.
5. Kaufman A, Divasto P, Jackson R, *et al:* Male rape victims: non institutionalized assault. Am J Psychiatry 137: 221–223, 1980.
6. United Nations Convention Against Torture and Other Cruel, Inhuman and Degrading Treatment or Punishment. General Assembly resolution 39/46 Available at: http://www.ircr.org/usr/ircr/home.nsf/unid. Accessed on May 1, 2004.
7. Lunde I, and Ortmann J: Prevalence and sequelae of sexual torture. Lancet 336: 289–291, 1990.
8. Peel M: Male sexual abuse in detention, in Peel M, and Iacopino V (Eds): *The Medical Documentation of Torture.* London, Greenwich Medical Media, 2002, pp 179–190.
9. Hardy C: An act of force: male rape victims. TORTURE: Q J Rehab Torture Victims Prev Torture 12: 19–23, 2002.
10. Steiner CE: Sexual assault on men in war. Lancet 349: 129, 1997.
11. James J: Silent conflict—helping the survivor of sexual violence. Emerg Nurse 9: 15–18, 2001.
12. Agger I: Sexual torture of political prisoners: an overview. J Traumatic Stress 2: 305–308, 1989.
13. Peel M, Mahtani A, Hinshelwood G, *et al:* The sexual abuse of men in detention in Sri Lanka. Lancet 355: 2069–2070, 2000.
14. King M, and Woollett E: Sexually assaulted males: 115 consulting a counselling service. Arch Sex Behav 26: 579–588, 1997.
15. Goldfeld AE, Mollica RF, Pesavento BH, *et al:* The physical and psychological sequelae of torture: symptomatology and diagnosis. JAMA 259: 2725–2729, 1988.
16. Basoglu M, Paker M, Paker O, *et al:* Psychological effects of torture: a comparison of tortured with nontortured political activists in Turkey. Am J Psychiatry 151: 76–81, 1994.
17. Eisenman DP, Gelberg L, Liu H, *et al:* Mental health and health-related quality of life among adult Latino primary care patients living in the United States with previous exposure to political violence. JAMA 203: 667–670, 2003.
18. Daugaard G, Petersen HD, Abildgaard U, *et al:* Sequelae to genital trauma in torture victims. Arch Androl 10: 245–248, 1983.
19. Lunde I, Rasmussen OV, Lindholm G, *et al:* Gonadal and sexual functions in tortured Greek men. Dan Med Bull 27: 243–245, 1980.
20. Munarriz RM, Yan QR, Nehra A, *et al:* Blunt trauma: the pathophysiology of hemodynamic injury leading to erectile dysfunction. J Urol 153: 1831–1840, 1995.
21. Munarriz RM, LaSalle MD, and Goldstein I: Penile revascularizaton for treatment of erectile dysfunction secondary to blunt perineal trauma. Urology 61: 222–223, 2003.

22. Richuutti VS, Haas CA, Seftel AD, *et al*: Pudendal nerve injury associated with avid bicycling. J Urol 162: 2099–2100, 1999.

23. Marceau L, Kleinman K, and Goldstein I: Does bicycling contribute to the risk of erectile dysfunction? Results from the Massachusetts Male Aging Study (MMAS). Int J Impotence Res 13: 298–302, 2001.

24. Schrader SM, Breitenstein MJ, Clark JC, *et al*: Nocturnal penile tumescence and rigidity testing in bicycling patrol officers. J Androl 23: 927–934, 2002.

25. Gholami SS, Gonzalez-Cadavid NF, Lin CS, *et al*: Peyronie's disease: a review. J Urol 169: 1234–1241, 2003.

26. Gresser U, and Gleiter CH: Erectile dysfunction: comparison of efficacy and side effects of the PDE-5 inhibitors sildenafil, vardenafil and tadalafil: review of the literature. Eur J Med Res 7: 435–446, 2002.

27. Goldstein I: Penile revascularization. Urol Clin North Am 14: 805–813, 1987.

28. Montorsi F, Salonia A, Deho' F, *et al*: Pharmacological management of erectile dysfunction. BJU Int 91: 446–454, 2003.

29. Milbank AJ, and Montague DK: Surgical management of erectile dysfunction. Endocrine 23: 161–165, 2004.

30. King M: Male rape. BMJ 301: 1345–1346, 1990.

Addressing Smoking Cessation in Methadone Programs

Yngvild Olsen, MD, MPH
Daniel P. Alford, MD, MPH
Nicholas J. Horton, ScD
Richard Saitz, MD, MPH

ABSTRACT. Little is known about how well methadone programs address smoking cessation. We describe the frequency of smoking cessation counseling, and factors affecting its provision in methadone programs. We conducted a cross-sectional survey of methadone patients and their counselors. Of 575 patients, 76% were eligible smokers. Although only 48% of patients reported receiving smoking cessation counseling within the previous six months, 97% of counselors reported providing it ($p < 0.0001$). Time with one counselor was significantly associated with pa-

Yngvild Olsen is Assistant Professor of Medicine, Division of General Internal Medicine, Johns Hopkins University School of Medicine.

Daniel P. Alford is Assistant Professor of Medicine, Boston University School of Medicine.

Nicholas J. Horton is Assistant Professor, Department of Mathematics, Smith College.

Richard Saitz is Associate Professor of Medicine and Co-Director of the Clinical Addiction, Research, and Education Unit, Boston University School of Medicine, and affiliated with the Department of Epidemiology, Boston University School of Public Health.

Address correspondence to: Yngvild Olsen, MD, MPH, Division of General Internal Medicine, Johns Hopkins School of Medicine, 2024 East Monument Street, Room 2-516, Baltimore, MD 21287 (E-mail: yolsen@jhsph.edu).

The authors gratefully acknowledge Suzette Levenson and the Boston University School of Public Health Data Coordinating Center, Maria Fung, Craig Noronha, Erin Finley, and Adebola Olanrewaju, and the patients and staff at the three participating methadone programs.

This study was funded by the American Cancer Society (IRG-72-001-26-IRG, Principal Investigator Daniel P. Alford).

Journal of Addictive Diseases, Vol. 24(2) 2005
http://www.haworthpress.com/web/JAD

Digital Object Identifier: 10.1300/J069v24n02_04

tient report of receiving smoking cessation counseling (OR 1.19 [95% CI 1.04-1.36]). Although addiction counseling is required in methadone programs, nicotine addiction is addressed less than half the time. Methadone programs should prioritize the provision of effective smoking cessation and facilitate continuity of patient-counselor relationships.

KEYWORDS. Smoking, smoking cessation, methadone maintenance, health services research, substance abuse treatment

INTRODUCTION

Cigarette smoking is a leading cause of preventable death among substance abusers in treatment,[1,2] and the prevalence of smoking in these populations often exceeds 80%.[3,4] Historically, treatment providers have hesitated to focus on smoking cessation because of fear of compromising sobriety.[5] However, several studies among alcoholics suggest that tackling both nicotine and alcohol dependence simultaneously leads to longer periods of sobriety, and less relapse to alcohol.[6,7] In the general population, counseling and pharmacotherapy for nicotine dependence are effective.[8,9] Several small studies suggest that smoking cessation therapy works among alcohol and drug dependent patients in treatment as well.[10-12]

Approximately 20% of the 900,000 heroin addicts in the United States receive methadone maintenance treatment.[13] According to federal regulation, methadone maintenance treatment includes regular addiction counseling.[14] Close to 95% of methadone maintenance patients smoke cigarettes.[15-17] Many smokers on methadone maintenance show an interest in quitting cigarettes,[3,15,16] but little is known about the extent to which smoking cessation counseling occurs in these settings. We undertook the current study to test two hypotheses; first, that smoking cessation counseling occurs infrequently in methadone programs, and second, that counselors and patients in methadone programs differ in their beliefs about the effect of smoking cessation on recovery from heroin addiction. We also assessed which factors are associated with patients receiving smoking cessation counseling as part of methadone maintenance treatment.

MATERIALS AND METHODS

Setting

The study was conducted at three large, urban methadone programs that provide comprehensive pharmacological maintenance therapy and counseling services to a large spectrum of opioid dependent persons in the greater Boston area. The Institutional Review Board at Boston University Medical Center approved the study. All subjects provided informed consent.

Subject Inclusion Criteria

Patients

Eligible subjects were current smokers, defined as having smoked at least 100 cigarettes in their lifetime and at least 1 cigarette in the last month, had met with their assigned addiction counselor at least once, and understood English.

Counselors

All counselors caring for at least one methadone patient in any one of the three study programs were eligible for inclusion. To ensure that patients at all program sites would have had a chance to meet with their counselor at least once, counselors who began work with eligible patients on or after the date subject enrollment began were excluded. Counselors who left study sites within one month prior to patient enrollment were included in the study, as many patients would have had at least one visit with these counselors prior to transitioning to new therapists.

Measures

Patient-Level Predictor Variables

Patient interviews assessed demographics, smoking and drug use history, readiness to quit smoking, and beliefs about smoking cessation and relapse to heroin. Using questions based on the trans-theoretical model of behavior change, patient readiness to quit smoking included assessment of motivation, importance, confidence, and interest in quit-

ting within 6 months, using 4-point Likert scales.[15,18] In addition, patients rated their readiness to change on a 5-level contemplation scale, using a previously validated 10-point contemplation ladder.[19] To assess patients' beliefs about smoking cessation counseling and relapse to heroin, we asked patients to rate their agreement with the statement "the stress of quitting smoking would make me relapse to heroin" using a 4-point Likert scale from *strongly disagree* to *strongly agree.*

Counselor-Level Predictor Variables

Counselor surveys assessed demographics, smoking history, and beliefs about smoking cessation and relapse to heroin. We asked counselors to rate their agreement with the statement "the stress of quitting smoking would make my methadone maintenance patients relapse to heroin" using a 4-point Likert scale from *strongly disagree* to *strongly agree.* As much as possible, the wording for all questions that appeared on both patient and counselor questionnaires was identical.

Patient- and Counselor-Level Outcome Variables

To assess patient-reported smoking cessation counseling, patients rated, on a 5-point Likert scale from *always* to *never*, the frequency with which their counselor had, over the previous 6 months: (1) asked whether they smoked, (2) asked whether they felt ready to quit, (3) discussed the adverse health effects of smoking, (4) advised them to quit, and (5) discussed smoking cessation options.

To assess counselor-reported provision of smoking cessation counseling, we used the same approach. Counselors reported, on a 5-point Likert scale from *always* to *never*, the frequency with which they had, over the previous 6 months: (1) asked patients whether they smoked, (2) asked patients if they felt ready to quit, (3) discussed with patients the adverse health effects of smoking, (4) advised patients to quit, and (5) discussed with patients smoking cessation options. Counselors were asked to respond to each item considering all of their smoking methadone patients.

Data Collection Procedures

Patients

From July through December 2000, four trained research assistants approached patients waiting for methadone dosing or counseling ses-

sions, and asked them to participate in a study about smoking cessation in methadone programs. Willing patients answered three screening questions to determine eligibility for entry into the study. Research assistants administered 23-item surveys by interview to enrolled subjects immediately following the screening interview, and recorded the data anonymously. Patients received $2 gift certificates towards the purchase of food as compensation for participation. Research assistants approached patients during a combination of dosing and counseling times in each clinic, thus ensuring recruitment of patients with and without take-home privileges and from early morning to evening hours of clinic operation.

Counselors

Two investigators met with addiction counselors at a usual weekly staff meeting at each site during the first 3 months of the study to distribute 20-item, self-administered surveys, ensure confidentiality, and collect completed questionnaires. Questionnaires were mailed or hand-delivered to counselors not present at these staff meetings, and collected several days later.

Analysis

In order to link patients and counselors, patients identified their primary addiction counselor. Proportions, means, and 95% confidence intervals (CI) were calculated to describe the sample. We defined a dichotomous summary measure of receipt or provision of smoking cessation counseling, for patients and counselors respectively, as having answered "always," "often," or "sometimes" to any one of the five questions involving receipt or provision of smoking cessation counseling. Having infrequently received or provided smoking cessation counseling, respectively, was then defined as having answered "rarely" or "never" to all five questions.

We dichotomized the five individual components of smoking cessation counseling into *ever* versus *rarely/never* in the same fashion as the summary measure. Comparisons between patient and counselor report of continuous variables (e.g., cigarette consumption) were made using two-sample t-tests, while comparisons of categorical variables (e.g., smoking cessation counseling) were made using chi-squared tests, both with an alpha level of 0.05.

Generalized estimating equation regression models were used to assess factors associated with patient report of counseling.[20,21] These models account for the clustering of patients within counselors. The GEE was fit using a logit link function and an exchangeable working correlation structure. In developing models, we considered for inclusion patient and counselor demographics, smoking histories, methadone program experience, and patient readiness to change smoking behavior, and chose other variables with potential clinical significance: length of time with a counselor, patient daily cigarette consumption, counselor smoking status, duration of time counselors had worked in the addictions field, counselor training in specific smoking cessation counseling, counselor caseload, percentage of patients that smoke, percentage of patients on methadone maintenance, patient readiness to quit as indicated by the modified contemplation ladder, and patient and counselor beliefs about smoking cessation and relapse to heroin. Models were also fit with other potential confounders: program site, patient age, patient and counselor gender, patient and counselor race/ethnicity, patient recent use of illicit drugs, patient employment status, and counselor education. No adjustments were made for multiple comparisons.

Kappa statistics were calculated to assess agreement between patient and counselor report of smoking cessation counseling, using first, the individual, dichotomized components of counseling and, second, the summary measure for both patients and matched counselors. These analyses were repeated stratified by counselor smoking status.

All analyses were done using SAS statistical software, Release 8.2 (SAS Institute Inc, Cary, NC).

RESULTS

Of 575 methadone patients screened, 438 (76%) were smokers eligible for inclusion into the study. All 438 eligible subjects completed questionnaires. There were 42 eligible counselors, all of whom completed surveys. At least half of both patients and counselors were female, both groups were mostly white, and on average in their 40s. Despite methadone treatment, 44% of patients reported using some form of illicit drug in the previous month. Patients had been with the same counselor for a median of 12 months. Although patients and counselors had a similar number of prior quit attempts, only 31% (13/42) of counselors were current smokers, and this group smoked significantly fewer cigarettes than patients ($p < 0.001$). Less than 25% of patients and

counselors felt that stress caused by quitting cigarettes would lead to relapse to heroin (Table 1).

Although no patients were actively taking steps to quit smoking cigarettes, 58% of the 408 patients responding to this question indicated that they were contemplating or determined to give up nicotine ("think I should quit but not ready" or "think I need to consider quitting"), and

TABLE 1. Patient and Counselor Demographic Characteristics and Smoking History

Characteristic	Patients (N = 438)	Counselors (N = 42)
Female, %	50	62
Race/ethnicity, %		
White	64	81
Black	21	10
Hispanic	13	7
Other	2	2
Age, mean years (± SD)	42 (SD 8)	43 (SD 12)
Educational level	Mean 12 years (SD 2)	Master's level 76%
Unemployed, %	53*	N/A
Income, %		N/A
< $20,000	46	
$20-$50,000	26	
> $50,000	9	
Illicit drug use in past 30 days, %	44	N/A
Time in clinic, %		N/A
0-< 12 months	21	
1-< 5 years	49	
5+ years	30	
Months with counselor, median (range)	12 (1-156)	N/A
Cigarettes smoked per day, median (range)	20 (0-70)	5 (1-25)†
No. of prior quit attempts, median (range)	2 (0-50)	3 (0-17)†
Years of smoking, mean (± SD)	26 (SD 9)	N/A
Stress of quitting smoking would cause relapse to heroin, % (95% CI)‡	20 (16-24)	21 (9-34)

*N = 430
†among the 13 (31%) of current smokers
‡somewhat to strongly agree

33% were thinking about how to change smoking patterns. Only 9% were pre-contemplative, in that they had no thought of quitting. Data from this question was missing for 7% of the sample.

Weekly counselor caseloads averaged 28 patients, with a mean of 86% on methadone maintenance. Counselors estimated that, on average, 78% of their patients smoked cigarettes. Only 24% of addiction counselors had received training in smoking cessation counseling.

Overall, 48% of patients reported receiving smoking cessation counseling at least sometimes from their counselors over a six-month period. In comparison, over the same time period, 97% of counselors reported providing smoking cessation counseling at least sometimes to patients ($p < 0.0001$). Of note, no significant chance-corrected agreement was found between patient and counselor reports of smoking cessation counseling ($\kappa = -0.01$, 95% CI -0.04-0.03). For example, although only 28% (120/436) of patients reported having been asked by their counselor at least sometimes within the past six months whether they smoke, 71% (30/42) of counselors reported having asked their patients about smoking at least sometimes during the same time period ($p < 0.0001$). Only 31% (137/437) of patients reported having been advised at least sometimes to quit, compared to 64% (27/42) of counselors stating that they had at least sometimes provided this advice ($p < 0.0001$). Similarly, significantly fewer patients reported at least sometimes discussing the adverse health effects of smoking (32% vs. 81%, $p < 0.0001$) and talking about smoking cessation options (26% vs. 81%, $p < 0.0001$) compared with counselor reports.

In analyses adjusted for clustering of patients within counselors and patient and counselor demographic and smoking characteristics, each additional year spent in a therapeutic relationship with one counselor was significantly associated with greater odds of patient report of having received counseling (OR 1.19 [95% CI 1.04-1.36]) (Table 2). Patients with master's level counselors had lower odds of reporting having received counseling compared with patients with non-master's level counselors (OR 0.35 [95% CI 0.20-0.63]). In addition, as counselors' percentage of methadone maintenance patients increased, the odds of patients reporting having received smoking cessation counseling decreased (OR 0.98 [95% CI 0.97-1.00]). Readiness to change (OR 1.43 [95% CI 0.59-3.47] for pre-contemplation, OR 0.67 [95% CI 0.36-1.27] for contemplation, OR 1.05 [95% CI 0.62-1.78] for determination, compared with preparation), counselor smoking status (OR 0.79 [95% CI 0.37-1.67] for never smoked, OR 0.78 [95% CI 0.52-1.17] for quit,

TABLE 2. Adjusted odds ratios of patient report of having received smoking cessation counseling and patient and counselor characteristics*

Characteristic	OR	95% CI
Clinic		
C	1.0 (ref)	
A	0.65	0.30-1.38
B	0.75	0.29-1.95
Patient age	1.01	0.98-1.04
Patient gender		
Male	1.0 (ref)	
Female	1.10	0.62-1.94
Race/ethnicity		
White	1.0 (ref)	
Black	0.74	0.41-1.34
Hispanic	0.76	0.39-1.50
Illicit drug use in past month		
Yes	1.0 (ref)	
No	1.14	0.66-1.96
Daily cigarette consumption	1.01	0.99-1.02
Duration seeing one counselor (yrs)†	1.19	1.04-1.36
Patient employment status		
Unemployed	1.0 (ref)	
Employed	1.37	0.90-2.09
Patient readiness to change stage		
Preparation	1.0 (ref)	
Determination	1.05	0.62-1.78
Contemplation	0.67	0.36-1.27
Pre-contemplation	1.43	0.59-3.47
Patient belief that smoking cessation causes relapse to heroin		
Disagree	1.0 (ref)	
Agree	1.11	0.65-1.90
Counselor Gender		
Male	1.0 (ref)	
Female	0.86	0.53-1.40

TABLE 2 (continued)

Characteristic	OR	95% CI
Counselor race/ethnicity		
White	1.0 (ref)	
Black/Hispanic	0.51	0.25-1.04
Counselor education†		
Non-master's	1.0 (ref)	
Master's	0.35	0.20-0.63
Months worked as addictions counselor	1.0008	0.998-1.004
Counselor smoking status		
Quit	1.0 (ref)	
Current	0.78	0.52-1.17
Never	0.79	0.37-1.67
Caseload (No. of patients counselors estimate seeing per week)	1.03	1.00-1.06
Percentage of smoking patients	0.99	0.98-1.00
Percentage of patients on methadone maintenance†	0.98	0.97-1.00
Having received training in smoking cessation counseling		
Yes	1.0 (ref)	
No	1.00	0.65-1.55
Counselor belief that smoking cessation causes relapse to heroin		
Disagree	1.0 (ref)	
Agree	1.16	0.71-1.90

*Parameter estimates derived from a clustered regression model of N = 431 (7 counselors with race as other were excluded)
†$p < 0.05$

compared with current smoker), and the belief that quitting smoking may cause relapse to heroin (OR 1.16 [95% CI 0.71-1.90]) were not significantly associated with patient report of having received smoking cessation counseling.

DISCUSSION

In a group of patients receiving regular addiction counseling, we found that effective smoking cessation counseling is not a routine part

of this therapy. Fewer than half of smokers reported receiving smoking cessation counseling from their counselors at least sometimes during a six-month period. Only about a quarter to one-third of patients reported having at least sometimes been advised to quit or engaged in discussions with their counselor about the adverse health consequences of smoking or options for smoking cessation. Less than a third of patients reported just being asked whether they smoke. These findings significantly contrasted with the majority of counselors who reported providing frequent smoking cessation counseling to their patients over the same time period.

The belief that stress caused by smoking cessation would compromise recovery from heroin addiction did not appear to act as a strong deterrent for patients and counselors to address nicotine dependence. Contrary to our hypothesis, patients and counselors agreed that quitting cigarettes would not cause relapse to heroin abuse. We also found that, contrary to prior work, counselors' smoking status was not significantly associated with patients' report of having received smoking cessation counseling. Bobo and Davis found that counselors in rural Nebraska who were former smokers were six times more likely to routinely urge their patients to quit smoking compared with counselors who currently smoked.[22] Public perception around smoking and smoking cessation has changed over the last ten years, which may explain the disparate results we found compared with those of Bobo and Davis.

Clinically surprising, patients' level of readiness to change did not affect patients' report of having received smoking cessation counseling. Given the empiric evidence for the importance of this clinical measure on smoking quit rates,[18] we would have expected to find a difference between the various readiness to change levels in patients' report of having received smoking cessation counseling. It may be that the limited number of subjects on either end of the readiness to change scale left the majority of responders in the middle of the scale where differences in their recall of counselor advice about smoking may not have been statistically evident.

Interestingly, length of time patients had spent with the same counselor was significantly associated with patients' report of having received smoking cessation counseling. In fact, patients had 1.19 times the odds of receiving smoking cessation counseling for every additional year spent in a therapeutic relationship with their counselor. One explanation for this could be that a counselor will only tackle smoking once he/she feels other perceived priorities have adequately been addressed. This follows earlier work by Bobo and Gilchrist on the timing of smok-

ing cessation for substance abusers in treatment. They reported that 53% of 311 staff members at several inpatient alcohol treatment centers favored addressing smoking cessation only after the first year of sobriety.[5] Alternatively, patients who are more advanced in their recovery may be more open to hearing and retaining smoking cessation counseling.

Controlling for caseload, we also found that as the percentage of methadone maintenance patients each counselor was assigned increased, the odds of patients reporting having received smoking cessation counseling decreased. It seems clinically plausible that busy counselors dealing with substance abuse issues as well as methadone may address smoking cessation differently than those counselors with a lower percentage of methadone maintained patients in their caseload. This borderline significant finding must be interpreted with caution, however, because of multiple comparisons.

Surprisingly, those patients with master's level counselors had 0.35 times the odds of reporting having received counseling, while simultaneously controlling for years as counselor and counselor age. No previous work has documented such a difference. This finding should be confirmed in other studies and explored further.

Several studies have documented patient interest in quitting cigarettes, both in the inpatient and outpatient settings. Orleans and Hutchinson reported that 42% of 118 patients in a residential treatment program had a strong interest in smoking cessation.[23] Richter et al. found that, among 550 methadone maintenance patients, 80% were somewhat or very interested in quitting smoking.[16] Other researchers, including those in this study, have noted similar findings.[3,15,24]

While our study did not assess the effect of a smoking cessation intervention, several reports document the positive impact of nicotine dependence therapy among patients receiving other substance abuse treatment.[10-12,25] To our knowledge, however, our study is one of the first to report simultaneous counselor-level and patient-level reports of the current state of smoking cessation counseling in outpatient addiction treatment clinics. It is also one of the few studies documenting patient reports of smoking cessation counseling.

Counselor reports of smoking cessation counseling have been documented before. In a 1993 survey of staff in recovery at a chemical dependency program in Nebraska, 44% of counselors reported having advised at least one patient to quit and 25% stated that they routinely did so.[22] Geographic and time differences between our study and that of Bobo and Davis may account for the lower rates of counselor-reported

advice to quit found in their study. Gill and Bennett noted that smoking cessation training heavily influenced counselor-reported practices with regards to nicotine dependence in their patients.[26] While over 50% of counselors with smoking cessation training addressed smoking with at least 50% of their patients, only 5% of those with little to no training did the same.

The large discrepancy between patient and counselor reports of smoking cessation counseling noted in our study may cause some to wonder where the truth lies. While under- and over-reporting of smoking cessation counseling may have occurred from patients and counselors, respectively, it seems clear that if counselors delivered smoking cessation counseling, it was not effective in reaching their patients. For effective behavior change, whether it involve diet, excessive drug use, or smoking, it is crucial that patients recall whether and what counseling they have received from their providers. In a study of alcoholic workers advised to quit alcohol by their health care providers, recall of the warning was associated with improved drinking outcomes two years later.[27] In a general medical population, recall of smoking cessation advice was associated with high numbers of quit attempts, particularly among patients with serious health problems.[28] In our study, only 24% of addiction counselors reported receiving training in smoking cessation counseling and therefore may not know how to deliver messages about pharmacotherapy or the long-term risks of smoking effectively.

There are limitations to our study. There may be different timeframes for recall for patients and counselors since they were asked about the previous six months but often a month or two apart. There also may be a social desirability bias leading counselors to over-report the amount of counseling they provided. In addition, our sample may not be representative of all methadone maintenance patients who smoke because our recruitment methods may have missed eligible patients who did not have to wait for methadone dosing or counseling sessions. It is likely that our sample represents the majority of patients in these settings, however, for four reasons: (1) most patients wait for at least a few minutes when attending dosing or counseling sessions; (2) we enrolled patients from all possible clinic hours; (3) most patients screened were eligible for the study; and (4) we enrolled all eligible patients screened. Although only three sites were included in the study, these public and private clinics cover a large number of patients on methadone in the greater Boston area, with patient populations similar to those in other urban centers where most methadone programs are located. Finally, counselors were asked about their patients as a group, whereas patients

were asked about particular counselors. Counselors and patients thus may have interpreted questions differently, which may explain some of the discrepancy noted in the frequency of smoking cessation counseling reported by both groups. However, both counselors and patients were asked about the frequency of counseling in the same fashion and using nearly identical language. Since multiple comparisons were involved in our analyses without adjustments made, results reaching statistical significance with p-values near $p = 0.05$ should be used with caution.

Methadone maintenance patients are at extremely high risk for morbidity and mortality from smoking-related conditions, for which effective preventive measures exist. In fact, in the long-term, they are more likely to die of the consequences of smoking than of addiction to other drugs. Fortunately, in methadone programs, the opportunities to address the high prevalence of smoking are numerous. Patients are interested and motivated to learn about smoking cessation, and, as our study shows, concerns about compromising recovery do not appear to be common among patients or counselors.

Although the vast majority of counselors in our study reported providing smoking cessation counseling, a minority of patients reported having received it. In addition, a small percentage of counselors reported having received specific training in smoking cessation-counseling techniques. These results imply that effective smoking cessation counseling is not routinely occurring in these patient-provider interactions. It appears that the most important factor associated with effective smoking cessation counseling is a patient's continuous relationship with one counselor. Methadone programs should prioritize smoking cessation as part of their overall addiction counseling, and should facilitate stable, long-term counselor-patient relationships to increase the occurrence of smoking cessation counseling. Methadone programs should also invest in training their counselors to ensure that smoking cessation counseling is effectively delivered in order to effect behavior change in their patients. Future research should focus on using qualitative and quantitative methods to identify the educational components needed by counselors for effective smoking cessation counseling, assess the effectiveness of smoking cessation counseling in methadone programs, and study better ways to disseminate interventions for the leading cause of mortality among adults receiving opioid dependence treatment: tobacco.

REFERENCES

1. Hurt RD, Offord KP, Croghan IT, Gomez-Dahl L, Kottke TE, Morse RM, Melton LJ, III. Mortality Following Inpatient Addictions Treatment: Role of Tobacco Use in a Community-Based Cohort. JAMA. 1996;275:1097-103.

2. Hser YI, McCarthy WJ, Anglin MD. Tobacco Use as a Distal Predictor of Mortality Among Long-Term Narcotics Addicts. Prev. Med. 1994;23:61-9.

3. Clarke JG, Stein MD, McGarry KA, Gogineni A. Interest in Smoking Cessation Among Injection Drug Users. Am. J. Addict. 2001;10:159-66.

4. McIlvain HE, Bobo JK. Tobacco Cessation with Patients Recovering from Alcohol and Other Substance Abuse. Prim. Care. 1999;26:671-89.

5. Bobo JK, Gilchrist LD. Urging the Alcoholic Client to Quit Smoking Cigarettes. Addict. Behav. 1983;8:297-305.

6. Burling TA, Marshall GD, Seidner AL. Smoking Cessation for Substance Abuse Inpatients. J. Subst. Abuse. 1991;3:269-76.

7. Hurt RD, Eberman KM, Croghan IT, Offord KP, Davis LJ, Jr, Morse RM, Palmen MA, Bruce BK. Nicotine Dependence Treatment During Inpatient Treatment for Other Addictions: A Prospective Intervention Trial. Alcohol Clin. Exp. Res. 1994; 18:867-72.

8. Kottke TE, Battista RN, DeFriese GH, Brekke ML. Attributes of Successful Smoking Cessation Interventions in Medical Practice. A Meta-Analysis of 39 Controlled Trials. JAMA. 1988;259:2883-9.

9. A Clinical Practice Guideline for Treating Tobacco Use and Dependence: A US Public Health Service Report. The Tobacco Use and Dependence Clinical Practice Guideline Panel, Staff, and Consortium Representatives. JAMA. 2000;283:3244-54.

10. Saxon AJ, McGuffin R, Walker RD. An Open Trial of Transdermal Nicotine Replacement Therapy for Smoking Cessation Among Alcohol- and Drug-Dependent Inpatients. J. Subst. Abuse Treat. 1997;14:333-7.

11. Campbell BK, Krumenacker J, Stark MJ. Smoking Cessation for Clients in Chemical Dependence Treatment: A Demonstration Project. J. Subst. Abuse Treat. 1998;15:313-8.

12. Joseph AM, Nichol KL, Willenbring ML, Korn JE, Lysaght LS. Beneficial Effects of Treatment of Nicotine Dependence During an Inpatient Substance Abuse Treatment Program. JAMA. 1990;263:3043-6.

13. Substance Abuse and Mental Health Services Administration. National Household Survey on Drug Abuse. Vol. 2001. Washington, DC: U.S. Government Printing Office, 1999.

14. Substance Abuse and Mental Health Services Administration. Approval and Monitoring of Narcotic Treatment Programs: A Guide on the Roles of Federal and State Agencies. Technical Assistance Publication Series 12. DHHS Publication No. (SMA) 94-2082. Washington, DC: U.S. Government Printing Office, 1994.

15. Clemmey P, Brooner R, Chutuape MA, Kidorf M, Stitzer M. Smoking Habits and Attitudes in a Methadone Maintenance Treatment Population. Drug Alcohol Depend. 1997;44:123-32.

16. Richter KP, Gibson CA, Ahluwalia JS, Schmelzle KH. Tobacco Use and Quit Attempts Among Methadone Maintenance Clients. Am. J. Public Health. 2001;91: 296-9.

17. Berger H, Schweigler M. Smoking Characteristics of Methadone Patients. JAMA. 1972;222:705.

18. Prochaska JO, DiClemente CC. Stages and Processes of Self-Change of Smoking: Toward an Integrative Model of Change. J. Consult. Clin. Psychol. 1983;51:390-5.

19. Biener L, Abrams DB. The Contemplation Ladder: Validation of a Measure of Readiness to Consider Smoking Cessation. Health Psychol. 1991;10:360-5.

20. Zeger SL, Liang KY. Longitudinal Data Analysis for Discrete and Continuous Outcomes. Biometrics. 1986;42:121-30.

21. Liang KY, Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models. Biometrika. 1986;73:13-22.

22. Bobo JK, Davis CM. Recovering Staff and Smoking in Chemical Dependency Programs in Rural Nebraska. J. Subst. Abuse Treat. 1993;10:221-7.

23. Orleans CT, Hutchinson D. Tailoring Nicotine Addiction Treatments for Chemical Dependency Patients. J. Subst. Abuse Treat. 1993;10:197-208.

24. Frosch DL, Shoptaw S, Jarvik ME, Rawson RA, Ling W. Interest in Smoking Cessation Among Methadone Maintained Outpatients. J. Addict. Dis. 1998;17:9-19.

25. Shoptaw S, Jarvik ME, Ling W, Rawson RA. Contingency Management for Tobacco Smoking in Methadone-Maintained Opiate Addicts. Addict. Behav. 1996;21: 409-12.

26. Gill BS, Bennett DL. Addiction Professionals' Attitudes Regarding Treatment of Nicotine Dependence. J. Subst. Abuse Treat. 2000;19:317-8.

27. Walsh DC, Hingson RW, Merrigan DM, Levenson SM, Coffman GA, Heeren T, Cupples LA. The Impact of a Physician's Warning on Recovery After Alcoholism Treatment. JAMA.1992;267:663-7.

28. McIlvain H, Susman JL, Davis, C, Gilbert C. Physician Counseling for Smoking Cessation: Is the Glass Half Empty? J. Fam. Pract. 1995;40:148-52.

chapter 8

The Challenges of Informed Consent for Low-Literate Populations

Michael K. Paasche-Orlow, MD, MA, MPH

The object of language is not to bemuse grammarians, but to convey ideas, and the more simply it accomplishes that object the more effectively it meets the needs of an energetic and practical people.

—H. L. Mencken,
The American Language, 1921.[1]

Many of the activities of health care and medical research are outside the bounds of normal social interaction. Clinicians, investigators, and the institutions they work for protect themselves against claims of impropriety with the doctrine of informed consent. For example, under normal circumstances, when a surgeon suggests an operation, the surgeon cannot proceed unless the patient agrees. This agreement must be based on a fair understanding of what to expect from the surgery. The surgeon has an obligation to teach the patient about the risks, benefits, and alternatives to the proposed procedure. With this information, the patient can decide to refuse or accept the operation.

Although patient comprehension is crucial to the informed consent process, elements of the consent process have consistently been found to be highly formal and complex.[2–12] This complexity presents a major barrier to comprehension for all patients but especially for approximately one quarter of the adult population in the United States with low literacy skills.[13]

Because informed consent is a communication process, many themes that emerge in this chapter mirror those of other chapters that explore the role of health literacy in physician-patient communication. Concern for the protection of rights for low-literate adults is especially salient in

the case of informed consent, because the ethical and legal obligations inherent in this type of communication process may not be fulfilled by typical patterns of practice. Essentially, low-literate adults are a vulnerable population with respect to informed consent.[14] This vulnerability is also a danger for providers. When the consent process is inadequate, health care providers not only fail in their professional and ethical obligation but are exposed to liability.[15–18]

While many features of the traditional informed-consent process are known to be challenging for low-literate patients, there is still much unexplored territory. This chapter evaluates the role of health literacy in various aspects of the informed-consent process. Specifically, the chapter discusses formal and informal consent; barriers experienced by low-literate patients, challenges to preserving patient rights, the reliability of informed-consent forms, and consent as a protection of human subjects. The chapter concludes with suggestions for future research and advocacy to promote the autonomy of low-literate adults.

INFORMAL AND FORMAL CONSENT

In extreme circumstances (eg, delirium, incompetence, or dangerousness), health care providers may have a positive duty and legal obligation to oppose a patient's stated wishes. However, generally, people do not cede their right of self-determination by becoming a patient. Negotiating the details of what health providers may do with patients during the course of treatment is the arena of informed consent. While much of the consent literature has focused on written documents, these represent only a small part of physician-patient interactions involving informed consent.[19] Most consent is elicited informally and sometimes nonverbally; indeed, it is the minority of instances in which informed consent is confirmed in writing. While the issue of low literacy tends to be raised exclusively for formal written consent documents, low-literate patients may have difficulty with comprehension in multiple aspects of the informed-consent process.[20]

Typically, higher medical risk (including risk of physical, economic, psychological, and social harm) is associated with a more formalized consent procedure. For example, invasive surgery is almost always accompanied by an executed informed-consent document, while many simple, low-risk procedures and tests may proceed with a head nod.[21] However, rules for what types of activities require formalized consent are difficult to discern. HIV and genetic testing are done subsequent to counseling and the completion of written informed-consent documents. However, this is not typical for other tests like syphilis testing, despite mandatory reporting laws and the implicit risk for loss of privacy.

The level of risk that will trigger the use of more formalized consent procedures is highly variable. Myriad factors—including local convention,

legal constraints, personal preference, and training—influence the type of informed-consent process that is considered appropriate for various clinical interactions. Patients cannot assume that the absence of a formal process implies the relative safety or appropriateness of clinical decisions. For example, medication regimens with important risks and side effects are often initiated without a signed document in primary care.

BARRIERS EXPERIENCED BY LOW-LITERATE PATIENTS IN THE INFORMED-CONSENT PROCESS

Theoretical and empirical evaluations of informed consent and decision-making have established the elements needed to ensure that patients have attained substantive and meaningful informed consent.[17,21–25] Table 8-1 exhibits these elements and notes specific ways patients with low literacy may have difficulty with the consent process.

Decision-Making Capacity

According to Table 8-1, a starting condition is assessment of decision-making capacity. Because low literacy has been shown to decrease performance on several commonly used neuropsychiatric tests, it is possible that some low-literate adults, particularly the elderly, will be wrongly classified as incompetent.[26–28] Misclassification may result in fewer attempts to obtain informed consent directly from the patient and greater reliance on family members or other proxies.

Process of Decision-Making

Next, Table 8-1 represents informed consent as a process that includes multiple levels of communication. An underlying feature of patient-physician communication involves establishing the model of relationship that will guide the informed-consent process; in paternalistic decisions the physician is dominant; the reverse is true in a consumer model, and in a collaborative decision-making relationship a balance is established.[29] While the informed consent process need not conform exactly to a theoretical model, insight into the dynamics of decision-making allows clinicians and patients to adjust their communication accordingly.

Because informed consent is an ongoing interactive process, features of the clinical encounter that interfere with physician-patient communication are also barriers to informed consent.[30] Among the typical communication barriers that may complicate physician–patient communication, several are common, particularly for low-literate patients (eg, socioeconomic differences, general physician-patient differences, and cultural differences).[31] Haste and the use of technical jargon by health care providers and investigators are common

TABLE 8-1

Areas of Concern for Low-Literate Adults in Informed Consent

	Elements of Informed Consent*		Possible Concerns
	Patient has decision-making capacity		Literacy may affect assessment of competence
	Process of decision-making†: Paternalism — "You're the doctor; you tell me what to do." Consumerism — "Tell me about this so I can make my decision." Collaboration — "Let's talk about this and decide together."		May have anxiety from collaborative decision-making or choose a Paternalistic Model to not be exposed as low literate
Information Exchange	■ Characteristics of the primary option being offered ■ Risks of primary option ■ Benefits of primary option ■ Degree of uncertainty ■ Alternatives to primary option	Often presented as numerical estimate	May be overly influenced by physician's opinion or have problem with numeracy or number and order of alternatives
	■ What is implied by consent		May have misconception about the legal implications of consent
Consent (Voluntariness)	■ Lack of coercion ■ Confirm authenticity of choice ■ Confirm understanding ■ Confirm preferences ■ Confirm decision		May be reluctant to admit lack of understanding, ask questions, or contradict the physician

*Elements of informed consent collected from theoretical and empirical research as described in text.

pitfalls.[32] Even nontechnical health terms are often interpreted differently by physicians and patients with little education.[33]

In addition, the nature of these barriers depends on the etiology of low functional literacy. Functional literacy is predicated on having adequate vision, reading ability, and cognitive capacity. Stigma may preclude open disclosure of such deficits. Many low-literate adults experience their illiteracy as shameful and use a range of techniques to hide their deficiency.[34,35] Indeed, they are often successful; most physicians are not able to identify which of their patients are functionally illiterate.[36–38] Furthermore, people have varied levels of insight about their functional limitations.

Paternalism as a model for physician-patient communication and decision-making may be an authentic choice embraced by patients for a myriad of reasons. However, if this hierarchical model is chosen because of a stigmatizing condition, the freedom with which this choice is made is limited. Generally, people who do not feel empowered to engage in a collaborative decision-making process with their providers should be supported to do so.

Information Exchange

In Table 8-1, the next element of informed consent involves information exchange. For decisions to be substantively informed, people need to be informed of and understand five content domains: (1) the essential features of the option under consideration, (2) the potential risks, (3) the potential benefits, (4) the degree of uncertainty surrounding these items, and (5) the alternatives. Unfortunately, there is a very large gap between theory and practice. In an evaluation of 540 hospitals' informed consent for procedure forms, only 26.4% of these forms included information on all four central content areas evaluated: nature of procedure, risks, benefits, and alternatives.[39]

Numerous studies have shown that low-literate patients have a lower fund of knowledge about their medical conditions than other patients.[40–43] Because of this, low-literate patients may need to overcome a larger informational hurdle to be in a position to comprehend these content domains. Furthermore, particular aspects of these content domains are commonly presented in a manner that may be problematic for low-literate patients. For example, clinicians often depict risks, benefits, and the degree of certainty in statistical terms. Because numeracy is an integral component of functional literacy, fulfilling these domains of informed consent may be a particular challenge for low-literate adults.

A sixth content domain that is easily overlooked relates to the basic meaning and implication of consent. Consent is an authorization that allows a provider or investigator to do something he or she would not

otherwise do. The significance of this authorization depends on the context, but patients and physicians may understand this aspect of informed consent differently. Patients need to understand that consent constitutes a limited authorization and should not get the impression that signing an informed-consent form involves signing away any rights they otherwise retain.[17] The meaning of the authorization put in motion by informed consent is often unarticulated and not evaluated in research. It is unclear if low-literate patients have more difficulty than other patients do in apprehending this aspect of informed consent.

Consent (Voluntariness)

The final element of informed consent in Table 8-1 deals with the consent itself. For decisions to be substantively voluntary, they must be made without coercion and must represent the patient's actual intentions; confirmation of the authenticity of a patient's decision has been recommended.[44] For example, comparison of the current decision being made to prior decisions can help make a person's values over time more explicit. While there is no requirement that a person's decisions correspond with their stated values, many people want their health care decisions to reflect their values; discordance may reflect a departure from prior ethical commitments or may simply be evidence of misunderstanding. There are subtle ways that the authenticity of a person's decision can be compromised. Cognitive research in decision-making has shown that patients' choices are sensitive to the order and number of alternatives they are given and that people often choose options that run counter to their stated desires.[45] While these findings have not yet been evaluated in patients with low literacy, such patients may have a low sense of their own decisional authority and may have more anxiety than other patients when presented with a set of alternatives.

Furthermore, patients are highly influenced by recommendations presented by physicians.[46] Low-literate patients with a lower sense of self-efficacy may have even more difficulty opposing the authority of the physician than other patients. Consequently, low-literate patients may incorrectly experience a physician's description of the primary option as a Hobson's choice—a situation in which you have to take the one poor option that is offered or get nothing at all.

Note that Table 8-1 represents informed consent as a discrete episode of decision-making and does not capture the complexity that is introduced by an ongoing process with multiple parties over time.[30,47] Time is needed for education and to allow for synthesis of complicated ideas and the application of personal values by the patient. Time is also needed for confirmation of patient understanding, preferences, and decisions by the health care provider. The process is not over when a document is signed. Opportunity for discussion with family

and other confidants may solidify a person's decision or provoke concern and introduce uncertainty. Questions and answers continue to be relevant to reinforce not only the central elements of the decision but also the physician-patient relationship itself. Asking patients to recount key information may be an effective method of confirming understanding.[48] While a form may be used at a particular time to document a decision, ongoing information exchange is crucial. The act of signing may give the impression that discussion has been closed off or irreparably sealed.

While each element of informed consent need not be the focus of explicit deliberation in every clinical decision,[49] observation studies regarding how many of the essential domains of informed consent are actually used in clinical decision-making have shown that norms clearly fail ethical standards.[19,21,50] For example, audiotape analysis of 81 ambulatory care visits revealed that in only 2% of the 262 clinical decisions evaluated was there any assessment of patient understanding.[51]

CHALLENGES TO PRESERVING PATIENT RIGHTS IN FORMALIZED INFORMED CONSENT

The act of signing a consent form is typically reserved for the most formal types of informed consent such as for invasive surgical procedures, the receipt of blood products, or the recruitment of human subjects for research. There are different ways to formally designate that a decision has been made. Physicians and patients can write and sign a contract or review and sign a preprinted contract. These documents may be cosigned, witnessed, copied, and filed.

These activities confirm that a decision was made at a specific time. Unfortunately, such formality has similar attributes to other signed legal agreements and contracts, and people may mistakenly apply attributes of other documents they sign to the informed-consent context. In the recruitment of human subjects, federal regulations preclude the use of exculpatory language in informed consent forms that would "waive or appear to waive any of the subject's rights" to pursue compensation for injury.[52] However, this is exactly the purpose of the liability waivers people sign in nonmedical contexts, and low-literate adults may associate the legal formalities of informed-consent forms with such a waiver of rights.[53]

The Federal Office for Human Research Protection (OHRP) encourages careful wording of informed-consent language to ensure that subjects do not think they give up any rights when they sign consent forms.[54] However, failure appears commonplace. As one researcher noted, "It is doubtful that many subjects understand these studied attempts to protect the institution. The researcher, who supposedly is seeking informed consent, rarely will be able to explain the statement.

Thus, the very essence of the relation of informed consent between subject and investigator is compromised."[55]

If comprehension has been compromised, the consent may be nullified. Medical malpractice case law supports functional illiteracy as a basis for overriding an executed consent form. For example, the Intermediate Court of Appeals of Hawaii, in a 1991 opinion nullifying an informed consent form due to a patient's low literacy, stated that it would "pervert the law of informed consent to allow a physician to discharge his or her affirmative duty by merely securing a signature—even that of a...confused or uneducated patient on an abstruse, jargon-ridden, and largely unintelligible preprinted consent form."[56]

The Food and Drug Administration (FDA) has also established a special obligation to protect the rights of illiterate individuals during the consent process and warns that:

> Clinical investigators should be cautious when enrolling subjects who may not truly understand what they have agreed to do. The Institutional Review Board (IRB) should consider illiterate persons as likely to be vulnerable to coercion and undue influence and should determine that appropriate additional safeguards are in place when enrollment of such persons is anticipated.[57]

Despite these concerns, the FDA offers no guidance for literacy assessment and does not elaborate criteria for assessment of "illiterate individuals." The only hint regarding their definition of illiteracy is from guidance that states: "Illiterate persons who understand English may have the consent read to them and 'make their mark,' if appropriate under applicable state law." This is neither adequate guidance for managing the barriers presented by low literacy nor an adequate operational definition of illiteracy. Most of the 40 to 44 million American adults estimated to be functionally illiterate are able to sign their names without difficulty.[13]

In most circumstances where formalized consent is used, both clinician and patient may benefit from the procedure. Where there is risk of bodily harm, possible dissatisfaction, high cost, abuse, or potential exposure to discrimination for the patient, there is also risk of potential liability for the provider. The overwhelming complexity of many formalized consent materials, however, is evidence that the central focus of formalization is not patient education but an attempt to avoid professional liability.

READABILITY OF INFORMED-CONSENT FORMS

Most adults admit to not reading consent forms.[58] This is not surprising as informed-consent forms are too long, have poor layout, use font sizes that are too small for many readers, and use unexplained medical jargon and legal patois. Despite decades of reports decrying this

problem, informed-consent forms have largely remained complex and are getting longer.[59,60]

Expanding disclosure obligations have exacerbated this trend.[61,62] For example, in an attempt to protect patients' privacy rights, the recent Health Insurance Portability and Accountability Act (HIPAA) has added a considerable amount of additional text to consent forms. The current informed-consent form HIPAA section required for the recruitment of human subjects at Johns Hopkins is comparatively brief at 727 words long, or roughly 2½ pages (www.hopkinsmedicine.org/irb/jhmirb). Because marginal readers read simple text at 80 to 160 words per minute, this one mandatory section will add about 9 minutes of reading time.[63] While it is possible that a low-literate reader could make it through the text, it is unlikely that poor readers will sustain such an effort.

Research on readability analysis was initiated more than 75 years ago to evaluate grade-school textbooks.[64,65] Since then, dozens of techniques have been developed. These analysis systems operate with word lists and evaluate cognitive and organizational complexity, and/or by means of formulas based on text characteristics, such as average word length (semantics) and average sentence length (syntactics). No word-list system has been validated for decades, and shifts in word use may have made these systems obsolete. An automated and combined approach with an updated word list would maximize the benefit of traditional readability measures.

One particular readability formula is the Flesch-Kincaid formula, which is based on average sentence length and the average number of syllables per word. Because the Flesch-Kincaid formula has been incorporated into Microsoft's word readability statistics, it has become broadly available and the most frequently used in the medical literature. Unfortunately, though it has been validated in adults up to a 16th-grade level, the formula is truncated erroneously at a 12th-grade level in Microsoft Word.[63,66–68] Consequently, many studies that have used Flesch-Kincaid analysis likely present falsely low evaluations.

Readability formulas do not ensure that text will be understandable. For example, they are not able to identify authors who know the operating parameters of the formula and purposefully write in short but arcane argot and cant. The formulas serve to highlight problem areas and should never be the only tool used to ensure that text is clearly written.[69–72]

While the appropriate readability level will vary from setting to setting, many institutions set a particular grade-level readability standard and give the impression that documents will be acceptable providing they are written at a lower readability level than the chosen standard.[73] In a survey of US medical schools, grade-level readability

TABLE 8-2

Examples of Informed-Consent Text Provided by Institutional Review Boards at US Medical Schools*

Readability Level	Voluntary Participation	New Information About Risks	No Direct Benefits	Involuntary Removal
Fourth grade†	"You don't have to be in this research study. You can agree to be in the study now and change your mind later. Your decision will not affect your regular care. Your doctor's attitude toward you will not change."	"We may learn about new things that might make you want to stop being in the study. If this happens, you will be informed. You can then decide if you want to continue to be in the study."	"There is no benefit to you from being in the study. Your taking part may help patients in the future."‡	"You may be taken out of the study if: 1. Staying in the study would be harmful. 2. You need treatment not allowed in this study. 3. You fail to follow instructions. 4. You become pregnant. 5. The study is canceled." ‡
Sixth grade†	"Taking part in this study is your choice. If you decide not to take part, this will not harm your relations with your doctors or with the University."	"We may learn new things during the study that you may need to know. We can also learn about things that might make you want to stop participating in the study. If so, you will be notified about any new information."	"You may receive no direct benefit from being in this study. However, your taking part may help patients get better care in the future."	§
Eighth grade†	"Participation in this study is entirely voluntary. You have the right to leave the study at any time. Leaving the study will not result in any penalty or loss of benefits to which you are entitled."	"We will tell you about new information that may affect your willingness to stay in this study."	"There is no direct benefit to you from being in this study. However, your participation may help others in the future as a result of knowledge gained from the research."	"The study doctors have the right to end your participation in this study for any of the following reasons: 1. It would be dangerous for you to continue. 2. You do not follow study procedures as directed by the study doctors. 3. The sponsor decides to end the study."
Tenth grade¶	"Your participation in this study is voluntary and you are free to withdraw at any time. Participation or withdrawal will not affect any rights to which you are entitled."	"We will tell you about new information that may affect your health welfare or willingness to stay in this study."	"There is no guarantee that you will receive direct benefit from your participation in this study."	"The study doctor, per the sponsor, may stop my participation in this study without my consent."

Continued

TABLE 8-2 (Continued)

Examples of Informed-Consent Text Provided by Institutional Review Boards at US Medical Schools*

Readability Level	Voluntary Participation	New Information About Risks	No Direct Benefits	Involuntary Removal
Twelfth grade¶	"Your participation in this study is strictly voluntary. You have the right to choose not to participate or to withdraw your participation at any point in this study without prejudice to your future health care or other services to which you are otherwise entitled."	"You will be promptly notified if any new information develops during the conduct of this research study, which may cause you to change your mind about continuing to participate. If new information becomes known that will affect you or might change your decision to be in this study, you will be informed by the investigator."	"There may be no direct benefit to me; however, information from this study may benefit other patients with similar medical problems in the future."	"You may be terminated from this study without your consent if you have serious side effects, you fail to follow your doctor's instructions, your disease gets worse, or the sponsor closes the study. If this should happen, your doctor can discuss other available treatment options with you."
College¶	"You voluntarily consent to participate in this research investigation. You may refuse to participate in this investigation or withdraw your consent and discontinue participation in this study without penalty and without affecting your future care or your ability to receive alternative medical treatment at the University."	"During the course of the study, you will be informed of any significant new findings (either good or bad), such as changes in the risks or benefits resulting from participation in the research or new alternatives to participation that might cause you to change your mind about continuing in the study. If new information is provided to you, your consent to continue participating in this study will be re-obtained."	"The research physician treats all subjects under a specific protocol to obtain generalizable knowledge and on the premise that you may or may not benefit from your participation in the study."	"Your participation in this research project may be terminated by your doctor without your consent if you are not benefiting from the treatment/procedure, or if the treatment/procedure is determined to be inappropriate to your case. You may also be terminated from participation at any time, at the study physician's discretion, for any reason he/she deems appropriate."

*All the examples are taken directly from medical-school Web sites, unless otherwise noted.

† The readability level is based on the Flesch-Kincaid readability scale.

‡ The passage was modified to present key concepts at a fourth-grade reading level.

§ No passage was found at this reading level.

¶ The readability level is based on the Fry readability formula.

Source: Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. *N Engl J Med.* 2003;348:721–726.

standards for informed consent documents ranged from a fifth-grade readability level to a 10th-grade level.[73] The most commonly cited grade level for a readability standard in this study was an eighth-grade level. However, this standard is neither based on the epidemiology of local rates of low literacy nor actually fulfilled in the template text presented by these same institutions.

Indeed, only 8% of US medical schools met their own target standards in the template language they promulgate. The template language is produced by the institution and is typically used as a boilerplate to build study-specific informed consent documents. See Table 8-2 for template text excerpts from US medical schools written at different grade levels. The mean observed Flesch-Kincaid readability level of the template text in this study was 10.6-2.8 ($P < .001$) grade levels higher than stated standards. The magnitude of this disparity was amplified by application of the Fry readability formula, which is based on the number of sentences and syllables in three 100-word passages. In a representative subsample of 24 texts, modal readability was 13th grade (range, sixth to 16th) and the mean was 13.0.[73] These examples show that institutional review boards charged with safeguarding people with limited literacy[51,52] may ironically play an inadvertent role in promulgating unreadable forms.

CONSENT AS A PROTECTION OF HUMAN SUBJECTS

The highest standards for consent are applied to the recruitment of human subjects for medical research. The OHRP has identified readability of informed-consent forms as a priority area in its review process. Informed consent for medical research is particularly sensitive because comprehension beyond the requisite of usual care is required and there are historic concerns about exploitation of vulnerable populations.[74]

A great deal has changed over the last 50 years in human subjects research. In a landmark study in 1966, Beecher reviewed the ethics of 50 studies and found that consent was mentioned in only 2 studies. He wrote: "If suitably approached, patients will accede, on the basis of trust, to about any request their physician may make."[75] Beecher supported mandating informed consent for research but also felt that such consent would be a safeguard for patient protection only in the hands of responsible investigators. Furthermore, he felt that the best informed-consent forms are meaningless if investigators do not engage the process with integrity.

A particular barrier to informed consent in research is the broad misunderstanding of the central concepts of medical research and the pivotal vocabulary commonly used to describe them. For example,

more than three quarters of subjects do not understand the word "randomly."[54,76,77] To achieve systemic change in the readability of informed-consent forms for research, institutional review boards will need to improve the template and sample text offered to investigators. Federal review appears to improve readability of informed-consent text offered to investigators; in fact, US medical schools that have undergone federal compliance oversight were found to have informed-consent form templates with better readability than those that had not undergone such review. In addition, models already exist to aid interested medical schools.[73] Institutions as diverse as the National Cancer Institute and the State University of New York Downstate Medical School have presented informed-consent templates and sample forms that are below an eighth-grade readability level.[78,79] These forms can serve as models for a national informed consent form template to markedly improve much of the current text offered by institutional review boards.

A randomized controlled trial of an easy-to-read informed consent form of 226 patients at 44 institutions revealed that an informed consent form with a decreased readability level was associated with less anxiety, higher satisfaction, and higher accrual rate.[80] Other studies have suggested that improved forms may decrease medical litigation[81] and that lowering readability levels of written materials improves patient satisfaction and comprehension for patients, even if they are fully literate.[82–85]

However, while lowering the readability level of informed-consent documents is crucial, it does not ensure comprehension. Subjects commonly think they will personally have medical benefits from phase I protocols, despite enrolling with consent documents that say otherwise.[86] This type of confusion will not be completely ameliorated by improving the readability of the text. Improving methods for presenting information to potential research subjects is an ongoing challenge and is itself an enduring topic of randomized trials.[87,88] Ultimately, confirmation of comprehension is the safest approach.

CONCLUSION

Informed consent can be a technique for empowering patients and research subjects. It is also a technique for clinicians, investigators, and institutions to use to manage risk and liability. Both agendas will be advanced by improvements in the informed-consent process. The overwhelming complexity of many consent forms gives the impression that medical institutions seek to disempower their clients with obfuscatory confabulations.

With regard to informed-consent forms, it must be remembered that these educational documents are intended to facilitate the

decision-making process.[39] An obvious barrier to fulfilling this purpose is the readability level of these documents. Many institutions have adopted an eighth-grade readability level as the upper limit of understandability, but in reality, this standard is too high. Furthermore, this standard is inconsistently adhered to in the template text presented by many institutions. As demonstrated in Table 8-2, the central concepts of informed consent can be written at a fourth grade level. Reasonable grade-level readability standards may remove the barrier of low literacy currently built into consent forms.

Quality control must be extended beyond the use of readability formulas. Readability formulas, such as the Flesch-Kincaid and Fry formulas, are ubiquitous, easy to use, and easy to fool. Adding a few periods and acronyms will improve a readability statistic more than actually advancing clarity or general understandability. Materials such as informed-consent documents should be developed and tested with target populations.[89–91]

Beyond shorter words and sentences, other techniques may help improve the comprehension of informed-consent materials. These techniques include modifying features such as font, layout, length, and conceptual complexity as well as using multimedia presentations, including video- and audiotape, and interactive computer programs.[7,71,92–103] There may also be a benefit to supplementing consent documents with systems, known as decision aids, that contextualize abstract risk concepts and compare the consequences of various options. These may take many forms, ranging from personalized risk profiles derived from questionnaires or computer programs that deliver tailored messages to interactive CD-ROM or Web-based instruction. Decision aids have been found to have a range of benefits in the communication of risk information, although they have not been specifically evaluated in low-literate populations. In particular, decision aids have been shown to improve knowledge, improve accuracy of risk/benefit assessment, promote shared decision-making, and improve concordance between patients' stated values and the decision to pursue cancer screening.[104–107] Future research may indicate that this approach to risk communication is useful in the informed consent process with low-literate patients.

The consent document and any ancillary materials, however, are simply tools of the communication process that must be present to ensure high-quality informed consent. The larger context of communication involves examining the assumptions and expectations that inform not only the specific decision that is being considered but also as the nature, purpose, and meaning of consent itself. Low-literate patients may be less likely to ask physicians questions to confirm their understanding. Research may confirm that a communication process in which patient questions are solicited and addressed is more likely to result in accurate patient understanding.

Unfortunately, the present culture of medical care allows routine abrogation of patient rights through a desultory and cursory approach to informed consent. Trainees are often sent to obtain a patient's consent without adequate training in the informed-consent process and without adequate knowledge of the procedure being offered.[108–113] Clinicians often view the process as "consenting a patient" rather than of helping a patient decide whether or not to pursue a particular option.[114] Part of this problem stems from the amount of time it may take to adequately perform the informed-consent process. It is possible that clinicians may be encouraged to invest the required amount of time if the consent process becomes a billable item, as has been done in Japan.[115]

A cultural shift toward a patient-centered model could significantly advance the opportunity for high-quality informed consent.[116,117] This may especially be true for low-literate patients who often know less about their medical conditions and may be less likely to question their health care providers. What are the patient's objectives for their care? What are the potential research subject's goals? If health professionals elicit responses to these core questions, they will be oriented to the patient's perspectives and level of understanding and have the opportunity to begin a dialogue that can lead to a successful informed-consent process.

Physicians and medical institutions have a responsibility and incentive to optimize the consent process. When informed-consent documents are at readability levels that exceed the reader's capacity, physicians and medical institutions fail a significant ethical precept and expose themselves to legal risk.[15,16] Beyond legal and ethical obligations, there is also a pragmatic benefit from having a fully informed constituency. Nonetheless, optimization of the informed-consent process should be driven primarily by the purpose of informed consent—namely, the education of patients—rather than fear of liability. After all, the transformation of the largely incomprehensible informed-consent process to clear, plain, and direct communication that conveys honesty and understanding is nothing less than what the "energetic and practical" people of our country deserve.

REFERENCES

1. Mencken HL. *The American Language: An Inquiry into the Development of English in the United States*. New York, NY: A.A. Knopf; 1921.
2. Fitzmaurice DA, Adams JL. A systematic review of patient information leaflets for hypertension. *J Hum Hypertens*. 2000;14:259–262.
3. Andrus MR, Roth MT. Health literacy: a review. *Pharmacother*. 2002;22:282–302.

4. Mohrmann CC, Coleman EA, Coon SK, et al. An analysis of printed breast cancer information for African American women. *J Cancer Educ.* 2000;15:23–27.
5. Gribble JN. Informed consent documents for BRCA1 and BRCA2 screening: how large is the readability gap? *Patient Educ Couns.* 1999; 38:175–183.
6. Grundner TM. On the readability of surgical consent forms. *N Engl J Med.* 1980;302:900–902.
7. Heinze-Lacey B, Saunders C, Sugar A. Improving the readability of informed consent documents. *IRB.* 1993;15:10–11.
8. Hearth-Holmes M, Murphy PW, Davis TC, et al. Literacy in patients with chronic disease: Systemic lupus erythematosus and the reading level of patient education materials. *J Rheumatol.* 1997;24:2335–2339.
9. Hopper KD, TenHave TR, Tully DA, Hall TE. The readability of currently used surgical/procedure consent forms in the United States. *Surgery.* 1998;123:496–503.
10. Meade CD, Byrd JC. Patient literacy and the readability of smoking education literature. *Am J Public Health.* 1989;79:204–206.
11. Morrow GR. How readable are subject consent forms? *JAMA.* 1980;244:56–58.
12. Wagner L, Davis S, Handelsman MM. In search of the abominable consent form: the impact of readability and personalization. *J Clin Psychol.* 1998;54:115–120.
13. Kirsch JS, Junegeblut A, Jenkins L, Kolstad A. Adult literacy in America: a first look at the results of the National Adult Literacy Survey (NALS). Washington, DC:US Department of Education; 1993.
14. Mazur D. Influence of the law on risk and informed consent. *BMJ.* 2003;327:731–734.
15. Health Literacy Project and consulting attorneys. *Literacy, Health, and the Law.* Philadelphia, Pa: Health Promotion Council of Southeastern Pennsylvania; 1996.
16. Pape T. Legal and ethical considerations of informed consent. *AORN J.* 1997;65:1122–1127.
17. Faden RR, Beauchamp TL. *A History and Theory of Informed Consent.* New York, NY: Oxford University Press; 1986.
18. Mello MM, Studdert DM, Brennan TA. The rise of litigation in human subjects research. *Ann Intern Med.* 2003;139:40–45.
19. Braddock CH, III, Fihn SD, Levinson W, Jonsen AR, Pearlman RA. How doctors and patients discuss routine clinical decisions: informed decision making in the outpatient setting. *J Gen Intern Med.* 1997;12:339–345.
20. Holmes-Rovner M, Wills CE. Improving informed consent: insights from behavioral decision research. *Med Care.* 2002;40:V30–V38.

21. Braddock CH III, Edwards KA, Hasenberg NM, Laidley TL, Levinson W. Informed decision making in outpatient practice: time to get back to basics. *JAMA*. 1999;282:2313–2320.
22. Jonsen AR. A map of informed consent. *Clin Res*.1975;23:277–279.
23. Nora LM, Benvenuti RJ III. Medicolegal aspects of informed consent. *Neurol Clin*. 1998;16:207–216.
24. Meisel A, Roth LH, Lidz CW. Toward a model of the legal doctrine of informed consent. *Am J Psychiatry*. 1977;134:285–289.
25. Meisel A, Kuczewski M. Legal and ethical myths about informed consent. *Arch Intern Med*. 1996;156:2521–2526.
26. Santacruz KS, Swagerty D. Early diagnosis of dementia. *Am Fam Physician*. 2001;63:703–708.
27. Weiss BD, Reed R, Kligman EW, Abyad A. Literacy and performance on the Mini-Mental State Examination. *J Am Geriatr Soc*. 1995;43:807–810.
28. Baker DW, Gazmararian JA, Sudano J, Patterson M, Parker RM, Williams MV. Health literacy and performance on the Mini-Mental Status Examination. *Aging Ment Health*. 2002;6:22–29.
29. Roter D, Hall JA. *Doctors Talking to Patients / Patients Talking to Doctors: Improving Communication in Medical Visits*. Westport, Conn: Auburn House; 1992.
30. Lidz CW, Meisel A, Osterweis M, Holden JL, Marx JH, Munetz MR. Barriers to informed consent. *Ann Intern Med*. 1983;99:539–543.
31. Quill TE. Recognizing and adjusting to barriers in doctor-patient communication. *Ann Intern Med*. 1989;111:51–57.
32. Bourhis RY, Roth S, MacQueen G. Communication in the hospital setting: a survey of medical and everyday language use amongst patients, nurses and doctors. *Soc Sci Med*. 1989;28:339–346.
33. Hadlow J, Pitts M. The understanding of common health terms by doctors, nurses and patients. *Soc Sci Med*. 1991;32:193–196.
34. Lazare A. Shame and humiliation in the medical encounter. *Arch Intern Med*. 1987;147:1653–1658.
35. Baker DW, Parker RM, Williams MV, et al. The health care experience of patients with low literacy. *Arch Fam Med*. 1996;5:329–334.
36. Bass PF III, Wilson JF, Griffith CH, Barnett DR. Residents' ability to identify patients with poor literacy skills. *Acad Med*. 2002; 77:1039–1041.
37. Lindau ST, Tomori C, Lyons T, Langseth L, Bennett CL, Garcia P. 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.
38. Montalto NJ, Spiegler GE. Functional health literacy in adults in a rural community health center. *WV Med J*. 2001;97:111–114.

39. Bottrell MM, Alpert H, Fischbach RL, Emanuel LL. Hospital informed consent for procedure forms: facilitating quality patient-physician interaction. *Arch Surg*. 2000;135:26–33.

40. 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;25:337–344.

41. 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;18:325–331.

42. Williams MV, Baker DW, Honig EG, Lee TM, Nowlan A. Inadequate literacy is a barrier to asthma knowledge and self-care. *Chest*. 1998;114:1008–1015.

43. Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease: a study of patients with hypertension and diabetes. *Arch Intern Med*. 1998;158(2):166–172.

44. Nelson RM, Merz JF. Voluntariness of consent for research: an empirical and conceptual review. *Med Care*. 2002;40:V69–V80.

45. Ubel PA. Is information always a good thing? Helping patients make "good" decisions. *Med Care*. 2002;40:V39–V44.

46. Gurmankin AD, Baron J, Hershey JC, Ubel PA. The role of physicians' recommendations in medical treatment decisions. *Med Decis Making*. 2002;22:262–271.

47. Lidz CW, Appelbaum PS, Meisel A. Two models of implementing informed consent. *Arch Intern Med*. 1988;148:1385–1389.

48. Pizzi LT, Goldfarb NI, Nash DB. Procedures for obtaining informed consent. In: Shojania KG, Duncan BW, McDonald KM, eds. *Making Health Care Safer: A Critical Analysis of Patient Safety Practices*. Rockville, Md: Agency for Healthcare Research and Quality; 2001:546–554.

49. Sprung CL, Winick BJ. Informed consent in theory and practice: legal and medical perspectives on the informed consent doctrine and a proposed reconceptualization. *Crit Care Med*. 1989;17:1346–1354.

50. Wu WC, Pearlman RA. Consent in medical decision making: the role of communication. *J Gen Intern Med*. 1988;3:9–14.

51. Department of Health and Human Services, Office for Human Research Protection. *Institutional Review Board Guidebook*. Washington, DC: Government Printing Office; 1993.

52. National Archives and Records Administration, Office of the Federal Register. *Code of Federal Regulations*. Protection of Human Subjects. Title 45, 46.109 and 50.20. Washington, DC: Government Printing Office; 1993.

53. Cassileth BR, Zupkis RV, Sutton-Smith K, March V. Informed consent: why are its goals imperfectly realized? *N Engl J Med*. 1980;302:896–900.

54. Waggoner WC, Sherman BB. Who understands? II: A survey of 27 words, phrases, or symbols used in proposed clinical research consent forms. *IRB*. 1996;18:8–10.

55. Curran WJ. Compensation for injured research subjects. Regulation by informed consent. *N Engl J .Med*. 1979;20:648–649.

56. *Keomaka v. Zakaib*, 8 Haw. App. 533, 811P.2d 487, 6-7 (1991).

57. FDA/Office of Science Coordination and Communication. Guidance for Institutional Review Boards and Clinical Investigators. Available at: http://www.fda.gov/oc/ohrt/irbs/informedconsent.html#illiterate. Accessed August 30, 2003.

58. Lavelle-Jones C, Byrne DJ, Rice P, Cuschieri A. Factors affecting quality of informed consent. *BMJ*. 1993;306:885–890.

59. LoVerde ME, Prochazka AV, Byyny RL. Research consent forms: continued unreadability and increasing length. *J Gen Intern Med*. 1989;4:410–412.

60. Sugarman J, McCrory DC, Powell D, et al. Empirical research on informed consent: an annotated bibliography. *Hastings Cent Rep*. 1999;29:S1–S42.

61. LeBlang TR. Informed consent and disclosure in the physician-patient relationship: expanding obligations for physicians in the United States. *Med Law*. 1995;14:429–444.

62. Petrila J. The emerging debate over the shape of informed consent: can the doctrine bear the weight? *Behav Sci Law*. 2003;21:121–133.

63. Gray WB Jr. *How to Measure Readability*. Philadelphia, Pa: Dorrance & Company; 1975.

64. Klare GR. *The Measurement of Readability*. Ames, Iowa: The Iowa State University Press; 1963.

65. Chall JS, Dale E. *Readability Revisited: The New Dale-Chall Readability Formula*. Cambridge, Mass: Brookline Books;1995.

66. McLaughlin GH. SMOG Grading—New Readability Formula. *J Reading*. 1969;12:639–646.

67. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy Enlisted Personnel. Research Branch Report. Chief of Naval Technical Training: Naval Air Station Memphis. 8–75. 1975.

68. Kincaid J, Gamble L. Ease of comprehension of standard and readable automobile insurance policies as a function of reading ability. *J Reading Behav*. 1977;9:85–87.

69. Hochhauser M. Some overlooked aspects of consent form readability. *IRB*. 1997;19:5–9.

70. Meade CD, Wittbrot R. Computerized readability analysis of written materials. *Comput Nurs*. 1988;6:30–36.

71. Peterson BT, Clancy SJ, Champion K, McLarty JW. Improving readability of consent forms: what the computers may not tell you. *IRB*. 1992;14:6–8.
72. Philipson SJ, Doyle MA, Gabram SG, Nightingale C, Philipson EH. Informed consent for research: a study to evaluate readability and processability to effect change. *J Investig Med*. 1995;43:459–467.
73. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. *N Engl J Med*. 2003;348:721–726.
74. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. 4–18-1979.
75. Beecher HK. Ethics and clinical research. *N Engl J Med*.1966;274: 1354–1360.
76. Lawson SL, Adamson HM. Informed consent readability: subject understanding of 15 common consent form phrases. *IRB*. 1995;17:16–19.
77. Waggoner WC, Mayo DM. Who understands? A survey of 25 words or phrases commonly used in proposed clinical research consent forms. *IRB*. 1995;17:6–9.
78. Comprehensive Working Group on Informed Consent in Cancer Clinical Trials. Simplification of Informed Consent Documents. Available at www.cancer.gov/clinical_trials. Accessed August 30, 2003.
79. Padberg RM, Flach J. National efforts to improve the informed consent process. *Semin Oncol Nurs*. 1999;15:138–144.
80. Coyne CA, Xu R, Raich P, et al. Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: a study of the Eastern Cooperative Oncology Group. *J Clin Oncol*. 2003;21:836–842.
81. Faden RR, Becker C, Lewis C, Freeman J, Faden AI. Disclosure of information to patients in medical care. *Med Care*. 1981;19:718–733.
82. Meade CD, Byrd JC, Lee M. Improving patient comprehension of literature on smoking. *Am J Public Health*. 1989;79:1411–1412.
83. Davis TC, Holcombe RF, Berkel HJ, Pramanik S, Divers SG. Informed consent for clinical trials: a comparative study of standard versus simplified forms. *J Natl Cancer Inst*. 1998;90(9):668–674.
84. Jolly BT, Scott JL, Sanford SM. Simplification of emergency department discharge instructions improves patient comprehension. *Ann Emerg Med*. 1995;26:443–446.
85. Young DR, Hooker DT, Freeberg FE. Informed consent documents: increasing comprehension by reducing reading level. *IRB*. 1990;12:1–5.
86. Lidz CW, Appelbaum PS. The therapeutic misconception: problems and solutions. *Med Care*. 2002;40:V55–V63.

87. Lavori PW, Sugarman J, Hays MT, Feussner JR. Improving informed consent in clinical trials: a duty to experiment. *Control Clin Trials*. 1999;20:187–193.
88. Peduzzi P, Guarino P, Donta ST, Engel CC Jr, Clauw DJ, Feussner JR. Research on informed consent: investigator-developed versus focus group-developed consent documents, a VA cooperative study. *Control Clin Trials*. 2002;23:184–197.
89. Taub HA, Baker MT, Sturr JF. Informed consent for research. Effects of readability, patient age, and education. *J Am Geriatr Soc*. 1986;34:601–606.
90. Taylor HA. Barriers to informed consent. *Semin Oncol Nurs*. 1999;15:89–95.
91. Doak CC, Doak LG, Root JH. *Teaching Patients with Low-Literacy Skills*, 2nd ed. Philadelphia, Pa: JB Lippincott; 1996.
92. Berto D, Peroni M, Milleri S, Spagnolo AG. Evaluation of the readability of information sheets for healthy volunteers in phase-I trials. *Eur J Clin Pharmacol*. 2000;56:371–374.
93. Jimison HB, Sher PP, Appleyard R, LeVernois Y. The use of multimedia in the informed consent process. *J Am Med Inform Assoc*. 1998;5:245–256.
94. Bjorn E, Rossel P, Holm S. Can the written information to research subjects be improved?—an empirical study. *J Med Ethics*. 1999;25:263–267.
95. Doak CC, Doak LG, Friedell GH, Meade CD. Improving comprehension for cancer patients with low literacy skills: strategies for clinicians. *CA Cancer J Clin*. 1998;48:151–162.
96. Loughrey L. Improving readability of hospital forms. *J Nurs Staff Dev*. 1986;2:37–38.
97. Meade CD, Howser DM. Consent forms: how to determine and improve their readability. *Oncol Nurs Forum*. 1992;19:1523–1528.
98. Meade CD. Producing videotapes for cancer education: methods and examples. *Oncol Nurs Forum*. 1996;23:837–846.
99. Philipson SJ, Doyle MA, Nightingale C, Bow L, Mather J, Philipson EH. Effectiveness of a writing improvement intervention program on the readability of the research informed consent document. *J Investig Med*. 1999;47:468–476.
100. Barbour GL, Blumenkrantz MJ. Videotape aids informed consent decision. *JAMA*. 1978;240:2741–2742.
101. Benitez O, Devaux D, Dausset J. Audiovisual documentation of oral consent: a new method of informed consent for illiterate populations. *Lancet*. 2002;359:1406–1407.
102. Dunn LB, Lindamer LA, Palmer BW, Schneiderman LJ, Jeste DV. Enhancing comprehension of consent for research in older patients with psychosis: a randomized study of a novel consent procedure. *Am J Psychiatry*. 2001;158:1911–1913.

103. Weston J, Hannah M, Downes J. Evaluating the benefits of a patient information video during the informed consent process. *Patient Educ Couns*. 1997;30:239–245.
104. Paling J. Strategies to help patients understand risks. *BMJ*. 2003;327:745–748.
105. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ*. 2003;327:741–744.
106. O'Connor AM, Stacey D, Entwistle V, et al. Decision aids for people facing health treatment or screening decisions (Cochrane Review). In: The Cochrane Library, Issue 1. Chichester, UK: John Wiley & Sons, Ltd; 2004.
107. Holmes-Rovner M, Valade D, Orlowski C, Draus C, Nabozny-Valerio B, Keiser S. Implementing shared decision-making in routine practice: barriers and opportunities. *Health Expect*. 2000;3:182–191.
108. Yahne CE, Edwards WS. Learning and teaching the process of informed consent. *Proc Annu Conf Res Med Educ*. 1986;25:30–35.
109. Cohen DL, McCullough LB, Kessel RW, Apostolides AY, Heiderich KJ, Alden ER. A national survey concerning the ethical aspects of informed consent and role of medical students. *J Med Educ*. 1988;63:821–829.
110. Johnson SM, Kurtz ME, Tomlinson T, Fleck L. Teaching the process of obtaining informed consent to medical students. *Acad Med*. 1992;67:598–600.
111. Huntley JS, Shields DA, Stallworthy NK. Consent obtained by the junior house officer—is it informed? *JR Soc Med*. 1998;91:528–530.
112. Wear S. Enhancing clinician provision of informed consent and counseling: some pedagogical strategies. *J Med Philos*. 1999;24:34–42.
113. Srinivasan J. Observing communication skills for informed consent: an examiner's experience. *Ann R Coll Physicians Surg Can*. 1999;32:437–440.
114. Hope T. Don't 'consent' patients, help them to decide. *Health Care Anal*. 1996;4:73–76.
115. Akabayashi A, Fetters MD. Paying for informed consent. *J Med Ethics*. 2000;26:212–214.
116. Bridson J, Hammond C, Leach A, Chester MR. Making consent patient centred. *BMJ*. 2003;327:1159–1161.
117. Sullivan M. The new subjective medicine: taking the patient's point of view on health care and health. *Soc Sci Med*. 2003;56:1595–1604.

The American Journal of Medicine (2005) 118, 175-180

THE AMERICAN JOURNAL *of* MEDICINE®

CLINICAL RESEARCH STUDY

Assessment of medical school institutional review board policies regarding compensation of subjects for research-related injury

Michael K. Paasche-Orlow, MD, MPH,[a] Frederick L. Brancati, MD, MHS[b]

[a]*From the Department of Medicine, Boston University School of Medicine, Boston, Massachusetts; and the* [b]*Departments of Medicine and Epidemiology, The Johns Hopkins University, Baltimore, Maryland.*

KEYWORDS:
Institutional review boards;
Injury;
Compensation;
Informed consent

PURPOSE: Although the Institute of Medicine (IOM) calls for research organizations to compensate subjects for research-related injury, administrators may fear that candid presentation of such policies would create financial risk. We hypothesized that informed consent language at U.S. medical schools would be particularly complex and fall short of IOM goals, especially for projects without industry sponsorship.

METHODS: Medical school websites (N = 123) were surveyed for informed consent language for research-related injury. Text was extracted from 113 sites (92%) and evaluated for details regarding financial liability for research-related injury. When sufficient template text was available (n = 106), the readability of liability policies was compared with the readability of other standardized passages using Flesch-Kincaid analysis.

RESULTS: Coverage for medical bills is offered at 61% (23/38) of schools when there is an industry sponsor as compared with 22% (22/102) when there is none ($P < 0.001$). When coverage is offered in studies with no industry sponsor, it is limited to emergency bills in half (11/22) of these policies. Seventy-two percent (81/113) of medical school consent forms specifically rule out the possibility of monetary compensation. The mean (± SD) reading grade level of liability text in consent forms is higher than that in other template paragraphs (11.5 ± 1.4 vs. 10.6 ± 1.4; $P = 0.0001$).

CONCLUSION: Federally funded research at most U.S. medical schools is conducted with consent form language that is particularly complex and that fails to protect subjects from the financial burden of research-related injury. Few schools meet IOM standards.

Despite repeated calls from national ethics commissions and the Institute of Medicine (IOM) for a federal policy to compensate subjects for research-related injury, no national policy has been established.[1–7] Local institutions have instead developed their own policies for managing research-related injuries. Indeed, the IOM advocates for research institutions to provide at least the cost of medical care and rehabilitation, without regard to fault, until federal protection is initiated. However, from a risk management perspective, candid presentation of such a policy might create financial risk. Consequently, lawyers seeking to mitigate financial risk may tend to draft complex informed consent language that appears to

Dr. Brancati was supported by a Mid-Career Investigator Award in Patient Oriented Research (K24DK62222).

Requests for reprints should be addressed to Michael K. Paasche-Orlow, MD, MPH, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, 91 East Concord Street, Suite 200, Boston, Massachusetts 02118.

E-mail address: michael.paascheorlow@bmc.org.

0002-9343/$ -see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.amjmed.2004.03.045

exclude the possibility of compensation and limits institutional liability.[8]

We therefore undertook a survey of U.S. medical schools to test the hypothesis that informed consent language about research-related injury would be particularly complex, and that the policies depicted in such texts would fail to protect injured subjects from medical expenses, would not offer other forms of monetary compensation for injury (e.g., compensation for job loss, disability, pain, and suffering), and would obscure the rights of subjects to pursue compensation. We further hypothesized that schools would offer better coverage for industry-sponsored research than for research sponsored by the National Institutes of Health (NIH), as industry provides an alternative source of funding, and that schools with more research activity and those with a public mandate might be more likely to offer coverage.

Methods

Data sources

All data were obtained from publicly available websites between 2001 and 2002. The websites of medical school institutional review boards (N = 123) were surveyed for any suggested text for research-related injuries in informed consent documentation. Of the 123 medical schools in the 50 states and the District of Columbia, we were unable to identify policies for research-related injuries in 10 for the following reasons: website restricted to intranet access (n = 6), website without mention of injury or compensation (n = 1), and no indication of existence of an institutional review board website (n = 3).

The NIH website yielded the rank order of medical school funding in 2002.[9] This was used as an indicator of research activity at the medical school.

Text abstraction

Each medical school's institutional review board website was examined for suggested text for informed consent documentation. Applicable templates, sample documents, and sample language were downloaded into Microsoft Word 2000 (Microsoft, Redmond, Washington). These materials were surveyed for any text related to injury and compensation for adult human research subjects with full decisional capacity.

Payment for medical care

Abstracted text regarding research-related injury was evaluated for indications of financial responsibility for the cost of medical care. Any distinctions between the cost of emergency care and other medical expenses, stipulations of financial liability, or specified billing procedures were noted.

Financial compensation for injury

Abstracted text regarding research-related injury was evaluated for any description of a system or policy for subjects to receive financial compensation of any kind (e.g., lost wages, disability, pain, and suffering).

Readability assessment

Readability was measured with a component of the automated readability statistics offered in Microsoft Word: the Flesch-Kincaid readability scale (range, 0 to 12th grade). The readability of the liability policy statement was compared with the average readability of three other template text paragraphs.[10] The Flesch-Kincaid system is a valid measure that assesses readability based on the average number of syllables per word and the average number of words per sentence.[11,12] There was sufficient template text from the informed consent form in 94% (n = 106) to allow analysis of readability.

Statistical analysis

Rates of compensation according to funding source, public versus private status, and tertiles of NIH funding rank were compared using chi-squared tests. Comparison of mean observed readability scores of compensation text with that of other template text was conducted with the Wilcoxon signed-rank test. All significance tests were two-tailed. Analyses were conducted with Stata, version 8 (College Station, Texas). This research was classified as exempt by the Johns Hopkins Institutional Review Board.

Results

Informed consent language for research-related injury was identified in 92% (n = 113) of medical school websites. Indemnification arrangements are negotiated between private sponsors and the office for research administration at the medical schools on a case-by-case basis. As such, many schools do not present template language for sponsored studies. Of the 113 medical schools, 90% (n = 102) provide template language for federally sponsored studies and 34% (n = 38) provide template language for industry-sponsored studies. Separate texts for federal- and industry-sponsored research are provided by 24% (n = 27) of the schools and 10% (n = 11) of the schools exclusively provide text for industry-sponsored studies.

Payment for medical care

Of 38 schools that present language templates for industry-sponsored studies, 23 (61%) declare that industry sponsors will pay the medical expenses of research-related injury. An optional format is presented in 34% (n = 13) to accommodate various policies, and 8% (n = 3) appear to presume that sponsors will not be offering free medical care.

Table 1 Characteristics of 22 programs with no industry sponsor that offer subjects financial protection for research-related injuries

Policy limits*	Typical language	Number of schools
Emergency coverage only	" . . . will pay the cost of emergency first aid for any injury that may happen as a result of your being in this study."	8
Reasonable and necessary	" . . . will reimburse you for the reasonable costs of medically necessary treatment."	4†
Emergency and short-term coverage	"Only immediate, essential, short-term medical treatment as determined by the participating hospital, will be available for the injury without charge to you personally."	3
Only medical coverage	"Medical treatment for physical injury directly resulting from the research procedure."	3
Unspecified	"You will not have to pay any charges resulting from the harmful effect or injury."	3
Nontherapeutic	" . . . injured as a direct result of research procedures not done primarily for your own benefit, you will receive treatment at no cost."	1

*Six of these policies relate solely to charges that remain subsequent to attempts to recover funds from subject's insurance policies. Policies listed in descending order of frequency.

†One of these excludes studies involving critically ill patients.

Only 13% (3/23) of the industry-sponsored policies that offer some form of coverage are explicitly limited to reimbursement of costs related to emergency care.

In contrast, of the 102 schools with NIH text, only 22 (22%) offer some form of financial support, 58 (57%) specify that no financial support is available for the medical bills from research-related injuries, and in the remaining 22 (22%) financial responsibility is unclear. Most policies that offer financial protection are clearly limited, with half (11/22) offering only the cost of emergency medical care (Table 1).

Overall, informed consent language in industry-sponsored research is significantly more likely than in NIH-sponsored research to offer free medical care for research-related injury (Table 2). This difference persists in analyses stratified by level of research activity and medical school type. Higher level of research intensiveness is associated significantly with language that is more likely to offer free care for research-related injuries. In contrast, there appears to be no difference in language between public and private medical schools.

At the 27 schools that present separate texts for industry and non–industry-sponsored liability language, there are 11 concordant policies, seven of which representing instances where both texts offer some payment for medical bills. In each of the 16 sets of discordant policies, financial support is offered in the industry-sponsored policy and not offered in the absence of an industry sponsor ($P < 0.005$).

Financial compensation for injury

No school offered to provide financial compensation for research-related injury. Notably, 72% (81/113) of informed consent forms specifically rule out financial compensation, using a variety of wording (Table 3). No mention of compensation for research-related injuries is made in 18% (n = 20) of schools. In 11% (n = 12), the language is not clear and sometimes leaves open the possibility that some subjects could receive monetary compensation in certain unspecified circumstances. For example, four schools indicate,

Table 2 Percentage of U.S. medical school policies offering payment of medical bills for injured subjects according to selected criteria

	n/N (%)	
	Non-industry-sponsored	Industry-sponsored
Overall sample	22/102 (22)	23/38 (61)*
Tertile rank of NIH funding		
Highest	12/33 (36)	14/16 (88)
Middle	6/37 (16)	4/11 (36)
Lowest	4/32 (13)†	5/11 (45)‡
Public medical school	11/63 (17)	13/21 (62)
Private medical school	11/39 (28)	10/17 (59)

NIH = National Institutes of Health.

*NIH-sponsored vs. industry-sponsored ($P < 0.001$).

†Highest, middle, lowest ($P = 0.02$).

‡Highest, middle, lowest ($P = 0.01$).

Table 3 Language used by U.S. medical schools in informed consent forms to avoid liability for research-related injury

Form of statement*	Typical language	Number of schools
Unavailable	"Compensation for an injury resulting from your participation in this research is not available."	21
Not provided	" . . . does not provide free medical treatment or other forms of compensation to persons injured as a result of participating in research . . . "	18
Contrary to policy	"It is not the policy of _____ to compensate . . . "	11
Frank denial	"____will not pay you for pain, worry, lost income, or nonmedical care costs that might occur . . . "	8
Not offered	"No monetary compensation will be offered."	5
Not budgeted	" . . . has not set aside funds for compensation . . . "	5
Absence of program	"There is no program in place for other monetary compensation."	5
Impermissible	" . . . financial compensation cannot be provided"	4
Lack of responsibility	" . . . assumes no obligation to pay any money . . . "	4

*Listed in descending order of frequency.

without explanation, that the decision regarding compensation is a matter of discretion.

Readability

The mean (± SD) readability of text regarding injury and compensation in informed consent forms was higher (11.5 ± 1.4) than the mean reading grade level of other template paragraphs (10.6 ± 1.4). Of 106 injury and compensation texts evaluated for readability, five (5%) were found to be below an eighth-grade reading level and 80 of these (75%) were found to be at least at a 12th-grade reading level, which is the upper limit of the automated Microsoft Word Flesch-Kincaid scale. Examples of policy language at various reading levels are listed in Table 4.

Discussion

Our findings demonstrate that the language employed in U.S. medical school informed consent forms depicts policies that generally fail to protect subjects from the financial

Table 4 Examples of text regarding research-related injury in informed consent forms from U.S. medical school institutional review board boilerplates

Reading level*	Text excerpt†
5th	If you have been injured, let us know right away. We will direct you to medical care. However, you and your insurance company will be billed for this treatment.‡
7th	If you are harmed by being in this study, we will provide or arrange care as needed. We do not have funds to pay for such care, however. Payment for such care will be your responsibility, or that of your health insurance company (or Medicaid, etc).
9th	We will give you emergency care if you are injured by this research. However, _____ University has not set aside funds to pay for this care or to compensate you if a mishap occurs.
11th	In the event of injury resulting from your participation in this study, treatment can be obtained at _____ College Hospital. You should understand that the costs of such treatment will be your responsibility. Financial compensation is not available.
>16th§	If physical injury resulting from participation in this research should occur, I understand that although compensation is not available, medical treatment will be available, including first aid, emergency treatment, and follow-up care as needed, and that my insurance carrier may be billed for the cost of such treatment. I further understand that in making such medical treatment available, or providing it, the persons conducting this research project are not admitting that my injury was their fault.

*Readability levels based on Flesch-Kincaid readability scale.

†Direct quotations from medical school websites with identifiers obscured except where noted.

‡No text found at this reading level; text prepared by authors.

§This passage measured manually with the Flesch-Kincaid formula: (.39)ASL + 11.8(ASW) − 15.59, where ASL = average sentence length (number of words/number of sentences) and ASW = average syllables per word (number of syllables/number of words). The formula has been validated in adults up to a 16th-grade level.[12,13]

burden of research-related injury and that fall short of IOM-touted ideals. Subjects at approximately three quarters of U.S. medical schools sign forms that reject institutional responsibility for the direct or indirect costs of research-related injury. In programs where subjects are offered some form of financial relief, this is typically limited to the bills associated with emergency care. Further, the language used to describe these policies is particularly complex and so is less likely to be understood by many potential subjects.

When medical schools host studies paid for by industrial sponsors, they generally ensure indemnification arrangements are in place. However, when they host research paid for by the NIH, most do not maintain this standard. In many institutions, this creates a double standard whereby the terms of liability shift from study to study depending on the funding source.

The strengths of our study that lend weight to these conclusions are the nearly complete sampling of U.S. medical schools and the use of standardized instruments for data abstraction. Nonetheless, several limitations should be kept in mind. First, data were obtained exclusively from websites. Although it is likely that the materials presented on medical school websites accurately reflect the materials used locally, additional documents were not examined. For example, only a third of the schools presented template text for privately sponsored studies. It is plausible that medical schools not presenting such materials might systematically endorse policies that are less protective; however, there is no particular reason to link the choice to post materials on a medical school's website with the details of compensation policy. Second, our analysis was limited to templates and samples rather than actual consent forms. Although we did not directly evaluate the possibility that investigators might introduce changes in the liability language, our experience suggests that such action is rare. Third, and most importantly, our analysis was limited to informed consent language and did not evaluate the actual practice pattern of investigators, administrators, and risk-managers. There may be a gap between the language and actual practice. Although schools may have unstated policies, an a priori rejection of financial responsibility paired with a subsequent plan to pay for research-related injuries is problematic. This type of arrangement uses the informed consent document as a device to dissuade injured subjects from seeking financial support.

There are several ways for informed consent documents to prevent people from achieving their rights as human subjects. Two of these barriers, which are specifically prohibited by federal statute, were commonly noted in our study. The first is the presence of overly complicated language.[14] Forty-eight percent of adults in the United States read at or below an eighth-grade reading level.[15] Since 1970, experts have recognized this concern and called for liability policies to be "prominently placed in clear, readable terms at the top or front of consent forms or be presented in a separately signed document".[16] As the mean reading level of informed consent language pertaining to liability examined in our study was higher than the 11th-grade level, it is clearly a prevalent and persistent problem.

The second barrier is the presence of exculpatory language, which would "waive or appear to waive any of the subject's rights" to pursue compensation for injury.[14] The Federal Office for Human Research Protection (OHRP) encourages careful wording of informed consent language for research-related injury and offers acceptable text, including phrases such as "not able to offer financial compensation" or "makes no commitment to provide free" medical care.[17] Much of the language we found appears more exculpatory than the OHRP examples, but even what is deemed acceptable by OHRP may give many subjects the mistaken impression that they have no recourse to pursue compensation beyond the institution's voluntarily chosen limits.[8]

Legal representatives for medical schools must be concerned about financial liability. Possibly, they feel that they do not want to encourage the participation of a litigious cohort of subjects with informed consent form language that might serve as a roadmap to compensation.[8] After the fact, risk-managers determine which injured subjects will be compensated *ex gratia* and which will have to sue. However, as noted by the Tuskegee Syphilis Study Ad Hoc Advisory Panel, "remitting injured subjects to the uncertainties of the law court is not a solution".[5] This arrangement creates hurdles to compensation and promotes an adversarial relationship between the injured subject and the institution.[19]

The whole discussion begs the question: what is a research-related injury? Institutions should be concerned by the vagaries of assigning causation in many circumstances. Any adverse event must be reported to the institutional review board; however, only a fraction of adverse events will be due to the research and only a fraction of these will be due to negligence. In a trial of cholesterol-lowering interventions, every myocardial infarction is an adverse event, but few, if any, may constitute a research-related injury. Further, data to characterize the scope of this problem are inadequate.

Research institutions carry insurance to protect themselves from claims of negligence; yet, injuries that result from research, but are not due to negligence, are typically not covered by institutional policies.[20] In addition, when policies are based on proof of negligence, investigators must manage contrasting obligations. Will they advance the interests of their subject or of the institution?

To avoid the concern of dual loyalties, decrease administrative obstacles, decrease cost, and fulfill the moral duty of compensating injured subjects, the IOM and others have advocated for a no-fault compensation system.[4–6,20–23] Indeed, the Veterans Administration and the Department of Defense have instituted formal programs to pay for the medical care of injured subjects in research they fund.[24,25] Similarly, policies to provide compensation for injured re-

search subjects are in place at the Council for International Organizations of Medical Sciences, the European Union, and other countries.[4,26] Such programs indicate a broad endorsement of the importance of compensating injured subjects. However, such programs cannot serve as direct models for U.S. policy as organizations such as the Veterans Administration are both the payer and the provider.

Several medical schools that offer some form of payment for medical services differentiate between therapeutic and nontherapeutic research. This distinction is hard to defend. The fact that a subject may potentially derive some benefit from a therapeutic protocol does not add force to the claim that subjects who are also patients should not be compensated for injury related to the research. Although the natural progression of a disease and a true research-related illness may sometimes be difficult to discern, this should not provide a categorical rejection of responsibility.[1,19]

We all benefit from medical research. People who become research subjects are exposed to the small but real chance of being injured. Our study provides evidence that most publicly funded research is conducted without a safety net to protect subjects from the financial burden of injuries. Research institutions should help shoulder the burden of risk for subjects in NIH-sponsored research to provide at least the cost of medical care and rehabilitation, without regard to fault, until federal protection is initiated. Such a policy will affirm our social contract with these invaluable volunteers, simplify the relationship between subjects and the physician-researchers who recruit them, and establish candor as the cornerstone of consent.[27,28]

Acknowledgment

The authors thank Dr. Paul Lietman, Johns Hopkins University School of Medicine, Baltimore, Maryland, for encouragement and review of the manuscript.

References

1. Childress JF. Compensating injured research subjects: I. The moral argument. *Hastings Cent Rep.* 1976;6:21–27.
2. President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, 42 USC (1982) §300v (1978).
3. Faden RR, Lederer SE, Moreno JD. US medical researchers, the Nuremberg Doctors Trial, and the Nuremberg Code. A review of findings of the Advisory Committee on Human Radiation Experiments. *JAMA.* 1996;276:1667–1671.
4. Federman DD, Hanna KE, Rodriguez LL. *Responsible Research: A Systems Approach to Protecting Research Participants*. Washington, D.C.: National Academies Press; 2003.
5. Tuskegee Syphilis Study Ad Hoc Advisory Panel. *Final Report*. Washington, D.C., Department of Health, Education, and Welfare; 1973.
6. Department of Health, Education, and Welfare. Secretary's Task Force on the Compensation of Injured Research Subjects. *Final Report.* Bethesda, Maryland: Department of Health, Education, and Welfare; 1977. Publication No. OS-77-003.
7. The National Bioethics Advisory Commission. *Ethical and Policy Issues in Research Involving Human Participants.* Bethesda, Maryland: National Bioethics Advisory Commission; 2001. Available at: http://www.georgetown.edu/research/nrcbl/nbac/pubs.html. Accessed June 9, 2003.
8. Curran WJ. Compensation for injured research subjects. Regulation by informed consent. *N Engl J Med.* 1979;301:648–649.
9. NIH Awards to Medical Schools by Rank. Bethesda, Maryland: National Institutes of Health; 2002. Available at: http://grants2.nih.gov/grants/award/rank/medttlnod.htm. Accessed January 28, 2003.
10. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. *N Engl J Med.* 2003;348:721–726.
11. Doak C, Doak L, Root J. *Teaching Patients with Low Literacy Skills*. 2nd ed. Philadelphia, Pennsylvania: J. B. Lippincott Company; 1996.
12. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. *Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel.* Research Branch Report. Memphis, Tennessee: Chief of Naval Technical Training, Naval Air Station Memphis; 1975:8–75.
13. Kincaid JP, Gamble LG. Ease of comprehension of standard and readable automobile insurance policies as a function of reading ability. *J Reading Behav.* 1977;9:85–87.
14. Protection of Human Subjects (1993) (codified at 45 CFR §46.109, §50.20).
15. Report of the National Work Group on Literacy and Health. Communicating with patients who have limited literacy skills. *J Fam Pract.* 1998;46:168–176.
16. Robertson JA. Compensating injured research subjects: II. The law. *Hastings Cent Rep.* 1976;6:29–31.
17. OHRP Guidance. 'Exculpatory Language' in informed consent. Available at: http://ohrp.osophs.dhhs.gov/humansubjects/guidance/exculp.htm. Accessed June 9, 2003.
18. OHRP Guidance. Tips on informed consent. Available at: http://ohrp.osophs.dhhs.gov/humansubjects/guidance/ictips.htm. Accessed June 9, 2003.
19. Medical research: civil liability and compensation for personal injury—a discussion paper. *BMJ.* 1980;280:1172–1175.
20. Marwick C. Compensation for injured research subjects. *JAMA.* 1998; 279:1854.
21. Who should pay for research harm? *Med World News.* 1976;17:92.
22. Guest S. Compensation for subjects of medical research: the moral rights of patients and the power of research ethics committees. *J Med Ethics.* 1997;23:181–185.
23. Studdert DM, Brennan TA. No-fault compensation for medical injuries: the prospect for error prevention. *JAMA.* 2001;286:217–223.
24. Treatment of Research-Related Injuries to Human Subjects (1998) (codified at 38 CFR §17.85).
25. Human Subjects Research Review Board. United States Army Medical Research and Materiel Command. Medical Care for Research Related Injury (Clause 3.01). Available at: http://mrmc.www.army.mil/index.asp?EntryURL=/larindex.asp. Accessed June 4, 2003.
26. Brody BA. Research ethics: international perspectives. *Camb Q Healthc Ethics.* 1997;6:376–384.
27. Beecher HK. Human studies. *Science.* 1969;164:1256–1258.
28. Havighurst CC. Compensating persons injured in human experimentation. *Science.* 1970;169:153–157.

JOURNAL OF WOMEN'S HEALTH
Volume 14, Number 9, 2005

Educational Attainment but Not Literacy Is Associated with HIV Risk Behavior among Incarcerated Women

MICHAEL K. PAASCHE-ORLOW, M.D., M.A., M.P.H.,[1]
JENNIFER G. CLARKE, M.D., M.P.H.,[2] MEGAN R. HEBERT, M.A.,[2]
MOIRA K. RAY, B.S.,[2] and MICHAEL D. STEIN, M.D.[2]

ABSTRACT

***Purpose*: To identify the educational factors associated with HIV risk behaviors among incarcerated women.**

***Methods*: We evaluated a cohort of female detainees at The Rhode Island Adult Correctional Institute between July and September 2004. Among the 423 women who participated in the study, 55% did not have a high school diploma, 29% had ≤8th grade reading capacity, 32% had a learning disability, 37% had problem drinking, and 61% (257/423) reported HIV risk behavior.**

***Results*: In multiple logistic regression, participants who had completed high school had the lowest likelihood of HIV risk behavior (adjusted odds ratio [OR] 0.35, 95% confidence interval [CI] 0.12-1.00). There was no association between participants' literacy level and HIV risk behavior (adjusted OR 2.02, 95% CI 0.83-4.92).**

***Conclusions*: Correctional education programs to reduce HIV risk behavior should focus on those with low educational attainment irrespective of literacy skills.**

INTRODUCTION

THE PREVALENCE OF LOW educational attainment, low literacy, and learning disabilities has been found to be particularly high among prisoners.[1–4] According to the U.S. Department of Justice's 1997 survey of over 1 million state inmates, 68% did not have a high school diploma.[2] Among the 1147 subjects assessed in the prison sample of the National Adult Literacy Survey of 1992, 68%–72% were found to have limited prose, document, and quantitative literacy and lack the skills needed for independent participation in society.[1,5]

Prisoners' educational attainment and literacy have been correlated with correctional outcomes.[1,2,6,7] Indeed, the potential for correctional education programs to reduce recidivism has been an intensive area of research.[8–16] However, almost no research has explored the role of educational factors among prisoners for health outcomes or evaluated the potential health benefit of correctional education.

Prisoners have a high disease burden and are important vectors for infectious diseases.[6,17,18] Female prisoners with short incarcerations in particular have been identified as a high-risk population for the spread of HIV and sexually

[1]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts.
[2]Division of General Internal Medicine, Department of Medicine, Brown University School of Medicine, Providence, Rhode Island.

This research was supported by the National Institute on Alcohol Abuse and Alcoholism (RO1AA14495). M.S. is a recipeint of a NIDA Mid-Career Investigator Award (DA00512).

transmitted diseases (STDs), and the rate of HIV infection is higher in women than men.[19–23,23–25] Addictions to drugs and alcohol have been identified as key variables in the propagation of such risks, and an array of prison-based and postincarceration intervention models have been developed and are being evaluated.[19,26–28] It is crucial to determine ways to reduce HIV risk behaviors among incarcerated women returning to the community. Lower educational attainment has been reported as a risk factor for higher rates of HIV risk behavior among female inmates, although evaluation of different educational factors has not been conducted.[29]

Low educational attainment and low literacy have been linked to low levels of knowledge and worse health outcomes for such chronic diseases as diabetes, asthma, and HIV infection.[30–35] However, the relationship of educational attainment and literacy with HIV risk behavior has not been evaluated. We sought to examine the relationship between educational factors and HIV risk behavior among women in short-term incarceration. In addition, we sought to provide an estimate of the independent relationship of specific educational factors and HIV risk behavior in an adjusted analysis. We hypothesized that low educational attainment, low literacy, a history of special education, and a history of a learning disability would all be associated with higher HIV risk behaviors.

MATERIALS AND METHODS

Study design

This is a cross-sectional analysis of data including HIV risk behavior, alcohol consumption, educational attainment, literacy, history of special education, and demographic information obtained by face-to-face interviews from a cohort of female detainees within 4 days of arrival at The Rhode Island Adult Correctional Institute (ACI). Study approval was obtained from the Miriam Hospital Institutional Review Board, the Office for Human Research Protection, and the Medical Research Advisory Group at the ACI prior to initiation. A Certificate of Confidentiality was obtained from the federal government to further ensure participant privacy. The warden of the women's facility agreed to help guarantee participant confidentiality and granted permission for all interactions with the women to occur one-on-one with research assistants in unmonitored rooms.

Study site

The Rhode Island ACI is a unified correction system serving as a combined prison and jail, holding all pretrial and sentenced inmates in the state. There are approximately 2000 female commitments to the ACI annually, of which 400 result in sentences. Similar to women prisoners throughout the country, most women in the ACI are charged with nonviolent crimes (79%), 31% with drug-related charges. Over 72% of the women are less than 40 years old. In Rhode Island, 56% of women released from prison are reincarcerated within 1 year. For women who are not sentenced, 69% are released within 4 days, and 45% of those women who are sentenced are released in 6 months or less.

Outcome variable

The outcome variable of interest was HIV risk behavior. This was defined as a dichotomous variable based on response to the following query: During the last 3 months, have you had sex without using a condom *OR* have you shared any part of injection drug equipment (needle, syringe, cotton, cooker, or rinse water) at least once a month?

Primary independent variables

Four separate educational variables were evaluated: educational attainment, literacy, special education, and having an individualized educational plan (IEP). Consistent with the definition used by the National Center for Educational Statistics and the Bureau of Labor Statistics, educational attainment was defined as the last completed year of school.[1] Subjects who did not have a high school diploma were asked if and where they may have taken classes for the general equivalence degree (GED). Literacy is the capacity to use "printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential."[1] Literacy was evaluated with the 66-word version of the Rapid Estimate of Adult Literacy in Medicine (REALM). The REALM is based on the pronunciation of 66 medically relevant words. It is the most commonly used literacy instrument in the medical lit-

erature and is well correlated with other measures of medical and general literacy (WRAT 0.88, SORT-R 0.96, PIAT-R 0.97, TOHFLA 0.84).[36,37]

Based on REALM scores, participants are categorized as having ≥6th grade, 7th–8th grade, or ≥9th grade reading capacity. A history of special education was defined by affirmative endorsement of the question: Did you ever go to special education classes or work with a resource teacher outside of your regular classroom? An IEP is a personal pedagogic plan developed for students who are found to have a learning disability. Such plans are adapted to accommodate specific learning deficits identified by neuropsychological testing, available free in the state of Rhode Island for the past 25 years. Such plans include (1) assessment of the child's level of educational performance, (2) delineation of short-term and annual instructional objectives, (3) statement of the child's capacity to learn in regular education programs and the need for specified educational services required, (4) specified dates for initiation and completion of specified educational services, and (5) statement of the objectives, evaluation procedures, and results of at least an annual assessment of these goals.[38] Having an IEP was defined by affirmative endorsement of the question: When you were in school, did you ever have an IEP, an individualized educational plan?

Other independent variables

Other specific variables assessed included age, race (black, white, or Hispanic), and problem drinking. Problem drinking was defined as having a score of ≥8 on the Alcohol Use Disorder Identification Test (AUDIT).[39,40] The AUDIT is a 10-item instrument that evaluates consumption, dependence, and consequences of alcohol.

Sample and procedure

From February 4, 2004, through July 19, 2004, women entering the ACI who were awaiting trial were recruited to participate. Research assistants reviewed traffic sheets (daily printouts of all female inmates committed to the facility) on a daily basis, Monday (which included weekend traffic) through Friday, and attempted to contact each woman. Research staff collected data on which women declined participation, were released prior to contact, or did not meet inclusion criteria. Eligibility criteria included English speaking, housed in general facility population, age 18 or older, not yet sentenced, and able to competently provide verbal consent. If a woman was unable to be screened secondary to being in segregation, being ill, or in acute withdrawal from drugs or alcohol or both, her status was followed until she was released or could be approached for participation.

Eligible women met with research assistants who introduced themselves as members of a research team from Rhode Island Hospital, performing a brief survey regarding educational experiences. It was emphasized that study participation was completely voluntary and that no identifying information was going to be recorded on the woman's questionnaire.

Analysis

Fisher exact test and two-sample *t* tests were used to compare subject characteristics by HIV risk behavior. Analysis of variance (ANOVA) was used to evaluate the relationships among the educational variables. Bivariable logistic regression analysis was used to relate the primary independent variables educational attainment (≤8th grade, 9th–11th grade, high school graduate and GED), literacy (≤6th grade, 7th–8th grade, or ≥9th grade), history of special education (yes vs. no), history of having an IEP (yes vs. no), and demographic variables to HIV risk behavior and are presented as unadjusted odds ratios (ORs). Age was evaluated in tertiles (youngest 18–30, middle 31–38, and oldest 39–64). In addition, a multivariate logistic regression model was fit that included race, age, and problem drinking. All analyses were carried out using Stata, version 8.0 (College Station, TX).

RESULTS

During the study period, there were 966 nonredundant commitments; 459 women were released before contact was made, and research assistants attempted to recruit 507 subjects. Among these inmates, the participation rate was 83% (423 of 507): 52 refused to participate, 9 were non-English speaking, 16 were in segregation or on the hospital ward, and 7 were incompetent. Participants in the study had an average age of 34 years, 63% were Caucasian, 55% did not have a high school diploma, 29% had ≤8th grade reading ca-

pacity, 32% had either a history of special education (26 %) or a history of an individualized educational plan (15%), 37% had problem drinking, and 61% (257 of 423) reported HIV risk behavior. Younger age, being Caucasian, having lower educational attainment, and having problem drinking were each associated with HIV risk behavior. There was no difference in age ($p = 0.09$) or distribution of race/ethnicity ($p = 0.38$) between nonparticipants and participants. Demographic characteristics and the relationship to HIV risk factors are summarized in Table 1.

Unadjusted ORs for HIV risk behavior reveal a graded relationship for educational attainment, with more education being significantly related to lower HIV risk behavior (Table 2). There was also a graded relationship between literacy level and HIV risk behavior. Participants with lower literacy had less HIV risk behavior; however, this relationship was not statistically significant. Participants who were older, African American, and did not have problem drinking had lower odds for HIV risk behavior.

These same trends persisted in a multiple logistic regression model of adjusted ORs for HIV risk behavior (Table 2). Whereas participants who had completed high school had the lowest likelihood of HIV risk behavior (adjusted OR 0.35, 95% confidence interval [CI] 0.12-1.00), participants who had the highest literacy level had the highest likelihood of HIV risk behavior (adjusted OR of 2.02, 95% CI 0.83-4.92). Older participants (adjusted or [aOR] 0.46, 95% CI 0.25-0.82) and African Americans (aOR 0.50, 95% CI 0.29-0.86) had lower rates of HIV risk behavior. A history of problem drinking continued to be associated with lower rates of HIV risk behavior; however, this was no longer significant (aOR, problem drinking, yes vs. no: 1.60, 95% CI 0.97 -2.64)

DISCUSSION

We found that educational attainment is associated with HIV risk behavior among incarcerated woman. Specifically, women with the highest likelihood of HIV risk behavior were those who had left school prior to 9th grade. Staying in school into high school was associated with a 65% decreased odds of HIV risk behavior, and having a high school diploma was associated with a 70% decreased odds of HIV risk behavior. Contrary to

TABLE 1. DEMOGRAPHIC CHARACTERISTICS

		HIV risk behavior[a]	
	All % (n = 423)	*Yes* % (n = 257)	*No* % (n = 164)
Age mean (SD), range	34 (9), 18–64	**33 (9), 18–58**[b]	**36 (9), 18–64**
Race/ethnicity			
Caucasian	63 (261)	**67 (170)**	**56 (90)**
African American	25 (103)	**19 (49)**	**33 (53)**
Hispanic	10 (42)	**10 (26)**	**10 (16)**
Level of education completed			
≤8th grade	9 (39)	**12 (31)**	**4 (7)**
9th–11th grade	46 (193)	**46 (117)**	**46 (75)**
High school graduate	45 (191)	**42 (109)**	**50 (82)**
Observed literacy level[c]			
≤6th grade	10 (38)	9 (21)	12 (17)
7th–8th grade	19 (69)	19 (42)	19 (27)
≥9th grade	71 (258)	72 (162)	69 (96)
Special education, % yes	26 (109)	29 (73)	22 (36)
IEP,[d] % yes	15 (62)	16 (40)	14 (22)
Problem drinking[e], % yes	37 (156)	**42 (107)**	**29 (47)**

[a]HIV risk defined as responding yes to the question: During the last 3 months, have you had sex without using a condom or have you shared any part of injection drug equipment at least once a month? Data missing for 2 subjects.

[b]Variables with statistically significant differences ($p < 0.05$) are shown in bold.

[c]Rapid Estimate of Adult Literacy in Medicine (REALM) raw scores range from 0 to 66 and correspond to levels of reading capacity: ≤6th grade (0–44), 7th–8th grade (45–60), ≥9th grade (61–66).

[d]IEP, individualized educational plan.

[e]Problem drinking defined as ≥8 on the Alcohol Use Disorders Inventory Test (AUDIT) (0–40).

TABLE 2. UNADJUSTED ORS (95% CI) FOR HIV RISK BEHAVIOR

	Odds of reporting HIV risk behavior[a]	*Adjusted odds of reporting HIV risk behavior*[a]
Age		
Youngest	1	1
Middle	**0.61 (0.37–0.99)**[b]	0.59 (0.33–1.05)
Oldest	**0.50 (0.31–0.82)**	**0.46 (0.25–0.82)**
Race/ethnicity		
Caucasian	1	1
African American	**0.49 (0.31–0.78)**	**0.50 (0.29–0.86)**
Hispanic	0.86 (0.44–1.69)	0.82 (0.37–1.84)
Level of education completed		
≤8th grade	1	1
9th–11th grade	**0.35 (0.15–0.85)**	0.39 (0.13–1.12)
High school graduate	**0.30 (0.12–0.73)**	**0.35 (0.12–1.00)**
Observed literacy level[c]		
≤6th grade	1	1
7th–8th grade	1.26 (0.56–2.82)	1.89 (0.74–4.81)
≥9th grade	1.37 (0.69–2.72)	2.02 (0.83–4.92)
History of special education		
Yes	1.43 (0.90–2.27)	1.66 (0.87–3.15)
No	1	1
History of IEP[d]		
Yes	1.17 (0.66–2.06)	0.78 (0.38–1.60)
No	1	1
History of problem drinking[e]		
Yes	**1.78 (1.16–2.71)**	1.60 (0.97–2.64)
No	1	1

[a]HIV risk defined as responding yes to the question: During the last 3 months, have you had sex without using a condom or have you shared any part of injection drug equipment at least once a month? Data missing for 2 subjects.
[b]Variables with statistically significant differences ($p < 0.05$) are shown in bold.
[c]Rapid Estimate of Adult Literacy in Medicine (REALM) raw scores range from 0 to 66 and correspond to levels of reading capacity: ≤6th grade (0–44), 7th–8th grade (45–60), ≥9th grade (61–66).
[d]IEP, individualized educational plan.
[e]Problem drinking defined as ≥8 on the Alcohol Use Disorders Inventory Test (AUDIT) (0–40).

our hypothesis, however, literacy, a history of special education, and a history of a learning disability were not associated with HIV risk behavior.

The central message of these failed hypotheses may be that health prevention messages have been relatively successful in reaching across literacy and special education barriers for people who had stayed in school at least into high school. Specifically, staying in school was helpful even if it did not advance literacy skills. Early identification of learning disabilities in states like Rhode Island, which have broadly supported the evaluation and education of students with learning disabilities, may alter the HIV risk behavior for such subjects by helping to keep people with learning disabilities in school.

As this is a cross-sectional study, we can provide no evidence of a causal relationship for our main reported association between low educational attainment and HIV risk behavior. However, we speculate that many of the reasons participants dropped out of school early were directly related to either sex-risk or drug-risk behaviors. For instance, addiction and unwanted sexual activity are causes for leaving school early.[41] Similarly, the social interactions for young woman outside of school may bring people in contact with older men who may themselves offer a higher-risk behavior profile. Alternatively, it is possible that participants who stayed in school were influenced by this environment to be risk averse. It is unlikely, however, that the beneficial influence of remaining in school is mediated through aspects of cognitive function or literacy skills, as neither having a learning disability nor low literacy was associated with HIV risk behavior.

The lack of concordance for HIV risk behavior between educational attainment and literacy represents the fact that these variables measure different phenomena. Direct measurement of liter-

acy provides the opportunity to focus on a narrow band of specific skills currently possessed by participants. In contrast, the level of educational attainment is a historical marker of overall social function, status, and behavior. People often advance in school without adequate education, a phenomenon known as social promotion, and other people are fully literate even if they drop out of school. At a statistical level, this can be represented in the current study by the fact that only 7% of the variance in literacy levels is attributable to the level of education (ANOVA $R^2 = 0.07$).

A broad literature has documented the association between educational attainment and an array of healthcare outcomes.[42] The mechanisms by which this association is elaborated are likely complex and include such factors as life course social stratification that clusters types of experiences over an individual's lifetime, biological factors that predispose people both to education and health advantages, and social dynamics that involve the broader cultural, economic, policy, and political environment.[43] Similar mechanisms have been evoked to explain the connection between literacy and health,[35,44] but specific delineation of risk factors and mechanisms for crucial health conditions, such as HIV, may help inform the development and implementation of successful interventions.

African American women were found to have half the odds of reporting HIV risk behavior. To our knowledge, no previous study has identified race as an independent risk factor for HIV risk behavior among female detainees. In a cohort of female inmates from the same facility in 1999, Rich et al.[25] reported the HIV prevalence among African American women to be 5.9, which was 2.7 (95% CI 1.7-4.1) times higher than the rate found in Caucasian women.[25] Potentially, the high HIV prevalence among African American women in the past helped advance the public health message of decreased risk behavior among African American detainees currently in the facility. Similar to previous studies, younger age was found in this study to be associated with HIV risk behavior among female inmates.[45]

Several limitations in this study warrant discussion. The HIV risk variable used in this study combines sexual and drug use risks. It is certainly possibly that female inmates have divergent patterns of behavior for these different types of risk behaviors and that educational factors may play distinct roles in different types of HIV risk behavior. Similarly, the HIV risk variable used in this study requires a certain threshold of risk (monthly). Such a threshold forces complex phenomena to be viewed in a dichotomous fashion. Although this threshold may help focus on the people who are at highest risk, this particular threshold has not been validated to be the most useful predictor of HIV risk behavior and was not developed for this study question. Similarly, although not always observed in correctional settings, it is generally thought that self-reports of previous stigmatizing behavior may lead to underreporting.[46,47] Other than the literacy test, all data are self-reported. We collected no documentation to substantiate reported level of educational attainment, special education, or IEP status. The most likely error introduced by this would be that participants would overstate their level of educational attainment, which is a socially desirable phenomenon. It is improbable that such overstatement would preferentially occur among participants who also report less HIV risk behavior.

Although highly correlated with several other literacy tests, the test used in this study (REALM) has not been used in other studies of female inmates. This limits prevalence comparisons with other reports. An additional limitation relates to the potential generalizability of these findings. This study was conducted in a single facility, and this cohort exhibits a higher rate of educational attainment and literacy than previous studies for female prison inmates. Although 55% of participants did not have a high school diploma, this is considerably less than the 64% of female state prison inmates in 1997 who did not have a high school diploma.[2] Whereas 29% of participants in this study were classified as having low literacy, fully 60% of the female inmates tested by Moody et al.[3] were classified as having low literacy. The prevalence of low literacy in the current study is closer to the rates found outside of prisons. The only data available to provide a direct comparison is the National Adult Literacy Survey of 1992, which reported that the lowest level of literacy was found in 23% of community-dwelling women and 43% of female inmates. Despite a higher than expected level of education and literacy, this cohort also had a higher rate of participants who had been in special education in elementary school (26%) than what has been reported previously (16%).[16] The total portion of participants in this study with a learning disability (32%), how-

ever, is comparable to previous reports (33%).[16] This likely represents the level of support provided in Rhode Island to maintain students in mainstream learning environments. These considerations make it important to retest this study's findings in other settings.

The key implication of our findings is that to maximize the impact of correctional education programs to reduce HIV risk behavior, it may be beneficial to focus efforts on those with lower educational attainment irrespective of literacy skills. Correctional education settings provide the opportunity to present messages about HIV risk reduction. Adult education programs, which are offered in most correctional education programs, have had a rapid growth of health-oriented curricula.[16,48] As the highest rates of HIV risk behavior were observed among people who had dropped out of school, incarcerated women who had dropped out of school are the most appropriate targets for interventions to reduce HIV risk behavior.

REFERENCES

1. Haigler KO, Harlow C, O'Connor P, Campbell A. Literacy behind prison walls: Profiles of the prison population from the National Adult Literacy Survey. National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education, 1994.
2. Harlow C. Special report: Education and correctional populations. Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice, 2003.
3. Moody KC, Holzer CE III, Roman MJ, et al. Prevalence of dyslexia among Texas prison inmates. Tex Med 2000;96:69.
4. Rasmussen K, Almvik R, Levander S. Attention deficit hyperactivity disorder, reading disability, and personality disorders in a prison population. J Am Acad Psychiatry Law 2001;29:186.
5. Sum A, Kirsch IS, Taggart R. The twin challenges of mediocrity and inequality: Literacy in the U.S. from an international perspective. Literacy in the U.S. from an international perspective. Policy Information Report. Princeton, NJ: Educational Testing Service, 2002.
6. Gaes G, Kendig N. The skill sets and health care needs of released offenders. From prison to home: The effect of incarceration and reentry on children, families, and communities. National Policy Conference of the Urban Institute and the U.S. Department of Health and Human Services, January 30–31, 2002.
7. Flanagan TJ. Prison Education Research Project: Final report. Criminal Justice Center, Sam Houston State University and the Windham School System of the Texas Department of Criminal Justice-Institutional Division, 1994. Available at *www.eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/26/66/0c.pdf* Accessed January 27, 2004.
8. Jenkins DH, Steurer SJ, Pendry J. A post-release follow-up of correctional education program completers released in 1990–1991. J Correctional Education 1995;46:20.
9. Kelso CE. Recidivism rates for two education programs' graduates compared to overall Washington state rates. Journal of Correctional Education 2000; 51:233.
10. Jancic M. Does correctional education have an effect on recidivism? J Correctional Education 1998;49:152.
11. Hrabowski FA, Robbi J. The benefits of correctional education. J Correctional Education 2002;53:96.
12. Steurer SJ, Smith LG. Education reduces crime: Three-state recidivism study. Lanham, MD: Correctional Education Association, 2003.
13. Wilson DB, Gallagher CA, MacKenzie DL. A meta-analysis of corrections-based education, vocation, and work programs for adult offenders. J Res Crime Delinquency 2000;37:347.
14. O'Neil M. Correctional higher education: Reduced recidivism? J Correctional Education 1990;41:28.
15. Cecil DK, Drapkin DA, MacKenzie DL, Hickman LJ. The effectiveness of adult basic education and life-skills programs in reducing recidivism: A review and assessment of the research. J Correctional Education 2000;51:207.
16. Newman AP, Lewis W, Beverstock C. Prison literacy: Implications for program and assessment policy. Technical report TR93-1. Bloomingtom, IA: ERIC Clearinghouse on Reading and Communication Skills. 1993.
17. National Commission on Correctional Health Care. The health status of soon-to-be-released prisoners: A report to Congress. Chicago: National Commission on Correction Health Care, 2002:vol 1.
18. Hammett TM, Roberts C, Kennedy S. Health-related issues in prisoner reentry. Crime Delinquency 2002; 47:390.
19. Guyon L, Brochu S, Parent I, Desjardins L. At-risk behaviors with regard to HIV and addiction among women inprison. Women Health 1999;29:49.
20. McClelland GM, Teplin LA, Abram KM, Jacobs N. HIV and AIDS risk behaviors among female jail detainees: Implications for public health policy. Am J Public Health 2002;92:818.
21. Bond L, Semaan S.At risk for HIV infection: Incarcerated women in a county jail in Philadelphia. Women Health 1996;24:27.
22. Desai AA, Latta ET, Spaulding A, Rich JD, Flanigan TP.The importance of routine HIV testing in the incarcerated population: The Rhode Island experience. AIDS Educ Prev 2002;14:45.
23. Macalino GE, Vlahov D, Sanford-Colby S, et al. Prevalence and incidence of HIV, hepatitis B virus, and hepatitis C virus infections among males in Rhode Island prisons. Am J Public Health 2004;94:1218.

24. Gyarmathy VA, Neaigus A, Szamado S. HIV risk behavior history of prison inmates in Hungary. AIDS Educ Prev 2003;15:561.
25. Rich JD, Dickinson BP, Macalino G, etal. Prevalence and incidence of HIV among incarcerated and reincarcerated women in Rhode Island. J Acquir Immune Defic Syndr 1999;22:161.
26. Hogben M, St Lawrence JS. HIV/STD risk reduction interventions in prison settings. J Womens Health Gender-Based Med 2000;9:587.
27. Magura S, Kang SY, Shapiro J, O'Day J. HIV risk among women injecting drug users who are in jail. Addiction 1993;88:1351.
28. Farley JL, Mitty JA, Lally MA, etal.Comprehensive medical care among HIV positive incarcerated women: the Rhode Island experience. J Womens Health Gender-Based Med 2000;9:51.
29. Mullings JL, Marquart JW, Brewer VE. Assessing the relationship between child sexual abuse and marginal living conditions on HIV/AIDS-related risk behavior among women prisoners. Child Abuse Negl 2000; 24:677.
30. Davis TC, Arnold C, Berkel HJ, Nandy I, Jackson RH, Glass J. Knowledge and attitude on screening mammography among low-literate, low-income women. Cancer 1996;78:1912.
31. 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;25:337.
32. 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.
33. Williams MV, Baker DW, Honig EG, Lee TM, Nowlan A. Inadequate literacy is a barrier to asthma knowledge and self-care. Chest 1998;114:1008.
34. Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. Arch Intern Med 1998;158:166.
35. DeWalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: A systematic review of the literature. J Gen Intern Med 2004; 19:1228.
36. Davis TC, Long SW, Jackson RH, et al. Rapid estimate of adult literacy in medicine: A shortened screening instrument. Fam Med 1993;25:391.
37. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005;20:165.
38. American Academy of Pediatrics. Committee on Children with Disabilities. The pediatrician's role in development and implementation of an Individual Education Plan (IEP) and/or an Individual Family Service Plan (IFSP). Pediatrics 1999;104:124.
39. Conigrave KM, Hall WD, Saunders JB. The AUDIT questionnaire: Choosing a cut-off score. Alcohol Use Disorder Identification Test. Addiction 1995;90:1349.
40. Babor TF, Kranzler HR, Lauerman RJ. Early detection of harmful alcohol consumption: Comparison of clinical, laboratory, and self-report screening procedures. Addict Behav 1989;14:139.
41. NIMH Multisite HIV Prevention Trial Group. A test of factors mediating the relationship between unwanted sexual activity during childhood and risky sexual practices among women enrolled in the NIMH Multisite HIV Prevention Trial. Women Health 2001;33:163.
42. Gwatkin DR. Literacy, education and health development: Policy implications. Health Policy Educ 1982; 3:109.
43. Mackenbach JP, Howden-Chapman P. New perspectives on socioeconomic in equalities in health. Perspect Biol Med 2003;46:428.
44. Baker DW. Reading between the lines: Deciphering the connections between literacy and health. J Gen Intern Med 1999;14:315.
45. Martin V, Cayla JA, Moris ML, Alonso LE, Perez R. Predictive factors of HIV-infection in injecting drug users upon incarceration. Eur J Epidemiol 1998;14: 327.
46. Hser YI. Self-reported drug use: Results of selected empirical investigations of validity. NIDA Res Monogr 1997;167:320.
47. Hser YI, Maglione M, Boyle K. Validity of self-report of drug use among STD patients, ER patients, and arrestees. Am J Drug Alcohol Abuse 1999;25:81.
48. Wagner Daniel A, Venezky RL. Adult literacy: The next generation. Educ Researcher 1999;28:21.

Address reprint requests to:
Dr. Michael Paasche-Orlow
Assistant Professor of Medicine
Section of General Internal Medicine
Department of Medicine
Boston University School of Medicine
91 East Concord Street, Suite 200
Boston, MA 02118

E-mail: mpo@bu.edu

Notices of Privacy Practices

A Survey of the Health Insurance Portability and Accountability Act of 1996 Documents Presented to Patients at US Hospitals

Michael K. Paasche-Orlow, MD, MPH, Dan M. Jacob, BA,† and Joshua N. Powell, AB‡*

Background: Federal regulation requires hospitals to present patients with a Notice of Privacy Practices (NPP) that contains all stipulated content items, is readable by patients, and posted on institutional web sites.

Objective: The objective of this study was to determine whether the NPP texts contain the required content, if readability is influenced by local literacy rates, and if readability or the presentation of NPP texts in other languages is influenced by local rates of English proficiency.

Research Design: The authors conducted a cross-sectional study of the web sites of 115 hospitals selected from the *US News and World Report* list: Best Hospitals in America.

Measures: English NPP texts were evaluated for 18 content items and readability using the Flesch-Kincaid scale, which assigns the minimal grade level required to read a text (range, 0–16).

Results: NPP texts were available for all hospitals (115 of 115). A Spanish-language NPP was available for 25% (29 of 115). All content items were evident in 76% (87 of 115) of hospitals' NPP texts. The average grade-level readability of NPP text was 12.3 (95% confidence interval, 12.0–12.7). Readability was not associated with the rate of local literacy ($P = 0.07$). Hospitals with a lower local rate of English proficiency had NPP texts that were more difficult to read ($P = 0.03$) and did not present NPP texts in other languages more frequently ($P = 0.15$).

Conclusions: Although NPP texts typically cover the stipulated content, they are written beyond the reading capacity of the majority of American adults. Explicit federal guidance is needed to help privacy lawyers draft NPP texts that are both comprehensive and comprehensible. The goals of the Health Insurance Portability and Accountability Act of 1996 Privacy Rule cannot be met with NPP texts patients cannot decipher.

From the *Department of Medicine, Boston University School of Medicine, Boston, Massachusetts; †Healthcare Analytics, LLC, New York, New York; and the ‡Department of Preventive Medicine and Public Health, University of Kansas Medical Center, Kansas City, Kansas.

Reprints: Michael Paasche-Orlow, MD, MPH, Assistant Professor of Medicine, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, 91 East Concord Street, Suite 200, Boston, MA 02118. E-mail: mpo@bu.edu.

ISSN: 0025-7079/05/4306-0558

Key Words: HIPAA, privacy practices, readability, medical record

(*Med Care* 2005;43: 558–564)

On April 14, 2003, all covered healthcare providers were required to begin providing patients with a Notice of Privacy Practices (NPP).[1] The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule specifies that the NPP must communicate: 1) how a patient's health information may be used and disclosed by healthcare providers;[2] 2) 6 federally mandated rights regarding patients' health information;[3] and 3) providers' responsibilities regarding their patients' health information. The HIPAA Privacy Rule specifically delineates 18 elements that must be included in every valid NPP and defines under what conditions it must be provided, to whom, and how. Furthermore, to emphasize the importance of privacy considerations, the NPP must be provided during the first service delivery and on request, be posted in a clear and prominent location, and, for covered providers that maintain a web site, must be posted on the web site.

The goal of HIPAA is to protect patients' privacy, and the NPP is supposed to explain how this works. As such, the NPP is intended to be a document that is easy for patients to understand. Health providers are required to customize the NPP for their practice environment and to use the NPP to inspire conversations between patients and providers regarding privacy rights. Indeed, the Privacy Rule establishes a firm "duty to provide the notice in plain language so that the average reader can understand the notice."[4]

In 1992, the National Adult Literacy Survey (NALS) sponsored by the US Department of Education profiled the functional English-language literacy skills of over 26,000 American adults and found that half of US adults have limited or low literacy skills.[5] This means that the average US adult is unable to use complex texts to accomplish everyday tasks and lacks the skills needed for full participation in our current society, includ-

ing successful interaction with the healthcare system.[6] Because the HIPAA Privacy Rule is specifically intended to empower patients to gain access to their healthcare information and fulfill their privacy rights, it is particularly important to ensure that the NPP is a document that patients can understand. We hypothesized that NPPs would be available on institutional web sites and would meet the mandated content requirements but would not be readable by patients. Because NPP texts are supposed to be crafted to meet local needs, we hypothesized that lower rates of local literacy would be associated with NPPs that are easier to read and that lower rates of English proficiency would be associated with the NPP texts that are easier to read and presented in other languages.

METHODS

Data Sources

We surveyed the Internet web sites of 115 American Hospitals to obtain an electronic copy of each hospital's NPP. The list of hospitals was preselected from the *US News and World Report* ranking of America's Best Hospitals 2002; it included the honor roll (top 17 hospitals in that ranking) and a random selection of half of the remaining hospitals on the list.[7] English NPP texts were obtained from institutional web sites for 98% (113 of 115) of the hospitals and 2% (2 of 115) were sent to the authors by institutional compliance officers on request. In addition, each web site was evaluated for the availability of a NPP in any language other than English.

Two markers for language skills were used to reflect local language patterns. The National Adult Literacy Survey of 1992 involved in-person literacy assessment of a nationally representative sample of 24,944 adults in America; the National Institute for Literacy web site provided synthetic estimates from these data by congressional district for each hospital.[8] This same web site provided the rate, according to the 1990 US Census, respondents reported verbal English fluency by congressional district for each hospital. Low verbal English fluency was indicated by a subject responding that they "speak English not well or not at all." To differentiate these factors in this article, the former is referred to as the local rate of low literacy and the latter as the local rate of low English proficiency. Hospital web sites were repeatedly accessed for NPP text from August 2003 to October 2004.

Survey of Required Content

The HIPAA Privacy Rule stipulates 18 required elements that must be included in a valid NPP. We used a checklist to evaluate adherence to statutory guidelines; the 18 required elements are presented in Table 1. Two additional elements examined were statements regarding medical research and any stated costs associated with obtaining an accounting of disclosures and copies of medical records. Each NPP was independently reviewed by 2 authors who recorded each element as present, absent, or present but clearly different from statutory intent. Each instance of disagreement among reviewers was reevaluated in a joint conference for final classification until agreement was reached.

Text Abstraction

English NPPs, downloaded from hospital web sites, were converted into Microsoft Word 2000 (Microsoft, Redmond, WA) documents. We exclusively evaluated the NPP intended for use among a hospital's general patient population. NPP texts intended for special populations (eg, employee health plans) or special circumstances (eg, research-subject recruitment) were excluded from consideration.

Readability Score and Word Length

Readability of NPP text was measured with the Flesch-Kincaid readability scale (range, 0–16th grade).[9] The Flesch-Kincaid formula is based on the average sentence length (ASL) and the average number of syllables per word (ASW): Flesch-Kincaid reading grade level = (0.39) ASL + 11.8 (ASW) − 15.59.[10] Although it has been validated in adults up to a 16th-grade level, the formula is truncated erroneously at a 12th-grade level in Microsoft Word.[9,11] Microsoft Word does report the ASL and ASW without restriction. To circumvent this programming limitation, we created a program in Microsoft Access to solve the Flesch-Kincaid formula when given ASL and ASW for a given block of text. This data was then truncated at a 16th-grade reading level.

Many medical systems include the names of affiliated practice sites in the NPP text. Such lists vary in length and are often presented at the end of the document and/or as part of the pledge statement. To avoid the influence of such lists on Flesch-Kincaid and word length analyses, lists of affiliated practice sites were removed from the document.

Readability was evaluated for the document as a whole as well as separately for the pledge, uses and disclosure, and rights sections of each NPP. In some cases, the subsections were not well-defined and reviewers independently made judgments to decide, for example, which text constituted the pledge section. Reviewers compared the reading grade level for each section and when results varied by less than half a grade level, the Flesch-Kincaid score assigned to the document would be the average of the 2 findings. When reviewers' results varied by more than half a grade level, documents were reexamined in a joint conference to resolve disparate findings. Figure 1 provides excerpts of NPP text for 2 different required elements at different reading levels. In addition, the word length of each NPP was recorded.

Statistical Analysis

Analysis of variance was used to test the association of readability and local rates of low literacy. The mean readability of NPP text for hospitals that present foreign-language NPPs versus those that do not was compared using Wilcoxon rank sum

TABLE 1. Description of the Required Elements Examined for Each Notice of Privacy Practices From 115 U.S. Hospitals

Element Title	Description	Element Present % (n)	Element Present but Altered % (n)
Header	NPP with the following header: "This Notice describes how medical information about you may be used and disclosed and how you can get access to this information. Please review it carefully."	100% (115)	8% (9)
Uses and disclosures			
Treatment	Description (including at least 1 example) of the types of uses and disclosures that the hospital is permitted to make for treatment	99% (114)	10% (12)
Payment	Description (including at least 1 example) of the types of uses and disclosures that the hospital is permitted to make for payment	99% (114)	5% (6)
Healthcare operations	Description (including at least 1 example) of the types of uses and disclosures that the hospital is permitted to make for healthcare operations	99% (114)	8% (9)
Disclosure without authorization	Description of each of the other purposes for which the hospital is permitted or required to use or disclose protected health information without the patient's written consent or authorization	100% (115)	0% (0)
Disclosure with authorization	Statement that other uses and disclosures will be made only with the individual's written authorization and that the individual may revoke such authorization	95% (109)	3% (3)
Appointments	If the hospital contacts individuals to provide appointment reminders, it must say so in its NPP	96% (110)	0% (0)
Fundraising	If the hospital contacts individuals to raise funds, it must say so in its NPP	93% (107)	1% (1)
Individual rights			
	Right to request restrictions on certain uses and disclosures	99% (114)	3% (3)
	Right to access, inspect, and copy	100% (115)	3% (3)
	Right to request an amendment	100% (115)	2% (2)
	Right to receive an Accounting of Disclosures	99% (114)	3% (4)
	Right to receive confidential communications	99% (114)	13% (14)
	Right to receive a paper copy of the NPP	97% (112)	4% (4)
Complaints	A statement that patients may complain to the hospital and to the HHS Secretary if they believe their privacy rights have been violated	100% (115)	4% (5)
Contact information	Must contain the name, or title, and telephone number of a person or office to contact for further information regarding the hospital's privacy practices	97% (112)	16% (18)
Effective date	Must contain an effective date	93% (107)	3% (4)
Revisions	Must have statement that the hospital may revise its notice	99% (114)	1% (1)

NPP indicates Notice of Privacy Practices.

test. Similarly, the Wilcoxon rank sum test was used to compare mean readability according to the local rate of low English proficiency evaluated as a dichotomous variable based on the observed mean value of 4.9% (>mean versus <mean). All significance tests were 2-tailed. Analyses were conducted with Stata version 8 (College Station, TX).

RESULTS

The NPP texts had a mean length of 2922 ± 988 words (range, 904–6410). Spanish NPP text was available at 25% (29 of 115) of hospital web sites, and 7 of these hospitals presented NPP texts in additional languages as well, including Arabic, Urdu, Creole, Portuguese, Chinese, Russian, Korean, Farsi, Japanese, and Greek.

Fulfillment of Content Requirements

All required elements were presented in 76% (87 of 115) of hospitals' NPP texts. An additional 19% (22 of 115) of the NPP texts contained 17 of the 18 required elements. Of the remaining 5% (6 of 115) of the hospitals, 3 had NPP texts

Grade Level	Use and Disclosure of PHI for Research	Accounting of Disclosures
5th	Can we use your information for research? Yes, but if your privacy will not be protected we must first ask your permission. All research must be approved by the Institutional Review Board. This board makes sure research is safe.†	You have the right to get a list of times we share your information. A record of any times we share your information will be kept for six years. Your first request in a 12-month period is free. After that, we may charge for additional requests. †
11th	We perform medical research here. Our clinical researchers may look at your health records as part of your current care, or to prepare or perform research. They may share your health information with other [Hospital Name] researchers. All patient research conducted at [Hospital Name] goes through a special process required by law that reviews protections for patients involved in research, including privacy. We will not use your health information or disclose it outside [Hospital Name] for research reasons without either getting your prior written approval or determining that your privacy is protected.	You have the right to request an accounting of disclosures that we made about you for non-treatment, non-payment, and non-operations purposes. The information we will provide you will be from the past 6 years and will not include information before April 14, 2003. The first list that you request within a 12-month period is free and any additional lists will be provided at a nominal fee. We will respond to your request within 60 days.
13th	Research is a systematic investigation designed to develop or contribute to generalizable knowledge about the causes and treatment of disease. Those involved in research may use or disclose protected health information under certain conditions. Generally, data collected from the care of patients in a research study needs to be linked to the patient's record, so that the accuracy of the data collected can be confirmed or additional follow-up can be obtained. Data from multiple patients can be aggregated and analyzed statistically. When the results of the study are discussed or published, the identities of individual patients involved in the study are not revealed. All research involving human subjects at [Hospital Name] is reviewed and approved by an Institutional Review Board (IRB). The IRB complies with federal regulations to ensure your health information is appropriately used, stored, and accessed. The Privacy Rule permits us to use or disclose patient health information for research purposes without further notice to or written authorization from you in three instances: 1) Reviews preparatory to research, when setting up a research "protocol." 2) Research on a deceased individual's records, subject to state law. 3) When the IRB has waived the authorization requirement because it has determined there is no significant risk to privacy rights.	You have the right to request an "accounting of disclosures." An accounting of disclosures is a list of the people and/or organizations we've given your medical information to, with a number of notable exceptions. Those exceptions include, but are not limited to: disclosures of your medical information for purposes of treatment, payment or healthcare operations, or disclosures we've made pursuant to a valid authorization.

All texts are direct excerpts unless otherwise indicated.
† Text written by authors as no text identified at this reading level

FIGURE 1. Sample language for 2 paragraphs from Notice of Privacy Practices documents at various levels of readability.

that were missing 2 elements, 2 had NPP texts that were missing 3 elements, and 1 had an NPP that was missing 4 elements. The most frequent item absent was a statement regarding fundraising, as seen in Table 1. All but 1 hospital included a statement regarding research. Although 93% (107 of 115) included a general statement explaining that there could be charges for copying medical records or obtaining more than 1 accounting of disclosures in a 12-month period, 9 hospitals specified details for such costs.

The most frequent elements that appeared in a format that was clearly different from statutory intent: treatment (12 of 115), healthcare operations (9 of 115), and the right to receive confidential communications (14 of 115), as seen in Table 1. For the 14 NPP texts that presented the right to receive confidential communication in a format that was different from the statutory intent, the topic was raised with statements such as "You can request how and where we communicate with you," without explicitly introducing the idea of confidentiality.

Readability

The NPP text had an overall mean readability that was higher than a 12th-grade reading level (12.3; 95% confidence interval [CI], 12.0–12.7). The section describing patients' rights had a mean reading level of 10.7 (95% CI, 10.3–11.0), the pledge section had a mean reading level of 11.6 (95% CI,

TABLE 2. Flesch-Kincaid Reading Grade Level for Notices of Privacy Practices (n = 115 Hospitals)

	Overall	Pledge*	Use and Disclosure†	Rights and Responsibilities‡
Mean	12.3	11.6	13.3	10.7
95% confidence interval	12.0–12.7	11.1–12.2	13.0–13.7	10.3–11.0
Median	12.5	11.4	13.6	10.5
Interquartile range	11.2–13.6	9.9–13.7	12.3–14.8	9.4–11.6

*Pledge: Defines the covered entity's duties with respect to health information. The pledge section typically states that the covered entity is required by law to maintain the privacy of protected health information and to abide by the terms of the notice (CFR §164.520 (b)(1)(v) (A) and (B)). It often includes other promises of quality care and confidentiality.

†Use and disclosure: Describes the types of uses and disclosures that the covered entity is permitted or required to make with or without patient authorization (CFR §164.520 (b)(1)(ii)).

‡ Rights and responsibilities: Defines the individual's rights with respect to protected health information and includes a brief description of how the individual may exercise those rights (CFR §164.520 (b)(1)(iv)).

11.1–12.2), and the uses and disclosures section had a reading level of 13.3 (95% CI, 13.0–13.7) (Table 2).

Factors Thought to Influence Readability

The local rate of low literacy (range, 10–50%; mean, 25.4%) was not significantly associated with overall NPP readability ($P = 0.16$). Hospitals with >4.9% local rate of low English proficiency (range, 0–34%; mean, 4.9%) were not more likely to have foreign-language NPP texts available on their web sites ($P = 0.15$) and were significantly more likely to present NPP texts that are more difficult to read (12.8 versus 12.0, $P = 0.03$).

DISCUSSION

Our findings suggest that most hospitals present NPP texts that include the required content but are written at a level of complexity that far surpasses the average US reading capacity. The readability of NPP text does not appear to be influenced by the local rate of literacy. Contrary to our hypothesis, hospitals serving populations with lower rates of English proficiency were likely to present NPP texts that were more difficult to read and were not more likely to present NPP texts in other languages.

The strengths of this study that lend weight to our conclusions are the broad sample of top quality US hospitals and the use of standardized instruments for data abstraction. Nonetheless, several limitations should be kept in mind. Data were obtained exclusively from web sites. Although it is likely that the NPP texts presented on institutional web sites accurately reflect local practice, additional materials were not examined. In addition, the HIPAA compliance of hospitals included in this survey may not reflect typical practice in US hospitals. It is possible that hospitals on the *US News and World Report* list of Best Hospitals in America invest more resources in HIPAA compliance and present NPP texts that are different than other hospitals. It is unclear if this might introduce a bias to over- or underestimate the central findings of this analysis.

We chose the Flesch-Kincaid scale for our main analysis primarily because of its convenience. Of the dozens of readability scales, the Flesch-Kincaid scale is the most widely available for computerized use, because it is embedded in Microsoft Word. Other advantages include its wide use in studies of readability, excellent repeatability, and high correlation with other established readability scales (r = 0.87–0.90).[10,11] Unlike previous research using Microsoft Word, which artificially truncates readability at the 12th-grade level, this study was conducted with scores extending to the 16th grade, reflecting the full range of the scale.[12]

The Flesch-Kincaid formula is based on word and sentence length. The formula cannot account for the complexity introduced by short but unfamiliar medical or legal terms.[13] Similarly, such an approach does not evaluate the overall reading difficulty introduced by factors such font size, layout, and document length.[14–18] All of these limitations tend to result in underestimation of the actual reading skill required to understand NPP texts.

In addition to the frequent use of medical jargon (eg, cadaveric) and legal patois (eg, pursuant to), the NPP texts are encumbered by an additional linguistic barrier that stems from the HIPAA itself.[19] The Privacy Rule attributes specific and idiosyncratic meaning to common words and phrases. A set of specially defined terms, which are the conceptual and linguistic building blocks of the Privacy Rule, is ubiquitous in NPP texts. For example, the words "protected health information," "organized healthcare arrangement," and "hybrid covered entity" have unique meaning in the context of HIPAA but little meaning to average readers of an NPP. This feature is further complicated by the common use of acronyms such as PHI to denote such defined terms.

Almost half of American adults read at or below the 8th-grade level.[20] Although most texts had over a 12th-

grade reading level, several institutions such as Abbot Northwestern Hospital in Minneapolis and the Children's Hospital and Medical Center of Seattle present NPP text written at an 8th-grade reading level and can provide initial guidance for interested institutions. In addition, the Health Resources and Services Administration (HRSA), in consultation with the Office for Civil Rights, and other offices and agencies within the US Department of Health & Human Services have presented extensive guidance on the development of plain English NPP texts.[21] Beyond writing NPP texts to convey key concepts simply and directly, HRSA recommends pretesting draft NPP language with members of the target audience and provides a protocol with which this can be done.[21]

Such simplification is likely to benefit a broad range of patients because plain language improves comprehension and retention for all readers.[14,22–25] The Office of Civil Rights has projected that approximately 613 million NPP texts will be distributed annually.[4] Marginal readers typically read simple texts at approximately 160 words per minute.[26] Because the average text was 2922 words long, it would take more than 18 minutes for marginal readers to read an NPP. Most patients will not sustain such an effort and will simply not read the document, as is the case with informed consent forms.[27] Alternative methods for communicating the concepts advanced by the Privacy Rule such as multimedia systems are also likely to be of substantial benefit as has been shown in research relating to informed consent and risk communication.[28–32]

It is unlikely that institutions will broadly incorporate plain English texts such as the examples presented in Figure 1 unless regulators confirm that such texts meet the content requirements. However, if regulators presented a range of plain English template text options that could be tailored for local use, marked improvement in readability would result. In lieu of presenting prepared text, establishment of a specific grade-level readability standard as a benchmark for readability paired with endorsing the HRSA recommendation to pretest NPP text with the target audience could guide improvement toward fulfilling the plain language requirement.

In reviewing modifications made to the Privacy Rule in response to public comment, the Office of Civil Rights made clear that "nothing in the final modifications relieve a covered entity of its duty to provide the entire notice in plain language so the average reader can understand it."[4] However, the hospitals in this survey present NPP texts that are not written in plain language suitable for an average reader. Violation of the plain language requirement of the HIPAA Privacy Rule is prevalent. This is largely the result of insufficient federal guidance. Explicit instructions are needed to help privacy lawyers draft NPP texts that are both comprehensive and comprehensible.

REFERENCES

1. Health Insurance Portability and Accountability Act of 1996. Public Law No. 104–191, 110 Stat. (codified as amended in scattered sections of 18, 26, 29 and 42 USC); 42 USC 1320d–1320d-8 (West Supp 1998); 1998.
2. Privacy of Individually Identifiable Health Information. C.F.R. title 164, §520 (2002).
3. Privacy of Individually Identifiable Health Information. C.F.R. title 164, §520, §522, §524, §526, §528, and §530(d)(1) (2002).
4. Office for Civil Rights DoHaHS. Standards for Privacy of Individually Identifiable Health Information: 45 CFR Parts 160 and 164, Final Rule. Federal Register 2002;67(157):53181–53273. Available at: http://www.hhs.gov/ocr/hipaa/privruletxt.txt. Accessed March 10, 2004.
5. Kirsch I, Jungeblut A, Jenkins L, et al. *Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey.* Washington, DC: National Center for Education Statistics, US Department of Education; 1993.
6. Nielsen-Bohlman LT, Panzer AM, Hamlin B, et al. Institute of Medicine. *Health Literacy: A Prescription to End Confusion.* Committee on Health Literacy, Board on Neuroscience and Behavioral Health. Washington, DC: National Academies Press; 2004.
7. *US News and World Report.* Available at: http://www.usnews.com/usnews/home.htm. Accessed December 10, 2003.
8. *The State of Literacy in America: Synthetic Estimates of Adult Literacy Proficiency at the Local, State and National Levels.* Available at: http://www.nifl.gov/reders/reder.htm. Accessed May 17, 2004.
9. Kincaid JP, Fishburne RP, Rogers RL, et al. *Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy Enlisted Personnel.* Research Branch Report, Chief of Naval Technical Training: Naval Air Station Memphis; 8–75; 1975.
10. Doak C, Doak L, Root J. *Teaching Patients with Low Literacy Skills,* 2nd ed. Philadelphia: JB Lippincott Co; 1996.
11. Kincaid J, Gamble L. Ease of comprehension of standard and readable automobile insurance policies as a function of reading ability. *Journal of Reading Behavior.* 1977;9:85–87.
12. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. *N Engl J Med.* 2003;348:721–726.
13. Beyer DR, Lauer MS, Davis S. Readability of informed-consent forms. *N Engl J Med.* 2003;348:2262–2263.
14. Bjorn E, Rossel P, Holm S. Can the written information to research subjects be improved? —An empirical study. *J Med Ethics.* 1999;25: 263–267.
15. Doak LG, Doak CC, Meade CD. Strategies to improve cancer education materials. *Oncol Nurs Forum.* 1996;23:1305–1312.
16. Heinze-Lacey B, Saunders C, Sugar A. Improving the readability of informed consent documents. *IRB.* 1993;15:10–11.
17. Meade CD, Byrd JC. Patient literacy and the readability of smoking education literature. *Am J Public Health.* 1989;79:204–206.
18. Peterson BT, Clancy SJ, Champion K, et al. Improving readability of consent forms: what the computers may not tell you. *IRB.* 1992;14:6–8.
19. Kutzko D, Boyer GL, Thoman DJ, et al. HIPAA in real time: practical implications of the Federal Privacy Rule. *Drake Law Review.* 2003;51: 403.
20. Report of the National Work Group on Literacy and Health. Communicating with patients who have limited literacy skills. *J Fam Pract.* 1998;46:168–176.
21. Health Resources and Services Administration. *Plain Language Principles and Thesaurus for Making HIPAA Privacy Notices More Readable.* Available at: http://www.hrsa.gov/language.htm#N. Accessed May 17, 2004.
22. Jolly BT, Scott JL, Sanford SM. Simplification of emergency department discharge instructions improves patient comprehension. *Ann Emerg Med.* 1995;26:443–446.
23. Wagner L, Davis S, Handelsman MM. In search of the abominable consent form: the impact of readability and personalization. *J Clin Psychol.* 1998;54:115–120.
24. Young DR, Hooker DT, Freeberg FE. Informed consent documents: increasing comprehension by reducing reading level. *IRB.* 1990;12:1–5.

25. Coyne CA, Xu R, Raich P, et al. Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: a study of the Eastern Cooperative Oncology Group. *J Clin Oncol.* 2003;21:836–842.
26. Gray WB Jr. *How to Measure Readability.* Philadelphia: Dorrance & Co; 1975.
27. Lavelle-Jones C, Byrne DJ, Rice P, et al. Factors affecting quality of informed consent. *BMJ.* 1993;306:885–890.
28. Barbour GL, Blumenkrantz MJ. Videotape aids informed consent decision. *JAMA.* 1978;240:2741–2742.
29. Weston J, Hannah M, Downes J. Evaluating the benefits of a patient information video during the informed consent process. *Patient Educ Couns.* 1997;30:239–245.
30. Benitez O, Devaux D, Dausset J. Audiovisual documentation of oral consent: a new method of informed consent for illiterate populations. *Lancet.* 2002;20:1406–1407.
31. Sugarman J, McCrory DC, Powell D, et al. Empirical research on informed consent. An annotated bibliography. *Hastings Cent Rep.* 1999; 29:S1–42.
32. Wofford JL, Currin D, Michielutte R, et al. The multimedia computer for low-literacy patient education: a pilot project of cancer risk perceptions. *Med Gen Med.* 2001;20:3:23.

REVIEWS

The Prevalence of Limited Health Literacy

Michael K. Paasche-Orlow, MD, MPH,[1] Ruth M. Parker, MD,[2] Julie A. Gazmararian, MPH, PhD,[3] Lynn T. Nielsen-Bohlman, PhD,[4] Rima R. Rudd, ScD[5]

[1]Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [2]Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA; [3]Department of Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, GA, USA; [4]Board on Neuroscience and Behavioral Health, Institute of Medicine, Washington, DC, USA; [5]Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, MA, USA.

OBJECTIVE: To systematically review U.S. studies examining the prevalence of limited health literacy and to synthesize these findings by evaluating demographic associations in pooled analyses.

DESIGN: We searched the literature for the period 1963 through January 2004 and identified 2,132 references related to a set of specified search terms. Of the 134 articles and published abstracts retrieved, 85 met inclusion criteria, which were 1) conducted in the United States with ≥ 25 adults, 2) addressed a hypothesis related to health care, 3) identified a measurement instrument, and 4) presented primary data. The authors extracted data to compare studies by population, methods, and results.

MAIN RESULTS: The 85 studies reviewed include data on 31,129 subjects, and report a prevalence of low health literacy between 0% and 68%. Pooled analyses of these data reveal that the weighted prevalence of low health literacy was 26% (95% confidence interval [CI], 22% to 29%) and of marginal health literacy was 20% (95% CI, 16% to 23%). Most studies used either the Rapid Estimate of Adult Literacy in Medicine (REALM) or versions of the Test of Functional Health Literacy in Adults (TOFHLA). The prevalence of low health literacy was not associated with gender (P=.38) or measurement instrument (P=.23) but was associated with level of education (P=.02), ethnicity (P=.0003), and age (P=.004).

CONCLUSIONS: A pooled analysis of published reports on health literacy cannot provide a nationally representative prevalence estimate. This systematic review exhibits that limited health literacy, as depicted in the medical literature, is prevalent and is consistently associated with education, ethnicity, and age. It is essential to simplify health services and improve health education. Such changes have the potential to improve the health of Americans and address the health disparities that exist today.

KEY WORDS: prevalence; functional health literacy; health literacy; literacy.
DOI: 10.1111/j.1525-1497.2005.40245.x
J GEN INTERN MED 2005; 20:175–184.

Health literacy is increasingly described as the currency for improving the quality of health and health care in America.[1-3] In *Health Literacy: A Prescription to End Confusion*, the Institute of Medicine (IOM) described the growing body of literature documenting the magnitude and associations of limited literacy, and made recommendations for promoting a health-literate society.[4] This report adopted the definition used in Healthy People 2010, which 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."[3] Multiple studies indicate that people with limited health literacy have worse health status and higher rates of hospitalization.[4-6] Medical and public health literature also highlight the high reading demands made on people in need of important health information. Over 300 published articles document that most health materials are beyond the comprehension skills of most Americans.[6]

The National Adult Literacy Survey (NALS), a nationally representative household survey conducted in 1992, profiled the functional English language literacy skills of over 26,000 American adults and found that half of U.S. adults have limited or low literacy skills.[7] This means that American adults with average literacy skills have difficulty using complex texts to accomplish everyday tasks and lack the skills needed for full participation in our current society.[8] The prevalence of limited health literacy in medical settings, however, has not been systematically reviewed.

In its report, the IOM committee presents a sample of published studies on the prevalence and demographic associations of low health literacy. The current article extends the background review conducted for the IOM report to a systematic review of the medical literature. This article summarizes the methods and findings of published studies on the prevalence of limited health literacy in health care contexts and synthesizes these findings by evaluating demographic associations in pooled analyses. Understanding this evidence will help practitioners, researchers, and funding institutions formulate solutions to the pressing issues that emerge from a mismatch between system demands and the average literacy skills of health care consumers.

METHODS

Research Questions

The objectives of this review are to examine 1) the prevalence of low and marginal health literacy in the medical literature; 2) the methods used in studies evaluating the prevalence of limited health literacy in medical care; and 3) the demographic factors associated with low health literacy.

Finding Relevant Studies

In January 2004, bibliographic database search terms were used for article retrieval. Search terms were *functional health*

Accepted for publication July 1, 2004

The authors have no conflicts of interest to report.

Address correspondence and requests for reprints to Dr. Paasche-Orlow: Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, 91 East Concord Street, Suite-200, Boston, MA 02118 (e-mail: mpo@bu.edu).

literacy, literacy [as a title word] *AND health, numeracy, TOFHLA, Rapid Estimate of Adult Literacy in Medicine, REALM AND read, Wide Range Achievement Test, WRAT, Slosson oral reading test, SORT AND read, Peabody Individual Achievement Test, PIAT, National Adult Reading Test, NART, AMNART, Woodcock-Johnson AND test, medical terminology AND achievement, MART AND read, literacy assessment for diabetes, and adult basic education test.* Databases searched were MEDLINE (1966–2004), CINAHL (1982–2003), PsycInfo (1887–2004), Linguistics and Language Behavior Abstracts (1973–2004), and Sociological Abstracts (1974–2004). After developing the search parameters, identifying databases to target, and pursuing references in consultation with the coauthors and through review of the annotated bibliographies of the National Center for Adult Learning and Literacy,[4,9] the lead author conducted the search, retrieval, and selection process.

Inclusion Criteria

References were included if the study was conducted in the United States, related to health care or a health services inquiry, involved 25 or more adults, and provided evidence of direct testing of subject literacy.

Study Selection

After screening 2,132 references, 134 articles and published abstracts were retrieved and 85 were included in this review. Those excluded did not present primary data, were not conducted in the United States, involved fewer than 25 adult subjects, were not part of a health services inquiry or conducted in a medical context, or did not provide literacy prevalence.

Data Abstraction

All studies were evaluated for participation rate, study design, subject selection criteria, setting and dates, literacy measure used, vision testing, cognitive testing, demographic characteristics (age, gender, ethnicity, highest level of education, and income), and prevalence of low literacy and marginal literacy. Some researchers used multiple instruments to evaluate literacy, alternative versions of these instruments, or altered instruments. For example, data extracted from studies using the full Test of Functional Health Literacy in Adults (TOFHLA), the abbreviated TOFHLA, and the shortened TOFHLA were all included.

Statistical Analysis

Low health literacy was defined as the rate of subjects scoring at an inadequate level on versions of TOFHLA or at the sixth grade level and below on other measures. Marginal literacy was defined as the rate of subjects scoring at the marginal level on versions of TOFHLA or at the seventh to eighth grade levels on other measures. Weighted analyses of variance were used to compare the mean rates of low literacy according to quartiles of demographic characteristics: age, gender, education, and ethnicity. The percentage of subjects who had not completed high school or received a general education degree (GED) was used as the indicator of education. The percentage of female subjects was used to represent gender. Due to variation in the presentation of data on ethnicity, the percentage of black subjects was used as the indicator of ethnicity. The Wilcoxon rank-sum test was used to compare the rate of low literacy between studies conducted with the Rapid Estimate of Adult Literacy in Medicine (REALM) versus versions of the TOFHLA, between the languages of test administration (Spanish vs English), between the total pooled estimate versus studies with greater than 300 subjects, and between the total pooled estimate versus studies not conducted with convenience samples. Prevalence data derived from different iterations of the REALM and TOFHLA were combined to make the comparison by testing instrument. Pooled analyses were conducted with weighted means, that is, each study influenced analyses in proportion to the size of the population in that study. All significance tests were two-tailed. Analyses were conducted with Stata software, version 8 (Stata Corporation, College Station, TX).

RESULTS

The 85 studies included data on 31,129 subjects. Pooled analyses of these data reveal that over one quarter of subjects (26%; 95% confidence interval [CI], 22% to 29%; range, 0 to 68), had low health literacy. An additional fifth of subjects (weighted mean of 20%; 95% CI, 16% to 23%; range, 11 to 65) had marginal health literacy.

Systematic review of the published data on health literacy does not provide a nationally representative sample. Over one third of subjects did not complete high school (weighted mean of 37%; 95% CI, 32 to 41) and about half of all subjects were black (weighted mean of 55%; 95% CI, 48 to 62). Table 1 presents the studies included in this review;[10-94] an appendix (available online at http://www.blackwellpublishing.com/products/journals/suppmat/jgi/jgi40245/jgi40245.htm) includes the literacy rates and demographic characteristics for each of the studies. The first section of Table 1 presents studies conducted with the REALM, the second section presents studies conducted with versions of the TOFHLA, and the third section presents studies conducted with all other measures.

Analysis of Study Design and Methods

More than three quarters of the studies (79%; 67/85) were from convenience samples of subjects. Exclusion of studies conducted with convenience samples did not significantly alter the mean rate of low health literacy (24%; 95% CI, 16% to 33%; range, 9 to 48) in comparison with the total pooled estimate (P=.90). The sample size ranged from 26 to 3,260, with an average of 366 subjects and a standard deviation of 60. Elimination of small studies (N<300) did not significantly alter the mean rate of low health literacy (25%; 95% CI, 19% to 30%; range, 9 to 48) in comparison with the total pooled estimate (P=.48). Participation rate could be calculated from published information in half the studies (54%; 46/85) and had a weighted mean of 63% (range, 48% to 100%).

Many studies specifically excluded subjects who did not speak English (18%; 15/85),[12,17,19,30,70] read English (8%; 7/85),[34,35,45,85-97,92] or have English as their primary (5%; 4/85)[48,61,69,81] or first language (4%; 3/85).[37,52,84] Spanish-speaking subjects were tested in Spanish in 11% of studies (9/85).[5,53,55,59,63-67]

Visual function was mentioned as a specific criterion in 20% of studies (17/85).[13,21,37,41,48,52,53,61-63,66,76,85-87,89,92] While 7 studies mention cognitive disorders as an exclusion criterion,[33,34,41,53,69,84,92] only 2 specified details for this

Table 1. The Prevalence of Health Literacy Skills Among Various Populations

Part 1. Analyses Using the Rapid Estimate of Adult Literacy in Medicine (REALM)

Source	Research Objectives	Primary Location	(Surveyed/Eligible) Participation Rate
Ahluwalia et al. 2002[10]	To test association of literacy and returning for randomization in a smoking cessation trial	Atlanta, GA	(787/847) 93%
Al-Tayyib et al. 2002[11]	To test association of literacy and quality of responses in health survey measurements	Baltimore, MD	(992/1,224) 81%
Arnold et al. 2001[12]	To test association of literacy and tobacco knowledge, attitudes, and practices in pregnant women	LA	(599/623) 96%
Arozullah et al. 2002[13]	To test association of literacy and preventable hospital admission	Chicago, IL	(198/NS)
Bass et al. 2002[14]	To evaluate medical residents' ability to identify patients with low literacy	KY	(182/NS)
Bass et al. 2002[15]	To evaluate knowledge, self-efficacy, empowerment, and literacy in patients with diabetes	KY	(104/NS)
Beers et al. 2003[16]	To evaluate the REALM instrument according to subject ethnicity	Philadelphia, PA	(1,805/NS)
Bennett et al. 1998[17]	To test the association of literacy and stage of presentation of prostate cancer	Shreveport, LA	(212/221) 96%
Bennett et al. 2003[18]	To evaluate a screening instrument for literacy in caregivers of pediatric patients	Philadelphia, PA	(98/100) 98%
Bryant et al. 2003[19]	To test the association of literacy and adequacy of warfarin anticoagulation	Chapel Hill, NC	(58/71) 82%
Christensen & Grace 1999[20]	To assess the prevalence of low literacy in an indigent psychiatric population	Gainesville, FL	(45/45) 100%
Conlin & Schumann 2002[21]	To assess the prevalence of low literacy in cardiac surgery patients	Spokane, WA	(30/34) 88%
Coyne et al. 2003[22]	To test comprehension, satisfaction, accrual, and anxiety of improved informed consent statement	44 sites	(204/226) 90%
Davis et al. 1991[23]	To validate the REALM-125	LA, AR	(207/223) 93%
Davis et al. 1993[24]	To validate the shortened REALM (66-word version)	LA, AR, TX	(203/215) 94%
Davis et al. 1994[25]	To assess the prevalence of low literacy in caregivers of pediatric patients	Shreveport, LA	(396/414) 96%
Davis et al. 1996[26]	To evaluate knowledge and attitudes regarding mammography in low-literate, low-income women	Shreveport, LA	(417/445) 94%
Davis et al. 1998[27]	To test whether improved polio immunization pamphlet would improve comprehension in caregivers of pediatric patients	Shreveport, LA	(610/646) 94%
Davis et al. 1998[28]	To test whether improved informed consent statement would improve comprehension	Shreveport, LA	(183/205) 89%
Duffy & Snyder 1999[29]	To assess the prevalence of low literacy in emergency room patients and caregivers	South Carolina	(110/NS)
Foltz & Sullivan 1996[30]	To assess the prevalence of low literacy in cancer patients	New Orleans, LA	(63/73) 86%
Fortenberry et al. 2001[31]	To evaluate the association of literacy and gonorrhea-related care	CO, IN, NY, AL	(809/1,035) 78%
Gannon & Hildebrandt 2002[32]	To assess the prevalence of low literacy in patients at a women's health clinic	Midwestern	(50/61) 82%
Galloway et al. 2003[33]	To assess the prevalence of low literacy in neurology clinic patients	OH	(99/113) 88%
Hayes 1998[34]	To compare the effect of usual emergency department discharge instructions to instructions designed for elderly patients	Midwestern	(60/NS)
Hayes 2000[35]	To assess the prevalence of low literacy among patients in a rural emergency department	KS	(195/200) 98%
Hearth-Holmes et al. 1997[36]	To assess the prevalence of low literacy in patients with systemic lupus erythematosis	Shreveport, LA	(94/100) 94%
Kaufman et al. 2001[37]	To evaluate the association of literacy and breastfeeding in women at a public health clinic	Albuquerque, NM	(61/NS)
Kim et al. 2001[38]	To assess the prevalence of literacy and shared decision making in patients with prostate cancer	Chicago, IL	(30/NS)
Li et al. 2000[39]	To assess predictors of compliance in women with breast cancer	Rural South	(39/55) 71%
Lindau et al. 2002[40]	To evaluate the association of literacy with cervical cancer prevention knowledge and health behaviors	Chicago, IL	(529/584) 91%
Mayeaux et al. 1995[41]	To evaluate the association of literacy and Mini-mental State Examination scores	Shreveport, LA	(105/115) 91%
McNeill et al. 2003[42]	To assess the prevalence of low literacy in patients with HIV infection	Eastern NC	(55/NS)
Moon et al. 1998[43]	To assess the prevalence of low literacy in caregivers of pediatric patients	Washington, DC	(633/662) 96%
Murphy et al. 2001[44]	To assess the prevalence of low literacy in neurology patients	Shreveport, LA	(520/531) 98%
Raymond et al. 2002[45]	To evaluate comprehension of a label for an emergency contraception pill product	8 cities in U.S.	(395/NS)
Rothman et al. 2003[46]	To evaluation literacy as a variable in a diabetes management program	Chapel Hill, NC	(218/NS)
Sharp et al. 2002[47]	To assess the prevalence of low literacy in African-American women seeking colposcopy	Chicago, IL	(130/136) 96%
Williams et al. 1998[48]	To evaluate the association between literacy and asthma knowledge and inhaler technique	Atlanta, GA	(483/595) 81%
Wilson & McLemore 1997[49]	To assess the prevalence of low literacy in hospitalized orthopedic patients	Detroit, MI	(26/NS)

(*Continued*)

Table 1 (continued)

Source	Research Objectives	Primary Location	(Surveyed/Eligible) Participation Rate
Part 1. Analyses Using the Rapid Estimate of Adult Literacy in Medicine (REALM)			
Wilson et al. 2003[50]	To assess the prevalence of low literacy among elderly African-American anticoagulation patients	Detroit, MI	(65/NS)
Part 2. Analyses Using the Test of Functional Health Literacy in Adults (TOFHLA)			
Artinian et al. 2001[51]	To assess the prevalence of low literacy in Veterans Administration clinic patients	Midwest	(92/NS)
Benson & Forman 2002[52]	To assess the prevalence of low literacy in retirement community residents	Albuquerque, NM	(93/123) 76%
Gazmararian et al. 1999[53]	To assess the prevalence of low literacy in new Medicare enrollees	Cleveland, OH, Houston, TX, Tampa, FL, South FL	(3,260/6,734) 48%
Gazmararian et al. 1999[54]	To assess the prevalence of low literacy in women Medicaid managed care plan enrollees	Memphis, TN	(406/825) 49%
Golin et al. 2002[55]	To evaluate predictors of adherence to antiretroviral therapy	Los Angeles, CA	(140/233) 60%
Kalichman et al. 1999[56]	To evaluate literacy and adherence to antiretroviral therapies in patients with HIV infection	Atlanta, GA	(318/NS)
Kalichman & Rompa 2000[57]	To assess the prevalence of low literacy, HIV-related knowledge, and health status in patients with HIV infection	Atlanta, GA	(339/NS)
Kalichman et al. 2000[58]	To assess the prevalence of low literacy and HIV-related knowledge in patients with HIV infection	Atlanta, GA	(294/NS)
Lasater et al. 2002[59]	To assess the prevalence of low literacy and adherence in anticoagulation clinic patients	Denver, CO	(196/NS)
Montalto & Spiegler 2001[60]	To assess the prevalence of low literacy in rural health center patients	Charleston, WV	(70/183) 38%
Nurss et al. 1997[61]	To assess the prevalence of low literacy in patient with diabetes	Atlanta, GA	(131/222) 59%
Paasche-Orlow et al. 2003[62]	To assess the prevalence of low literacy in patients with asthma	Baltimore, MD	(80/NS)
Schillinger et al. 2002[63]	To evaluate literacy and diabetes outcomes	San Francisco, CA	(413/449) 92%
Shea et al. 2003[64]	To evaluate literacy and patient satisfaction	Philadelphia, PA	(2,494/NS)
Williams et al. 1995[65]	To assess the prevalence of low literacy in emergency department patients	Atlanta, GA	
		Los Angeles, CA	(2,659/2,856) 93%
Williams et al. 1998[66]	To assess the prevalence of low literacy in medical clinic patients	Atlanta, GA	
		Los Angeles, CA	(580/636) 91%
Win et al. 2003[67]	To assess the prevalence of low literacy in anticoagulation clinic patients	San Francisco, CA	(141/NS)
Part 3. Analyses Using Other Tests			
Coles et al. 1978[68]	To assess the prevalence of low literacy in hospitalized psychiatric patients	NJ	(48/NS)
Cooley et al. 1995[69]	To assess the prevalence of low literacy in Veterans Administration oncology clinic patients	Philadelphia, PA	(63/72) 88%
Currier et al. 2001[70]	To assess the prevalence of low literacy in psychiatric clinic patients	Los Angeles, CA	(53/55) 96%
Davis et al. 1991[23]	To validate the 125-word REALM test	AR, LA	(207/223) 93%
Davis et al. 1993[71]	To assess the prevalence of low literacy in patient substance use treatment centers	Shreveport, LA	(114/125) 91%
Davis et al. 1993[24]	To validate the 66-word REALM	AR, LA, TX	(203/215) 94%
Davis et al. 1994[25]	To assess the prevalence of low literacy among caregivers of pediatric patients	Shreveport, LA	(396/414) 96%
Fredrickson et al. 1995[72]	To assess the prevalence of low literacy in caregivers of pediatric patients	KS	(646/NS) 96%
Hanson-Divers 1997[73]	To validate the Medical Terminology Achievement Reading Test (MART)	NC	(405/NS)
Hartman et al. 1997[74]	To evaluate a nutritional education program designed for patients with low literacy	Twin Cities, MN	(204/NS)
Jackson et al. 1991[75]	To assess the prevalence of low literacy in primary care patients	Northwest LA	(528/544) 97%
Jackson et al. 1994[76]	To assess the prevalence of low literacy among older patients	NS	(272/281) 97%
Johnson & Fisher 1996[77]	To compare three reading tests in drug and alcohol users	NS	(123/NS)
Johnson et al. 1996[78]	To assess the prevalence of low literacy among drug users in an HIV/AIDS prevention study	AK, OH, CO, CA, LA	(412/NS)
Jubelirer et al. 1994[79]	To assess the prevalence of low literacy in oncology clinic patients	WV	(100/110) 91%
Kicklighter & Stein 1993[80]	To assess the prevalence of low literacy in patients with diabetes	Atlanta, GA	(58/NS)
Klinge & Dorsey 1993[81]	To assess the prevalence of low literacy in forensic psychiatric patients	Atascadero, CA	(350/NS)
Larson & Schumacher 1992[82]	To assess the prevalence of low literacy in Veterans Administration arthritis center patients	Philadelphia, PA	(100/103) 97%
Letz et al. 2003[83]	To validate the Neuro-behavioral Evaluation System version 3-Adult Reading Test (NES3-ART)	Atlanta, GA	(280/NS)
		Boston, MA	

(*Continued*)

Table 1.

Part 3. Analysis Using other tests (Continued)

Source	Research Objectives	Primary Location	(Surveyed/Eligible) Participation Rate
Manly et al. 2003[84]	To assess the prevalence of literacy and memory decline in ethnically diverse elders	New York, NY	(136/NS)
Meade & Byrd 1989[85]	To assess the prevalence of low literacy in smoking cessation patients	Milwaukee, WI	(258/NS)
Meade et al. 1989[86]	To evaluate simplification of smoking cessation literature and patient comprehension	Milwaukee, WI	(129/NS)
Meade et al. 1994[87]	To evaluate the effect of printed versus videotaped colon cancer education materials	Milwaukee, MI	(1,100/NS)
Miller et al. 1996[88]	To validate the Deaconess Informed Consent Comprehension test	St. Louis, MO	(269/NS)
Spandorfer et al. 1995[89]	To evaluate comprehension of discharge planning materials in discharged patients	Philadelphia, PA	(217/228) 95%
TenHave et al. 1997[90]	To assess the prevalence of low literacy in African-American patients in cardiovascular nutritional education	Washington, DC	(339/NS)
Weiss et al. 1992[91]	To evaluate the relationship between literacy and health status	Tucson, AZ	(193/197) 98%
Weiss et al. 1994[5]	To evaluate the relationship between literacy and health care costs	Tucson, AZ	(402/NS)
Wydra 2001[92]	To evaluate the effect of a self-care interactive multimedia program on fatigue in cancer patients	Los Angeles, CA Lebanon, NH Philadelphia, PA San Antonio, TX	(174/NS)
Zaslow et al. 2001[93]	To evaluate the relationship between literacy, depressive symptoms, and employment in families receiving welfare	Atlanta, GA	(351/427) 82%

determination and evaluated cognitive function among included subjects as a covariate.[34,53]

Testing Instruments

Several instruments tested multiple aspects of literacy including prose and document literacy, subdomains of reading capacity, and numeracy. Of the 14 literacy skills assessment instruments used by studies in this review, 9 were used in more than 1 study and are outlined in Table 2.[95-102] Most of these instruments are validated and have been used for the assessment of literacy skills in multiple contexts. Several instruments, developed for specific health contexts, are not yet well validated and have limited clinical data.

Several instruments, such as the Adult Basic Learning Examination (ABLE), evaluate comprehension of written text (prose literacy), capacity to use and understand tables and forms (document literacy), and arithmetic skills (numeracy). However, studies conducted with instruments that include more than 1 domain of literacy typically presented only a single summary measure. Five of the instruments are exclusively tests of word pronunciation, which is a component of prose literacy. While some of the instruments include subtests that focus on various domains of literacy, they were commonly employed in a restricted form. For example, pronunciation, but not spelling or arithmetic, was evaluated in most studies conducted with the Wide Range Achievement Test-Revised (WRAT-R).[94]

The majority of tests must be completed in English. Only 3 tests, the TOFHLA, the Instrumento Para Diagnostical Lecturas/Instrument for Diagnosis of Reading (IDL), and the Test of Adult Basic Education (TABE), provide the option for Spanish language testing. No other languages are accommodated by any of the instruments in this study. Fully 68% (58/85) of the studies used either the REALM or versions of the TOFHLA. Studies conducted with the REALM had similar rates of low literacy (22%; 95% CI, 17 to 27) as studies conducted with the TOFHLA (28%; 95% CI, 22 to 34).

Demographic Associations

The most common demographic features reported to be associated with health literacy were education level, age,[16,36,43,48,50-53,63,65,66,79-81] ethnicity,[11,12,16,17,30,36,40,43,51,53,57,63,72,78,83,84] geographic location, and income.[11,51,53,56,57,83] Studies reporting multivariate regression used a variety of different covariates. Most frequently, education and ethnicity remained significant in regression analyses.[36,43,51,53]

The rate of high school completion was significantly associated with the rate of low literacy (P=.02). For example, studies in the top quartile of high school graduation rate had the lowest prevalence of low literacy (10.7; 95% CI, 8.5 to 13.0).

The rate of black subjects was significantly associated with the rate of low literacy (P=.0003). For example, studies with the highest quartile of black subjects had the highest prevalence of low literacy (31.3; 95% CI, 26.8 to 35.7).

The average weighted age was 42.9 with a standard deviation of 1.49 years. Subject age was significantly associated with the rate of low literacy (P=.004). For example, studies in the lowest quartile of average age had the lowest prevalence of low literacy (15.9; 95% CI, 7.7 to 24.1). Studies with an average age over 50 years old (the top 2 quartiles together) had a prevalence of low literacy of 37.9 (95% CI, 31.6 to 44.2).

Overall, more women participated than men (65%; 95% CI, 59 to 70). The percentage of female subjects was not associated with the rate of low literacy (P=.38).

Data for Spanish language testing were separately reported for 5% of subjects (1,504/31,129). Subjects tested in Spanish had a higher rate of low literacy than those tested in English (44%, 95% CI, 26% to 62% vs 26%, 95% CI, 22% to 29%; P=.002). The combined weighted mean of low and marginal literacy for subjects tested in Spanish in the studies pre-

Table 2. Attributes of Literacy Assessment Instruments Used by at Least Two Studies in This Review

Test Name (Acronym)	Aspect of Literacy Tested	Scale	Correlation	Time Needed to Administer Test	Attributes	Number of Studies*
Peabody Individual Achievement Test (PIAT)[94,95]	Comprehension, 82 items; Pronunciation, 100 words; Spelling, 100 words; Numeracy, 100 items; Written expression	0–12th grade	Kaufman 0.84 Wechlsler 0.50	Untimed, approximately 1 hour	Training and materials needed; Long administration; Can be used with all ages	7
Rapid Estimate of Adult Literacy in Medicine, 66-word version (REALM-66)[24,96]	Pronunciation of medically relevant vocabulary, 66 words	4 categories: ≤ 3rd grade, 4th–6th grade, 7th–8th grade, ≥ 9th grade	WRAT 0.88 SORT-R 0.96 PIAT-R 0.97 TOFHLA 0.84	2–3 minutes	Quick and nonthreatening; Minimal training needed; Most frequently used	37
Slosson Oral Reading Test-Revised (SORT-R)[94,97]	Pronunciation, 200 words	10 categories: < 1st grade, 1–8, ≥ 9	PIAT-R 0.83–0.94 WRMT 0.90	5–10 minutes	Can be used in all ages (≥ 4 y); Minimal training	6
Test of Adult Basic Education (TABE)[91,98]	Reading, 50 items; Spelling, 20 items; Language, 55 items; Numeracy, 75 items	0.0–12.9th grade	GED 0.55–0.64	> 1.5 hour	Long; Training and materials needed; Available in Spanish; Automated scoring available	2
Test of Functional Literacy in Adults (TOFHLA)[99,100]	Prose literacy, 36-cloze items; Numeracy, 17 items	Inadequate, Marginal, Adequate	WRAT 0.74 REALM 0.84	18–22 minutes	Long; Tests document, prose, and numeracy; Spanish available; 14-point font available	9
TOFHLA abbreviated[53,100]	Prose literacy, 36-cloze items; Numeracy, 17 items	Inadequate, Marginal, Adequate		12 minutes	Long; Tests prose and numeracy; Spanish available 14-point font available	2
Short Test of Functional Literacy in Adults (STOFHLA)[99,100]	Prose literacy, 36-cloze items	Inadequate, Marginal, Adequate	REALM 0.81 TOFHLA 0.91	7 minutes	Prose only; Spanish available; 14-point font available	5
Wide Range Achievement Test-Revised (WRAT-R)[94,101]	Pronunciation, 57 words; Spelling; Arithmetic	3rd–12th grade	Stanford achievement test reading comprehension score 0.83 PIAT 0.62–0.91	3–5 minutes for the reading subtest	Brief; Can be used in ages 5–75; Age norms available; Reports specific grade level; Can be difficult for testers; Most limited to reading subtest	19
Woodcock Reading Mastery Test-Revised (WRMT-R) [94,102]	Pronunciation; Passage comprehension, cloze-type; Calculation; Applied problems; Dictation; Writing samples; Science Social studies; Humanities	0.5–16.9th grade	SORT-R 0.94 WRAT-R 0.91	50–60 minutes: 15 minutes for writing subtest, 5 minutes for other 8 subtests	Short form correlates well with full test (0.98); Can be used in ages 2–95; Norms for all educational levels available; Only letter-word pronunciation and passage comprehension subtests used in studies in this review	5

Number of studies in Table 1 using the particular instrument. The combined total is more than 85 as studies may have used more than one instrument.

sented in this review was 62% (95% CI, 55% to 68%; range, 54 to 71).

Comment

One in four subjects in the studies presented in this review had low health literacy and nearly half had low or marginal health literacy. The instruments used to measure literacy, populations sampled, and study methods varied across studies. Despite these methodological differences, the level of health literacy was consistently associated with level of education, ethnicity, and age. The level of health literacy was not associated with gender, or with data collection instrument (REALM or TOFHLA).

The strengths of this study that lend weight to our conclusions are the large sample size and the use of validated literacy assessment instruments in nearly all studies. However, this systematic review has several limitations. This article presents a systematic review of the published literature on the prevalence of limited health literacy. While a systematic review of the medical literature on literacy does not provide a nationally representative prevalence estimate, the NALS provides an opportunity to compare the results of this article to a nationally representative household survey of general literacy skills. The NALS assessment exhibited that 21% to 23% of American adults scored in the lowest of 5 skill levels and an additional 25% to 28% scored 1 level better.[7] People who score at level 1 or 2 of the NALS assessment lack the skills needed for full participation in our current society.[8] Direct correlation between the NALS scale and the scales used by the various instruments collected in this review is not possible. Yet, the similarities between the NALS findings and the prevalence estimates presented in this systematic review underscore the importance of basic literacy skills in health literacy and lend credence to the central findings of this article. Nevertheless, several important features of the literature on health literacy promote the possibility that the point estimates presented overestimate the true prevalence of limited health literacy. Publication bias may have limited the presentation of data on populations without high rates of limited health literacy. Similarly, it is likely that investigators interested in literacy would conduct such research in settings that have high rates of limited health literacy. In this review, it is apparent that investigators often conduct research in medical settings that cater to subjects with a low level of socioeconomic status. This may partially account for the overrepresentation of black subjects. Fully 55% of the subjects in the pooled analysis were black, 37% had not graduated high school, and the average age was 42.9 ± 1.49. These parameters have to be kept in mind in order to interpret the main findings of this article.

However, while subjects with low health literacy may be thought to be overrepresented in cited reports, it is notable that exclusion of studies conducted with subjects recruited by convenience sampling did not alter the rate of low health literacy presented. Further, while low income was associated with low literacy in some studies, income data were frequently not reported and in other studies this relationship was not exhibited. In addition, we were unable to include a summary measure of income in this analysis because of the multiple techniques used to report income data among the reviewed studies. It is also important to note that summary conclusions for demographic associations reported in this study are unadjusted. There may be systematic confounding among the demographic characteristics summarized in this article or with other unmeasured features. However, the association between health literacy and ethnicity found in this systematic review suggests the importance of incorporating health literacy improvements in efforts related to addressing health disparities.[103]

There are at least 4 important reasons that this literature review may actually present a conservative assessment of the prevalence of limited health literacy. First, studies reviewed for this analysis focused almost exclusively on aspects of reading and numeracy. However, all domains of literacy—listening, speaking, writing, reading, and numeracy—are relevant to health literacy. The IOM report supports a broad concept of health literacy which includes not only the 5 skills above but cultural and conceptual knowledge of health as well.[4] Oral literacy skills of listening and speaking are essential to patient-provider interactions, public health communication, and understanding direct to consumer marketing.[4,104,105] Future research on measuring health literacy should enhance our ability to capture these other important components of health literacy.

Second, the current analysis was based on a limited pool of data for languages other than English. As people commonly maneuver through the health care system with family and friends, another important aspect of the prevalence of low health literacy relates to concepts such as linguistically isolated households and social support.[106,107] Approximately 47 million individuals in the United States (18% of the total population) speak a language other than English at home. This rate has increased between 1990 and 2000.[106] Studies excluding subjects who are not native English speakers may exclude this portion of the American population, and are likely to underestimate the prevalence of low health literacy. Furthermore, the challenge of evaluating health literacy in languages that have direct phoneme-grapheme correspondence, such as Spanish and Haitian-Creole, may be difficult to surmount as the quickest and most commonly used instruments are based on word pronunciation tests.[108]

Third, most studies reviewed did not evaluate vision or cognition. Vision and cognitive capacity contribute to health literacy, and accurate assessment with the instruments used in these studies assumes normal or corrected vision and normal cognition. Vision and cognition must be tested, especially in populations such as the elderly where such deficits are known to be common and underreported.[109] The failure to evaluate cognitive capacity likely yielded an underestimate of low health literacy in studies with older populations. Word pronunciation, used in many of the studies included in this review, is commonly retained in the face of significant dementia and has even been used as a marker for premorbid intelligence in demented cohorts.[110]

Fourth, the studies may have been influenced by participation bias. People with limited literacy may participate less frequently in research.[51,111,112] Such a bias is clearly a concern as half of the studies did not disclose information to calculate the participation rate and the weighted participation rate for the remaining studies was 63% (range, 48% to 100%).

In 2003, the U.S. Department of Education initiated the National Assessment of Adult Literacy (NAAL), which contains expanded health-related components.[113] This second national

literacy assessment of American adults will provide data on the percentage of persons with inadequate or marginal literacy skills who can perform specific health literacy tasks related to clinical, prevention, and navigation activities. The NAAL focuses on adults' ability to use prose, documents, and numbers to accomplish specific tasks and will be used to gauge progress in meeting the Healthy People 2010 objective related to improving health literacy.

The Agency for Healthcare Research and Quality (AHRQ) recently performed an evidenced-based review of health literacy interventions and the influence of literacy on health outcomes and disparities. However, the AHRQ did not address the question of prevalence.[114] The current systematic review summarizing the prevalence of health literacy skills in American adults as depicted by reports in the medical literature will complement the AHRQ study. However, the focus on patients' literacy skills in these reviews reflects the state of the current literature, and should not distract attention from the overwhelming complexity of the health care system. The discourse of health literacy should address both the high literacy demands of complex systems and the skills required by individuals to navigate systems of care to self-manage chronic conditions and promote their health.

Conclusions

A pooled analysis of published reports on health literacy cannot provide a nationally representative prevalence estimate. This systematic review exhibits that limited health literacy, as depicted in the medical literature, is prevalent and is consistently associated with education, ethnicity, and age. It is essential to simplify health services and improve health education. Such changes have the potential to improve the health of Americans and address the health disparities that exist today.

The conclusions and opinions in this article are not necessarily those of the Institute of Medicine or its Committee on Health Literacy. The authors gratefully acknowledge the work of Allison M. Panzer, Institute of Medicine, National Academies, in amassing background material for this project.

REFERENCES

1. **Adams K, Corrigan JM**, eds. **Institute of Medicine.** Priority Areas for National Action: Transforming Health Care Quality. Committee on Identifying Priority Areas for Quality Improvement. Washington, DC: National Academies Press; 2003.
2. **Carmona RH.** Health literacy in America: the role of health care professionals. Prepared remarks given at the American Medical Association House of Delegates Meeting, Chicago, IL, June 14, 2003. Available at: http://www.surgeongeneral.gov/news/speeches/ama061403.htm. Accessed February 19, 2004.
3. **Ratzan SC, Parker RM.** Introduction. In: Selden CR, Zorn M, Ratzan SC, Parker RM, eds. National Library of Medicine Current Bibliographies in Medicine: Health Literacy. Vol. NLM, Pub. no. CBM 2000–1. 2000. Bethesda, MD: National Institutes of Health, U.S. Department of Health and Human Services. Available at: http://www.nlm.nih.gov/pubs/cbm/hliteracy.html. Accessed January 15, 2004.
4. **Nielsen-Bohlman LT, Panzer AM, Hamlin B, Kindig DA**, eds. 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. April 2004. Available at: http://www.nap.edu/catalog/10883.html/. Accessed April 19, 2004.
5. **Weiss BD, Blanchard JS, McGee DL, et al.** Illiteracy among Medicaid recipients and its relationship to health care costs. J Health Care Poor Underserved. 1994;5:99–111.
6. **Rudd RE, Moeykens BA, Colton TC.** Health and Literacy: A review of medical and public health literature. In: Comings J, Garners B, Smith C, eds. Health and Literacy. New York, NY: Jossey-Bass; 1999.
7. **Kirsch I, Jungeblut A, Jenkins L, Kolstad A.** Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey. Washington, DC: National Center for Education Statistics, U.S. Department of Education; 1993.
8. **Sum A, Kirsch IS, Taggart R.** The Twin Challenges of Mediocrity and Inequality: Literacy in the U.S. from an International Perspective. Policy Information Report. Princeton, NJ: Educational Testing Service; 2002.
9. **NSCALL.** National Center for the Study of Adult Learning and Literacy: Health Literacy Literature. Available at: http://www.hsph.harvard.edu/healthliteracy/literature.html. Accessed December 1, 2003.
10. **Ahluwalia JS, Richter K, Mayo MS, et al.** African American smokers interested and eligible for a smoking cessation clinical trial: predictors of not returning for randomization. Ann Epidemiol. 2002;12:206–12.
11. **Al-Tayyib AA, Rogers SM, Gribble JN, Villarroel M, Turner CF.** Effect of low medical literacy on health survey measurements. Am J Public Health. 2002;92:1478–81.
12. **Arnold CL, Davis TC, Berkel HJ, Jackson RH, Nandy I, London S.** Smoking status, reading level, and knowledge of tobacco effects among low-income pregnant women. Prev Med. 2001;32:313–20.
13. **Arozullah AM, Lee S, Khan T, Kurup S.** Low health literacy increases the risk of preventable hospital admission. J Gen Intern Med. 2003;18(suppl 1):221.
14. **Bass PF III, Wilson JF, Griffith CH, Barnett DR.** Residents' ability to identify patients with poor literacy skills. Acad Med. 2002;77:1039–41.
15. **Bass PF, Moore MA, Rising W, Ritchie CS.** Differences in knowledge, self-efficacy, empowerment, and literacy among diabetic patients. J Gen Intern Med. 2003;18(suppl 1):168.
16. **Beers BB, McDonald VJ, Quistberg DA, Ravenell KL, Asch DA, Shea JA.** Disparities in health literacy between African American and non-African American primary care patients.. J Gen Intern Med. 2003; 18(suppl 1):169.
17. **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;16:3101–4.
18. **Bennett IM, Robbins S, Al Shamali N, Haecker T.** Screening for low literacy among adult caregivers of pediatric patients. Fam Med. 2003;35: 585–90.
19. **Bryant B, Malone R, Ayscue D, DeWalt DA, Pignone MP.** The effect of literacy and anticoagulation knowledge on the adequacy of warfarin anticoagulation for patients with atrial fibrillation. J Gen Intern Med. 2003;18(suppl 1):169.
20. **Christensen RC, Grace GD.** The prevalence of low literacy in an indigent psychiatric population. Psychiatr Serv. 1999;50:262–3.
21. **Conlin KK, Schumann L.** Literacy in the health care system: a study on open heart surgery patients. J Am Acad Nurse Pract. 2002;14:38–42.
22. **Coyne CA, Xu R, Raich P, et al.** Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: a study of the Eastern Cooperative Oncology Group. J Clin Oncol. 2003;21:836–42.
23. **Davis TC, Crouch MA, Long SW, et al.** Rapid assessment of literacy levels of adult primary care patients. Fam Med. 1991;23:433–5.
24. **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.
25. **Davis TC, Mayeaux EJ, Fredrickson D, Bocchini JA Jr, Jackson RH, Murphy PW.** Reading ability of parents compared with reading level of pediatric patient education materials. Pediatrics. 1994;93:460–8.
26. **Davis TC, Arnold C, Berkel HJ, Nandy I, Jackson RH, Glass J.** Knowledge and attitude on screening mammography among low-literate, low-income women. Cancer. 1996;78:1912–20.
27. **Davis TC, Fredrickson DD, Arnold C, Murphy PW, Herbst M, Bocchini JA.** A polio immunization pamphlet with increased appeal and simplified language does not improve comprehension to an acceptable level. Patient Educ Couns. 1998;33:25–37.
28. **Davis TC, Holcombe RF, Berkel HJ, Pramanik S, Divers SG.** Informed consent for clinical trials: a comparative study of standard versus simplified forms. J Natl Cancer Inst. 1998;90:668–74.
29. **Duffy MM, Snyder K.** Can ED patients read your patient education materials? J Emerg Nurs. 1999;25:294–7.

30. **Foltz A, Sullivan J.** Reading level, learning presentation preference, and desire for information among cancer patients. J Cancer Educ. 1996; 11:32–8.
31. **Fortenberry JD, McFarlane MM, Hennessy M, et al.** Relation of health literacy to gonorrhoea related care. Sex Transm Infect. 2001;77:206–11.
32. **Galloway G, Murphy P, Chesson AL, Martinez K.** MDA and AAEM informational brochures: can patients read them? J Neurosci Nurs. 2003; 35:171–4.
33. **Gannon W, Hildebrandt E.** A winning combination: women, literacy, and participation in health care. Health Care Women Int. 2002;23:754–60.
34. **Hayes KS.** Randomized trial of geragogy-based medication instruction in the emergency department. Nurs Res. 1998;47:211–8.
35. **Hayes KS.** Literacy for health information of adult patients and caregivers in a rural emergency department. Clin Excell Nurse Pract. 2000; 4:35–40.
36. **Hearth-Holmes M, Murphy PW, Davis TC, et al.** Literacy in patients with a chronic disease: systemic lupus erythematosus and the reading level of patient education materials. J Rheumatol. 1997;24:2335–9.
37. **Kaufman H, Skipper B, Small L, Terry T, McGrew M.** Effect of literacy on breast-feeding outcomes. South Med J. 2001;94:293–6.
38. **Kim SP, Knight SJ, Tomori C, et al.** Health literacy and shared decision making for prostate cancer patients with low socioeconomic status. Cancer Invest. 2001;19:684–91.
39. **Li BD, Brown WA, Ampil FL, Burton GV, Yu H, McDonald JC.** Patient compliance is critical for equivalent clinical outcomes for breast cancer treated by breast-conservation therapy. Ann Surg. 2000;231:883–9.
40. **Lindau ST, Tomori C, Lyons T, Langseth L, Bennett CL, Garcia P.** The association of health literacy with cervical cancer prevention knowledge and health behaviors in a multiethnic cohort of women. Am J Obstet Gynecol. 2002;186:938–43.
41. **Mayeaux EJ Jr, Davis TC, Jackson RH, et al.** Literacy and self-reported educational levels in relation to Mini-mental State Examination scores. Fam Med. 1995;27:658–62.
42. **McNeill EL, Estrada CA, Pinn T, Baro AL, Collins C, Byrd JC.** Impact of health literacy on patients with human immunodeficiency virus infection. J Gen Intern Med. 2003;18(suppl 1):182.
43. **Moon RY, Cheng TL, Patel KM, Baumhaft K, Scheidt PC.** Parental literacy level and understanding of medical information. Pediatrics. 1998; 102:e25.
44. **Murphy PW, Chesson AL, Berman SA, Arnold CL, Galloway G.** Neurology patient education materials: do our educational aids fit our patients' needs? J Neurosci Nurs. 2001;33:99–104.
45. **Raymond EG, Dalebout SM, Camp SI.** Comprehension of a prototype over-the-counter label for an emergency contraceptive pill product. Obstet Gynecol. 2002;100:342–9.
46. **Rothman RL, Pignone MP, Malone R, Bryant B, DeWalt DA, Crigler B.** A longitudinal analysis of the relationship between literacy and metabolic control in patients with diabetes. J Gen Intern Med. 2003;18(suppl 1):155.
47. **Sharp LK, Zurawski JM, Roland PY, O'Toole C, Hines J.** Health literacy, cervical cancer risk factors, and distress in low-income African-American women seeking colposcopy. Ethn Dis. 2002;12:541–6.
48. **Williams MV, Baker DW, Honig EG, Lee TM, Nowlan A.** Inadequate literacy is a barrier to asthma knowledge and self-care. Chest. 1998;114:1008–15.
49. **Wilson FL, McLemore R.** Patient literacy levels: a consideration when designing patient education programs. Rehabil Nurs. 1997;22:311–7.
50. **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.
51. **Artinian NT, Lange MP, Templin TN, Stallwood LG, Hermann CE.** Functional health literacy in an urban primary care clinic. Internet J Adv Nurs Pract. 2001;5:11.
52. **Benson JG, Forman WB.** Comprehension of written health care information in an affluent geriatric retirement community: use of the test of functional health literacy. Gerontology. 2002;48:93–7.
53. **Gazmararian JA, Baker DW, Williams MV, et al.** Health literacy among Medicare enrollees in a managed care organization. JAMA. 1999;281: 545–51.
54. **Gazmararian JA, Parker RM, Baker DW.** Reading skills and family planning knowledge and practices in a low-income managed-care population. Obstet Gynecol. 1999;93:239–44.
55. **Golin CE, Liu H, Hays RD, et al.** A prospective study of predictors of adherence to combination antiretroviral medication. J Gen Intern Med. 2002;17:756–65.
56. **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.
57. **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;25:337–44.
58. **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.
59. **Lasater L, Davidson A, Mehler P.** Patient literacy, adherence, and anticoagulation therapy outcomes: a preliminary report. J Gen Intern Med. 2003;18:179.
60. **Montalto NJ, Spiegler GE.** Functional health literacy in adults in a rural community health center. W V Med J. 2001;97:111–4.
61. **Nurss JR, El-Kebbi IM, Gallina DL, et al.** Diabetes in urban African Americans: functional health literacy of municipal hospital outpatients with diabetes. Diabetes Educ. 1997;23:563–8.
62. **Paasche-Orlow M, Brancati F, Rand C, Krishnan J.** Education of patients with asthma and low literacy. J Gen Intern Med. 2003;18(Suppl 1):227.
63. **Schillinger D, Grumbach K, Piette J, et al.** Association of health literacy with diabetes outcomes. JAMA. 2002;288:475–82.
64. **Shea JA, Guerra C, Weiner J, Aguirre A, Schaffer M, Asch DA.** Health literacy and patient satisfaction. J Gen Intern Med. 2003;18(suppl 1):187.
65. **Williams MV, Parker RM, Baker DW, et al.** Inadequate functional health literacy among patients at two public hospitals. JAMA. 1995;274: 1677–82.
66. **Williams MV, Baker DW, Parker RM, Nurss JR.** Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. Arch Intern Med. 1998;158:166–72.
67. **Win K, Machtinger E, Wang F, Chan LL, Rodriguez ME, Schillinger D.** Understanding of warfarin therapy and stroke among ethnically diverse anticoagulation patients at a public hospital. J Gen Intern Med. 2003;18(suppl 1):278.
68. **Coles GS, Roth L, Pollack IW.** Literacy skills of long-term hospitalized mental patients. Hosp Community Psychiatry. 1978;29:512–6.
69. **Cooley ME, Moriarty H, Berger MS, Selm-Orr D, Coyle B, Short T.** Patient literacy and the readability of written cancer educational materials. Oncol Nurs Forum. 1995;22:1345–51.
70. **Currier GW, Sitzman R, Trenton A.** Literacy in the psychiatric emergency room. J Nerv Ment Dis. 2001;189:56–8.
71. **Davis TC, Jackson RH, George RB, et al.** Reading ability in patients in substance misuse treatment centers. Int J Addict. 1993;28: 571–82.
72. **Fredrickson DD, Washington RL, Pham N, Jackson T, Wiltshire J, Jecha LD.** Reading grade levels and health behaviors of parents at child clinics. Kans Med. 1995;96:127–9.
73. **Hanson-Divers EC.** Developing a medical achievement reading test to evaluate patient literacy skills: a preliminary study. J Health Care Poor Underserved. 1997;8:56–69.
74. **Hartman TJ, McCarthy PR, Park RJ, Schuster E, Kushi LH.** Results of a community-based low-literacy nutrition education program. J Community Health. 1997;22:325–41.
75. **Jackson RH, Davis TC, Bairnsfather LE, George RB, Crouch MA, Gault H.** Patient reading ability: an overlooked problem in health care. South Med J. 1991;84:1172–5.
76. **Jackson RH, Davis TC, Murphy P, Bairnsfather LE, George RB.** Reading deficiencies in older patients. Am J Med Sci. 1994;308:79–82.
77. **Johnson ME, Fisher DG.** Evaluating three reading tests for use with alcohol and other drug-abusing populations. Alcohol Clin Exp Res. 1996;20:1125–9.
78. **Johnson ME, Fisher DG, Davis DC, et al.** Assessing reading level of drug users for HIV and AIDS prevention purposes. AIDS Educ Prev. 1996;8:323–34.
79. **Jubelirer SJ, Linton JC, Magnetti SM.** Reading versus comprehension: implications for patient education and consent in an outpatient oncology clinic. J Cancer Educ. 1994;9:26–9.
80. **Kicklighter JR, Stein MA.** Factors influencing diabetic clients' ability to read and comprehend printed diabetic diet material. Diabetes Educ. 1993;19:40–6.

81. **Klinge V, Dorsey J.** Correlates of the Woodcock-Johnson Reading Comprehension and Kaufman Brief Intelligence Test in a forensic psychiatric population. J Clin Psychol. 1993;49:593–8.
82. **Larson I, Schumacher HR.** Comparison of literacy level of patients in a VA arthritis center with the reading level required by educational materials. Arthritis Care Res. 1992;5:13–6.
83. **Letz R, DiIorio CK, Shafer PO, Yeager KA, Henry TR, Schomer DL.** A computer-based reading test for use as an index of premorbid general intellectual level in North American English-speaking adults. Neurotoxicology. 2003;24:503–12.
84. **Manly JJ, Touradji P, Tang MX, Stern Y.** Literacy and memory decline among ethnically diverse elders. J Clin Exp Neuropsychol. 2003;25: 680–90.
85. **Meade CD, Byrd JC.** Patient literacy and the readability of smoking education literature. Am J Public Health. 1989;79:204–6.
86. **Meade CD, Byrd JC, Lee M.** Improving patient comprehension of literature on smoking. Am J Public Health. 1989;79:1411–2.
87. **Meade CD, McKinney WP, Barnas GP.** Educating patients with limited literacy skills: the effectiveness of printed and videotaped materials about colon cancer. Am J Public Health. 1994;84:119–21.
88. **Miller CK, O'Donnell DC, Searight HR, Barbarash RA.** The Deaconess Informed Consent Comprehension Test: an assessment tool for clinical research subjects. Pharmacotherapy. 1996;16:872–8.
89. **Spandorfer JM, Karras DJ, Hughes LA, Caputo C.** Comprehension of discharge instructions by patients in an urban emergency department. Ann Emerg Med. 1995;25:71–4.
90. **TenHave TR, Van Horn B, Kumanyika S, Askov E, Matthews Y, Adams-Campbell LL.** Literacy assessment in a cardiovascular nutrition education setting. Patient Educ Couns. 1997;31:139–50.
91. **Weiss BD, Hart G, McGee DL, D'Estelle S.** Health status of illiterate adults: relation between literacy and health status among persons with low literacy skills. J Am Board Fam Pract. 1992;53:257–64.
92. **Wydra EW.** The effectiveness of a self-care management interactive multimedia module. Oncol Nurs Forum. 2001;28:1399–407.
93. **Zaslow MJ, Hair EC, Dion MR, Ahluwalia SK, Sargent J.** Maternal depressive symptoms and low literacy as potential barriers to employment in a sample of families receiving welfare: are there two-generational implications? Women Health. 2001;32:211–51.
94. **Spreen O, Strauss E.** A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, 2nd ed. New York, NY: Oxford University Press; 1998.
95. **Markwardt FC Jr.** Peabody Individual Achievement Test-Revised. Circle Pines, MN: American Guidance Service. Available at: http://www.agsnet.com/Group.asp?nMarketInfoID=31&nCategoryInfoID=2503&nGroupInfoID=a29060. Accessed February 24, 2004.
96. **Doak C, Doak L, Root J.** Teaching Patients with Low Literacy Skills, 2nd ed. Philadelphia, PA: J. B. Lippincott Company; 1996.
97. **Slosson RL.** Slosson Oral Reading Test. East Aurora, NY: Slosson Educational Publications, Inc. Available at: http://www.slosson.com/productCat6886.ctlg. Accessed February 24, 2004.
98. **Sabatini JP.** McGraw Hill Contemporary: Tests of Adult Basic Education (TABE). Available at: http://www.ctb.com/products/product_summary.jsp?FOLDER%3C%3Efolder_id=1408474395217279&bmUID=1104292646342. Accessed February 10, 2004.
99. **Nurss JR, Parker RM, Williams MV, Baker DW.** Test of Functional Literacy in Adults, 2nd ed.. Snow Camp, NC: Peppercorn Books & Press; 2001.
100. **Nurss JR, Parker RM, Williams MV, Baker DW.** Test of Functional Literacy in Adults. Greensboro, NC: Peppercorn Books and Press, Inc. Available at: http://www.peppercornbooks.com/catalog/advanced_search_result.php?osCsid=e9102eea98709128d04e2ad9b77able7&search_id=1&search_in_description=1&inc_subact=1&keywords=TOFHLA&osCsid=e9102eea98709128d04e2ad9b77able7. Accessed February 24, 2004.
101. **Wilkinson GS.** Harcourt: Wide Range Achievement Test 3 (WRAT3). Available at: http://harcourtassessment.com/haiweb/Cultures/en-US/Products/Product+Detail.htm?CS_ProductID=015-8987-90X&CS_Category=AchievementBasicSkills&CS_Catalog=TPC-USCatalog. Accessed February 10, 2004.
102. **Woodcock RW.** Woodcock Reading Mastery Tests. Circle Pines, MN: American Guidance Service, Inc. Available at: http://www.agsnet.com/Group.asp?nGroupInfoID=a16640. Accessed February 24, 2004.
103. **Parker RM, Ratzan SC, Lurie N.** Health literacy: a policy challenge for advancing high-quality health care. Health Aff (Millwood). 2003;22: 147–53.
104. **Weiss BD, Coyne C.** Communicating with patients who cannot read. N Engl J Med. 1997;337:272–4.
105. **Finan N.** Visual literacy in images used for medical education and health promotion. J Audiov Media Med. 2002;25:16–23.
106. **Shin HB, Bruno R.** Language Use and English-speaking Ability: 2000. Issued October 2003. Available at: http://www.census.gov/prod/2003pubs/c2kbr-29.pdf. Accessed January 22, 2004.
107. **Lee SY, Arozullah AM, Cho YI.** Health literacy, social support, and health: a research agenda. Soc Sci Med. 2004;58:1309–21.
108. **Nurss JR, Baker DW, Davis TC, Parker RM, Williams MV.** Difficulties in functional health literacy screening in Spanish-speaking adults. J Reading. 1995;38:632–7.
109. **Friedman SM, Munoz B, Rubin GS, West SK, Bandeen-Roche K, Fried LP.** Characteristics of discrepancies between self-reported visual function and measured reading speed Salisbury Eye Evaluation Project Team. Invest Ophthalmol Vis Sci. 1999;40:858–64.
110. **Meyer J, Xu G, Thornby J, Chowdhury M, Quach M.** Longitudinal analysis of abnormal domains comprising mild cognitive impairment (MCI) during aging. J Neurol Sci. 2002;201:19–25.
111. **Baker DW, Parker RM, Williams MV, et al.** The health care experience of patients with low literacy. Arch Fam Med. 1996;5:329–34.
112. **Parikh NS, Parker RM, Nurss JR, Baker DW, Williams MV.** Shame and health literacy: the unspoken connection. Patient Educ Couns. 1996; 27:33–9.
113. **National Assessments of Adult Literacy: National Center of Education Statistics, Institute for Education Sciences, U.S. Department of Education**. Available at: http://nces.ed.gov/naal/. Accessed February 19, 2004.
114. **De Walt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP.** Literacy and health outcomes: a systematic review of the Literature. J Gen Intern Med. 2004;19:1228–39.

Tailored Education May Reduce Health Literacy Disparities in Asthma Self-Management

Michael K. Paasche-Orlow, Kristin A. Riekert, Andrew Bilderback, Arjun Chanmugam, Peter Hill, Cynthia S. Rand, Fred L. Brancati, and Jerry A. Krishnan

Department of Medicine, Boston University School of Medicine, Boston, Massachusetts; and Departments of Medicine, Emergency Medicine, and Epidemiology, Johns Hopkins University, Baltimore, Maryland

Rationale: **Although inadequate health literacy has been associated with lower asthma medication knowledge and worse metered-dose inhaler (MDI) technique, the relationship between health literacy and the capacity to learn asthma self-management skills is unknown.**
Objectives: **In this prospective cohort study of adults hospitalized for severe asthma exacerbations at two inner-city hospitals, we examined the relationship between inadequate health literacy and difficulties learning and retaining instructions about discharge medications and appropriate MDI technique.**
Methods: **At hospital discharge, participants received one-on-one, 30-min, guideline-based, written and oral instruction about their asthma discharge regimen as well as appropriate MDI technique.**
Measurements and Main Results: **Seventy-three patients were enrolled. Inadequate health literacy was identified in 16 (22%) participants. Before instruction, inadequate health literacy was associated with lower asthma medication knowledge (5.2/10 vs. 7.2/10, $p < 0.001$) and worse MDI technique (3.2/6 vs. 3.9/6, $p = 0.03$). However, inadequate health literacy was not associated with difficulty learning ($p = 0.33$) or retaining ($p = 0.35$) instructions about the discharge regimen. Similarly, inadequate health literacy was not associated with difficulty learning ($p = 0.26$) or retaining ($p = 0.97$) appropriate MDI technique. Results were similar in multivariable models adjusted for demographic characteristics and asthma severity indicators.**
Conclusions: **These findings suggest that inadequate health literacy is a surmountable barrier to learning and remembering key asthma self-management skills.**

Keywords: asthma; education; functional health literacy

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" (1–4) and is most often measured by reading comprehension of health-related information (5, 6). Multiple studies indicate that inadequate health literacy is associated with worse health status and higher rates of hospitalization across a number of patient populations (4, 7, 8), including patients with diabetes mellitus, patients with HIV infection, and the elderly (9–12). However, there are relatively few data about the effects of inadequate health literacy in patients with asthma, a common chronic respiratory disorder affecting 5 to 10% of the U.S. population (13). In cross-sectional studies, inadequate health literacy has been linked to lower asthma-related knowledge (14, 15) and improper use of metered dose inhalers (MDIs) (14). Although national asthma guidelines recommend patient education to improve patient–physician partnerships for care (16), no studies have evaluated the extent to which inadequate health literacy serves as a barrier for patients to learn and retain asthma self-management skills (17).

The objective of this study was to examine the relationship between inadequate health literacy and difficulties learning and retaining instructions about discharge medications and appropriate MDI technique. We recruited an inner-city predominately African-American patient population hospitalized for asthma exacerbations, a group at high risk for low health literacy and a two- to threefold risk of death from asthma exacerbations (18–20). We hypothesized that inadequate health literacy is associated with lower knowledge of asthma medications, improper use of MDI technique, as well as difficulty learning and retaining instructions about discharge medications and proper MDI technique. These hypotheses represent secondary goals of a study also intended to examine adherence to corticosteroid therapy after hospital discharge (21). Some results of the current study have been previously reported in the form of an abstract (22).

METHODS

Study Design

This was a prospective cohort study conducted from April 2001 through October 2002. Adults (age, $\geq$ 18 yr) admitted with a physician diagnosis of asthma exacerbation to two inner-city academic medical centers were screened for eligibility. Patients with other chronic lung diseases, those with a contraindication to corticosteroids, patients (or physicians) who declined consent, and investigators' patients were excluded. The study was approved by the Johns Hopkins Institutional Review Board.

Study Procedures

At hospital discharge (discharge study visit), eligible subjects were approached about participating in this study. Patients who provided written, informed consent were enrolled. Participants completed an interviewer-administered survey about sociodemographics (age, sex, ethnicity, education, income), smoking history (never, ever, current), and asthma-related health care use (emergency department or hospitalization) in the past year. On the basis of a review of medical records and interview responses, subjects were classified as having a history of near-fatal asthma exacerbation if they had ever been intubated due to asthma (yes vs. no).

Health literacy was measured with the Short Test of Functional Health Literacy in Adults and classified as having inadequate health literacy (yes [score $\leq$ 16/36] vs. no [score $>$ 16/36]), based on published recommendations (23, 24). The Short Test of Functional Health Literacy in Adults is a test of reading comprehension that takes 5 to 7 min to administer. The test has a set of sentences from medical scenarios with key words missing. Subjects select words to complete the sentences from a list provided. There are 36 items and each correct response is awarded 1 point (range of possible scores, 0–36). The test has high internal consistency (Cronbach's $\alpha = 0.97$) and is well correlated with the Rapid Estimate of Adult Literacy in Medicine (Spearman correlation, 0.81)

(*Received in original form September 28, 2004; accepted in final form July 3, 2005*)

Supported by the Parker B. Francis Fellowship Award and the National Institutes of Health (HL67850, HL063333).

Correspondence and requests for reprints should be addressed to Michael Paasche-Orlow, M.D., M.P.H., Assistant Professor of Medicine, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, 91 East Concord Street, Suite 200, Boston, MA 02118. E-mail: mpo@bu.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 172. pp 980–986, 2005
Originally Published in Press as DOI: 10.1164/rccm.200409-1291OC on August 4, 2005
Internet address: www.atsjournals.org

and the full Test of Functional Health Literacy in Adults (Spearman correlation, 0.91) (25).

In addition, we also assessed patients' understanding of asthma medications. There were no previously validated measures to assess patients' understanding of asthma medications (inhaled corticosteroids, oral corticosteroids, and short-acting bronchodilators). Therefore, we developed 10 items based on existing asthma knowledge scales, professional opinion, and the desire for each item to be directly related to medication use. Participants were given 1 point for each correctly answered item. Based on a total possible score of 10, we calculated a mean asthma medication knowledge score (total score/10). Internal consistency reliability was adequate (Cronbach $\alpha = 0.62$) and comparable to another published asthma knowledge scale (26).

Participants were asked to demonstrate use of an MDI. Participants' MDI technique was assessed on the basis of the following six criteria (16, 27–30): (*1*) shaking the MDI, (*2*) exhaling before actuation, (*3*) closing lips tightly around the mouthpiece, (*4*) pressing down once on the canister, (*5*) taking a full/deep breath without triggering the auditory "whistle" indicator of the spacer, and (*6*) holding breath for at least 5 s. One point was given for each step for a total possible score of 6 as follows: 0 (unable to perform any step) to 6 (perfect technique; Figure 1). The asthma discharge regimen was standardized to include both oral corticosteroids (prednisone: 20 mg/tablet, 2 tablets each morning for 7 d), inhaled corticosteroids (fluticasone MDI: 220 μg/puff, 2 puffs twice/d), and a short-acting bronchodilator (albuterol MDI: 2 puffs every 4 h as needed). These medications and a spacer for use with MDIs were provided for free to study participants. Participants took part in a one-on-one, 30-min-long, guideline-based, written and oral asthma-education session about this discharge regimen. This session included education about appropriate MDI technique; the research assistant provided both verbal instructions and demonstrated proper technique to study participants. After the education intervention and before discharge, participants were asked to state the following: (*1*) the name of each medication, (*2*) the number of tablets or actuations to be taken each time, (*3*) the number of times per day the medication should be taken, and (*4*) the prescribed duration of medication use. Teaching was repeated until participants could successfully demonstrate mastery of the asthma discharge regimen (defined as correctly describing items 1–4 for all three of the medicines) and MDI technique (30). The number of rounds of teaching necessary to achieve mastery of the asthma discharge regimen and MDI technique was recorded.

Participants returned for a 2-wk follow-up visit. At the follow-up visit, participant's MDI technique and knowledge of the asthma discharge regimen were reevaluated. In addition, at follow-up, asthma symptom control was assessed using the six symptom items in the Asthma Control Questionnaire (31). The total possible asthma symptom control was (0, well controlled, to 6, extremely poorly controlled).

In a subset of participants (n = 46) enrolled in the current study, we also electronically measured patients' use of inhaled and oral corticosteroids after hospital discharge. Adherence to inhaled corticosteroids was assessed using the Doser CT (Meditrac, Inc., Hudson, MA), which records the number of actuations of the fluticasone MDI each day (32, 33). Adherence to oral corticosteroids was assessed using the MEMS TrackCap (AARDEX Ltd., Union City, CA), which records the number of prednisone bottle openings each day (34–36). Adherence was defined as percentage of prescribed use after hospital discharge (use/prescribed use × 100%) and was calculated for both inhaled and oral corticosteroids. Poor adherence to corticosteroid therapy was defined *a priori* as electronically measured adherence to inhaled corticosteroids of less than 50% during the 2-wk period after hospital discharge or electronically measured adherence to oral corticosteroid therapy of less than 50% from hospital discharge through Day 7 (oral corticosteroids were prescribed for the first 7 d after discharge).

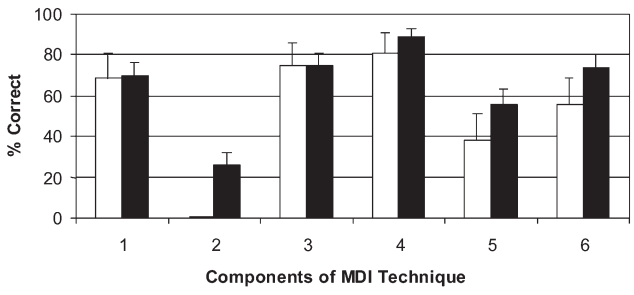


Figure 1. Proportion (%) of patients correctly performing each component of metered-dose inhaler (MDI) technique according to literacy level. Inadequate health literacy: *white bars* = yes; *black bars* = no. *1*: shaking the inhaler (p = 0.91); *2*: exhaling before actuation (p = 0.02); *3*: closing lips around mouthpiece (p = 0.97); *4*: pressing down once (p = 0.38); *5*: taking a full/deep breath without triggering "whistle" (p = 0.19); *6*: holding breath > 5 s (p = 0.18).

The data collection and educational intervention were conducted by a trained research assistant using standardized text and questionnaires. To minimize interviewer bias, the health literacy status of participants was not available to the interviewer at the follow-up visit. Also, electronic adherence data were collected after the interviewer-administered survey. Participants were informed that medication use was assessed, but were not told about precise monitoring abilities (i.e., that daily medication use was electronically measured).

Thus, at the discharge study visit, data regarding sociodemographics, asthma care, smoking history, morbidity, health literacy, asthma medication knowledge, and MDI technique were collected before any educational intervention. After the first and subsequent rounds of education, knowledge of the asthma discharge regimen and MDI technique were assessed. The education was repeated until participants could demonstrate mastery of the asthma discharge regimen and MDI technique. At the follow-up visit, knowledge of the asthma discharge regimen and MDI technique were tested again, adherence data for corticosteroid therapy were collected, and asthma symptom control was assessed.

Statistical Analyses

Descriptive statistics used means, medians, and proportions. Wilcoxon rank sum, matched pairs signed rank, or χ^2 tests were used in bivariate analyses. Logistic regression models were used to determine if inadequate health literacy was an independent predictor of the following outcomes: better asthma medication knowledge (yes [⩾ mean score] vs. no), better MDI technique (yes [⩾ mean score] vs. no), mastery of the discharge regimen after one round (yes vs. no), poor adherence to corticosteroid therapy (adherence < 50%: yes vs. no) (21), and better asthma symptom control (yes [⩾ mean score] vs. no). To determine if knowledge about medications mediated the relationship between inadequate health literacy and inferior self-management practices, we constructed additional models in which we included "better asthma medication knowledge" as a covariate in models to predict "better MDI technique" and "poor adherence to corticosteroid therapy."

Each regression model included the identical set of demographic and asthma severity indicators as independent variables: inadequate health literacy (yes vs. no; primary predictor of interest), age (quartiles), sex, ethnicity (African American vs. white), education level (high school graduate or equivalent degree; yes vs. no), income (annual combined household income ⩽ $19,999: yes [⩽ mean score] vs. no), history of near-fatal asthma (yes vs. no), asthma hospitalization in the prior 12 mo (yes vs. no), having a physician for asthma care (yes vs. no), and prior emergency department visits for asthma in the past 12 mo (yes vs. no). Results were similar when we analyzed outcomes as continuous or as binary variables; we presented the latter results to facilitate interpretation. To avoid overfitting the regression models, we elected to use a backward stepwise procedure in the multivariable logistic regression models to identify the most influential predictors for each of the outcomes ($p < 0.2$ for removal) (37). A two-tailed p value of less than 0.05 defined statistical significance. Computations were performed using STATA, version 7.0. (StataCorp, College Station, TX). Additional details regarding the methods are available in the online supplement.

RESULTS

Of 136 patients screened for eligibility, 36 met the following exclusion criteria: history of another chronic lung disease (n = 20), discharged to location other than home (n = 8), clinic patient of an investigator (n = 2), and contraindication to corticosteroids

(n = 6). Of 100 eligible patients, 8 patients were excluded because the physician declined consent and 19 were excluded because the patient declined consent.

Seventy-three patients (73% of eligible) were enrolled. Participants' mean age was 41 yr and they were mostly African American, female, unmarried, and employed. There was substantial evidence of previous asthma morbidity among participants, with more than three-quarters reporting at least one other emergency department visit for an asthma exacerbation in the previous 12 mo and nearly half with a history of near-fatal asthma. Fifty-six participants (77% of enrolled) returned for the 2-wk follow-up visit. Those not returning for the 2-wk visit were younger, had lower asthma medication knowledge scores, and tended to be African American. However, health literacy was similar, whether or not patients returned for the follow-up visit (Table 1).

Health Literacy

At the discharge study visit, 16 (22%) participants had inadequate health literacy. Sociodemographic characteristics (including age) were similar in patients with and without inadequate health literacy (Table 2). However, participants with inadequate health literacy were significantly more likely to have been hospitalized for asthma in the past 12 mo (81 vs. 52%, p = 0.04) and there was a trend toward a higher rate of near-fatal asthma in patients with inadequate health literacy (63 vs. 37%, p = 0.07).

Asthma Medication Knowledge Questionnaire

At the discharge study visit, the mean asthma medication knowledge score was 6.8 (of 10, SD = 2.1; Table 3). Participants with inadequate literacy had lower Asthma Medication Knowledge scores (odds ratio [OR] for better asthma medication knowledge score, 0.11; 95% confidence interval [CI], 0.03–0.42; p = 0.001; Table 4). Participants with inadequate health literacy were less likely to correctly answer every question, with significant differences noted for 5 of 10 items. In a multivariate logistic model, inadequate health literacy was the only significant independent predictor of asthma medication knowledge (OR for better asthma medication knowledge score, 0.08; 95% CI, 0.02–0.38; p = 0.002).

MDI Technique

The mean MDI technique score at the discharge study visit was 3.8 (of 6, SD=1.4). Failure to exhale before actuation of the MDI was the most frequently missed step. Pressing once on the canister at the time of actuation was the step that was most frequently done correctly. At the discharge study visit, subjects with inadequate health literacy were less likely to have superior MDI technique (OR for better MDI technique, 0.27; 95% CI, 0.08–0.87; p = 0.03). Having a history of near-fatal asthma was also associated with worse MDI technique at discharge study visit (OR for better MDI technique, 0.20; 95% CI, 0.1–0.7; p < 0.01; Table 4). In a multivariate model, inadequate health literacy was the only significant independent predictor of MDI technique (OR for better MDI technique, 0.29; 95% CI, 0.08–1.00; p = 0.05). When asthma medication knowledge was included as a covariate, inadequate health literacy was no longer a significant predictor of MDI technique (OR for better MDI technique, 0.53; 95% CI, 0.14–2.01), and asthma medication knowledge became the sole independent predictor of MDI technique (OR for better MDI technique, 1.48; 95% CI, 1.08–2.03).

After a single round of education, 59% (10/16) of subjects with inadequate health literacy and 73% (41/57) of subjects with higher health literacy exhibited mastery of MDI technique (p = 0.26). Overall, 21% (15) of subjects required one additional round of education, 10% (7) required two additional rounds of education, and no subject needed more than two additional rounds of education to exhibit mastery of MDI technique. Inadequate health literacy was not associated with the number of rounds of education needed to exhibit master of MDI technique (p = 0.79).

At the follow-up visit, the overall mean MDI technique score was 4.8 (SE = 0.14), which was significantly improved from the discharge study visit measurement (change = 1.1 [SE = 0.21], p < 0.001). There was a greater increase in the proportion of subjects achieving better MDI technique in patients with inadequate health literacy (p = 0.02; Figure 2), suggesting that this subgroup benefited most from the tailored education. At 2 wk, the proportion of participants with better MDI technique was no longer associated with health literacy (p = 0.57).

TABLE 1. PATIENT CHARACTERISTICS FOR SUBJECTS WITH AND WITHOUT 2-WK DATA

	All (*n = 73*)	With 2-wk Data (*n = 56*)	Without 2-wk Data (*n = 17*)	p Value‡
Sociodemographics				
Age, mean yr (SD, range)	40.9 (10.9)	43.0 (11.1)	33.9 (7.1)	< 0.01
Female	48 (66)	35 (63)	13 (76)	0.29
Race				0.09
African American	58 (79)	42 (75)	16 (94)	
White	15 (21)	14 (25)	1 (6)	
High school graduate or GED	44 (60)	32 (57)	12 (71)	0.32
Income ⩾ $19,999*	36 (65)	27 (63)	9 (75)	0.43
Asthma-related health care use				
Hospitalization past 12 mo†	42 (58)	35 (63)	7 (44)	0.18
ED visit past 12 mo†	55 (77)	43 (78)	12 (75)	0.79
Near-fatal asthma	31 (42)	23 (41)	8 (47)	0.66
Inadequate health literacy	16 (22)	12 (21)	4 (24)	0.85
Cigarette smoking history				0.43
Never smoker	32 (44)	26 (46)	6 (35)	
Past smoker	20 (27)	16 (29)	4 (24)	
Current smoker	21 (29)	14 (25)	7 (41)	
Physician for asthma care	37 (51)	30 (54)	7 (41)	0.37
Asthma knowledge score (SD)	6.9 (2.0)	7.2 (1.9)	6.2 (2.2)	0.07

Definition of abbreviations: ED = emergency department; GED = general equivalency diploma.
Values are given as numbers with percentages in parentheses unless otherwise noted.
* Limited data due to nonresponse (n = 55).
† Excludes current hospitalization/ED visit.
‡ Wilcoxon rank sum test or χ^2 test.

TABLE 2. PATIENT CHARACTERISTICS BY LEVEL OF HEALTH LITERACY

	Inadequate Health Literacy		
	Yes (*n* = *16*)	No (*n* = *57*)	p Value‡
Sociodemographics			
Age, mean yr (SD, range)	41.1 (12.9)	40.8 (10.5)	0.92
Female	9 (56)	39 (68)	0.37
Race			
African American	15 (94)	43 (75)	
White	1 (6)	14 (25)	0.11
High school graduate or GED	9 (56)	35 (61)	0.71
Income ⩽ $19,999*	8 (57)	28 (68)	0.45
Asthma-related health care use			
Hospitalization past 12 mo†	13 (81)	29 (52)	0.04
ED visit past 12 mo†	14 (88)	41 (75)	0.28
Near-fatal asthma	10 (63)	21 (37)	0.07
Cigarette smoking history			0.31
Never smoker	6 (38)	26 (46)	
Past smoker	3 (19)	17 (30)	
Current smoker	7 (44)	14 (25)	
Physician for asthma care	7 (44)	30 (53)	0.53

For definition of abbreviations, *see* Table 1.

Values are given as numbers with percentages in parentheses unless otherwise noted.

* Limited data due to nonresponse (n = 55).

† Excludes current hospitalization/ED visit.

‡ Wilcoxon rank sum test or χ^2 test.

Understanding of Discharge Regimen

After a single round of education, 69% (11/16) of subjects with inadequate health literacy and 68% (39/57) of subjects with higher health literacy were able to exhibit mastery of the discharge regimen. The need for supplemental education to exhibit mastery of the discharge regimen at the discharge study visit was not associated with health literacy in bivariate and multivariable analyses (p = 0.73 and p = 0.33, respectively). Overall, 17 (25%) subjects required one additional round of education, one subject required two additional rounds of education, one subject required three additional rounds of education, and no subject required more than three additional rounds of education to exhibit mastery of the discharge regimen. At the follow-up visit, the proportion of patients having mastery of the discharge regimen was also similar between groups based on health literacy (p = 0.35; Figure 2).

Adherence

Data on adherence to corticosteroid therapy and asthma symptom control were available in a substudy of 46 participants (63% of participants for the current study). About half (22/46, 48%) of the patients had poor adherence to corticosteroid therapy. Inadequate health literacy was not associated with poor adherence to corticosteroid therapy (OR, 0.89; 95% CI, 0.2–3.3; p = 0.86). In multivariable analyses, health literacy was not associated with adherence to corticosteroid therapy, whether (p = 0.45) or not (p = 0.45) the model included asthma medication knowledge.

Asthma Symptom Control

At the follow-up visit, asthma symptom control was similar in patients with and without inadequate health literacy (mean [95% CI]: 1.6 [0.95–2.31] vs. 1.5 [1.22–1.75]; p = 0.69). In multivariable analyses, inadequate health literacy was not associated with asthma symptom control at the follow-up visit (p = 0.84).

DISCUSSION

Deficiencies in asthma medication knowledge, MDI technique, and mastery of the discharge regimen were common in inner-city adults hospitalized for asthma exacerbations. Observed deficiencies in asthma self-management skills could not be explained by sociodemographic characteristics (age, sex, ethnicity, and education) or past asthma-related health care use but were independently associated with inadequate health literacy. Inadequate health literacy was associated with a greater likelihood of hospitalization for asthma exacerbations in the past 12 mo and significantly less knowledge of asthma medication and improper

TABLE 3. ASTHMA MEDICATION KNOWLEDGE ACCORDING TO LITERACY LEVEL

	Inadequate Health Literacy		
Question (correct answer)	Yes, % correct (*n* = *16*)	No, % correct (*n* = *57*)	p Value‡
Even if I have no symptoms from my asthma, I should take (say β_2-agonist they are on)* every day so that I can stop asthma attacks from starting (False)	10	60	< 0.01
If I am told to take two puffs, twice a day, then I should press down on the inhaler two times before I start to take a breath (False)	40	87	< 0.01
The longer I wait to treat an asthma attack after it begins, the easier it is to stop the attack (False)	75	96	< 0.01
When I am prescribed steroid tapers like prednisone for an asthma attack, I can stop taking them as soon as I start feeling well even if I have not finished the taper (False)	65	87	0.02
(Say inhaled steroid they are on)† should be taken only when I have asthma symptoms (False)	40	69	0.02
After I press down on the inhaler, I should take a normal, regular-sized breath (False)	25	49	0.06
(Say β_2-agonist they are on)* should be taken only when I have asthma symptoms (True)	60	74	0.21
(Say β_2-agonist they are on)* works within minutes to help you breathe better (True)	85	90	0.55
(Say inhaled steroid they are on)† takes days or weeks before it starts to help you breathe better (True)	50	55	0.68
Even if I have no symptoms from my asthma, I should take (say inhaled steroid they are on)† every day so that I can stop asthma attacks from starting (True)	70	73	0.78
Total test score, mean (95% CI)	5.2 (4.3–6.2)	7.2 (6.8–7.6)	< 0.01

Definition of abbreviation: CI = confidence interval.

Each question has a true, false, do not know format.

* If participant did not list a short-acting β_2-agonist as a medication on hospital admission medication, then the phrase "medicines such as Ventolin, Proventil, Maxair, Alupent, or Albuterol" was used.

† If the participant did not list an inhaled steroid as a medication on hospital admission, then the phrase "medicines like Flovent, Pulmicort, Azmacort, Vanceril, or Advair" was used.

‡ χ^2 test and Wilcoxon rank sum test.

TABLE 4. PREDICTORS OF BETTER ASTHMA MEDICATION KNOWLEDGE AT DISCHARGE STUDY VISIT, BETTER METERED-DOSE INHALER TECHNIQUE AT DISCHARGE STUDY VISIT, AND MASTERY OF THE DISCHARGE REGIMEN AFTER INITIAL ROUND OF INSTRUCTION (BIVARIATE ANALYSES)

	Better Asthma Medication Knowledge			Better MDI Technique			Mastery of Medication Discharge Regimen		
	OR	95% CI	P Value	OR	95% CI	p Value	OR	95% CI	P Value
Inadequate health literacy	0.1	0.0–0.4	< 0.01	0.3	0.1–0.9	0.03	0.7	0.2–2.6	0.64
Sociodemographics									
Age, quartiles	1.1	0.7–1.6	0.77	1.1	0.7–1.7	0.58	1.2	0.7–2.2	0.49
Female	1.1	0.4–2.9	0.85	1.7	0.6–4.4	0.32	1.0	0.3–2.9	1.00
White (Black is reference)	0.3	0.1–1.0	0.06	0.4	0.1–1.4	0.14	0.4	0.1–1.8	0.22
High school Graduate or GED	0.9	0.4–2.4	0.88	1.4	0.5–3.5	0.54	0.7	0.3–2.1	0.56
Asthma-related health care use									
Hospitalization past 12 mo*	1.3	0.5–3.3	0.60	0.6	0.2–1.8	0.36	0.7	0.3–2.1	0.58
ED visit past 12 mo*	2.1	0.7–6.4	0.20	1.0	0.9–1.0	0.23	0.8	0.2–2.8	0.71
Near-fatal asthma, ever	0.8	0.3–2.1	0.69	0.2	0.1–0.7	< 0.01	0.7	0.2–1.8	0.41
Cigarettes, smoked in past	0.6	0.2–1.8	0.37	0.8	0.3–2.5	0.68	0.6	0.2–2.0	0.42
Cigarettes, currently smoking	0.1	0.0–0.4	< 0.01	0.3	0.1–1.0	0.05	0.6	0.2–1.9	0.35
Asthma physician	2.5	1.1–5.7	0.03	1.6	0.6–4.2	0.30	1.0	0.4–2.8	0.97

Definition of abbreviations: CI = confidence interval; ED = emergency department; GED = graduate equivalency diploma; MDI = metered-dose inhaler; OR = odds ratio.

Asthma medication knowledge was dichotomized such that a score of ⩾ 7/10 was considered better knowledge; MDI technique at the discharge study visit was dichotomized such that a score of ⩾ 4/6 was considered better technique; mastery of the discharge regimen was dichotomized such that a score of 4/4 after initial instruction was considered mastery.

* Excludes current hospitalization/ED visit.

MDI technique. Patients with inadequate health literacy also were more likely to have lower asthma medication knowledge scores and improper MDI technique before the educational intervention. Surprisingly, inadequate health literacy was not associated with difficulty learning or retaining instructions about discharge regimen and proper MDI technique.

The link between inadequate health literacy and less knowledge of asthma medication and improper MDI technique reported here is consistent with prior reports (14, 15). A few studies in other disease models present evidence related to the role of health literacy in learning and retaining self-management skills in patients with other chronic diseases. For example, simplified materials have been shown to improve knowledge (38–44), health care use (45), and health behaviors (46, 47) for patients with diabetes mellitus, rheumatoid arthritis, obesity, and hypertension, irrespective of health literacy. In addition, subjects with low health literacy have been able to learn and retain self-management skills in small studies of heart failure, obesity, and osteoarthritis (48–50).

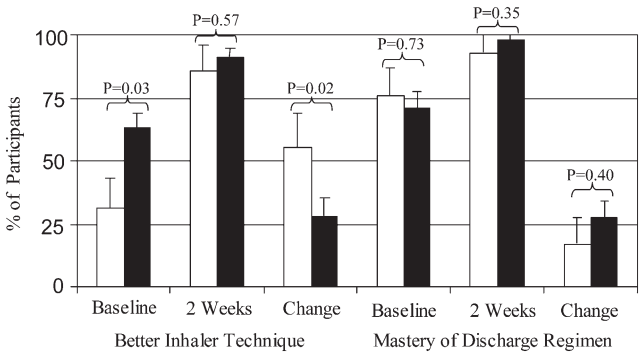


Figure 2. Better MDI technique score (% with ⩾ 4/6 correctly done) and mastery of discharge medication regimen (% exhibiting 4/4 questions correctly answered for all three medications in the standardized regimen) by literacy category at discharge study visit and at 2 wk. Inadequate health literacy: *white bars* = yes; *black bars* = no.

Results of our study, however, are the first to provide data in asthma and indicate that interventions using tailored education can successfully overcome barriers related to inadequate health literacy and improve asthma self-management skills. These observations are reasons for patients, clinicians, and policymakers to be optimistic about the benefits of providing tailored education to vulnerable patient populations. Additional studies, however, are needed to determine whether these short-term gains in self-management are retained at subsequent points in time or whether "refresher" courses are necessary. Sustained success with asthma self-management may, for example, be related to specific health beliefs, which may be related to health literacy (51, 52). Also, future research should include interventions focused on other guideline-recommended self-management skills (e.g., allergen avoidance and other triggers), which are critical to maintaining effective asthma control.

We provided intensive asthma medication education at hospital discharge, which included both oral and written instructions during a one-on-one training session until mastery was achieved ("teach to goal" strategy). Although this study cannot clarify which aspects of the education were critical to ensure learning and retention (combined use of oral and written instruction, one-on-one personalized training, teach-to-goal until mastery was achieved, exhibiting appropriate MDI technique), we speculate that the physical exhibition of MDI technique by study personnel and structured assessment of participants' understanding served to reinforce what was learned and overcome the barrier of inadequate health literacy (30). Because resources needed for this type of intensive inpatient education may not be routinely available, additional research is needed to identify key components of our multimodality intervention.

It is important to underscore the high prevalence of poor MDI technique and poor understanding of the discharge regimen we observed. Despite our intensive program to improve patients' self-management skills provided in this study, fully 28% did not understand the discharge regimen and 30% had not mastered

the MDI technique after one full round of education. These observations suggest that dramatic gains from the use of effective therapies in asthma are not reaching this high-risk population and highlight the need for evaluating patients' comprehension when providing discharge instructions and the development of systematic approaches to tailored inpatient asthma education.

Asthma self-management skills and use of acute medical services for asthma have been linked to various sociodemographic factors (18, 19, 53–56). In particular, fewer years of education completed has been implicated as an important risk factor in patients with asthma for cigarette smoking, worse MDI technique, and higher rates of emergency room visits, hospitalization, and mortality (19, 53–58). In this study, patients' level of education was not an independent predictor of asthma medication knowledge, MDI technique, or understanding of the discharge regimen. Instead, we found that health literacy was predictive of asthma medication knowledge and MDI technique. Moreover, our findings indicate that asthma medication knowledge plays a role in the causal pathway linking inadequate health literacy to poor MDI technique.

The prospective study design, uniform medication regimen after hospital discharge, and standardized education before discharge provided the unique opportunity to determine whether inadequate literacy served as a barrier to learning and retaining asthma self-management skills. Nonetheless, several potential limitations should be kept in mind. Although rates of follow-up were high (77%), bias may have been introduced due to incomplete follow-up, because not returning to the 2-wk follow-up visit was associated with age, race, and asthma medication knowledge. Also, without a control group, we cannot definitively conclude that our educational intervention was responsible for all the gains in asthma self-management that were noted. For example, it is possible but not certain that the lack of association we report between literacy and adherence was due to our educational intervention. Our data do, however, provide estimates of effect sizes that can be used to develop more definitive studies in this area (e.g., randomized clinical trial with an attention control group). Although interviewer bias is unlikely due to several safeguards put in place in our protocol, it is difficult to exclude completely the possibility that knowledge of health literacy status of participants influenced the teaching or assessment of self-management. Also, although we deliberately targeted an inner-city predominantly African-American population, a group at highest risk of complications resulting from asthma exacerbations, further research is needed to determine whether our findings are generalizable to other vulnerable patient populations (e.g., elderly with asthma).

Prior hospitalization strongly predicts risks of reexacerbations and death in the subsequent 12 mo (59–60). These poor outcomes are particularly common among African-American and inner-city patients for whom both hospitalizations and deaths related to asthma exacerbations are two to three times more likely than in other patient populations (64–66). These observations have led to a variety of inpatient programs to improve patients' self-management after hospital discharge (67, 68). The current study demonstrates that inpatient programs that embrace an intensive education program similar to ours are likely to overcome disparities in asthma self-management skills associated with inadequate health literacy.

In conclusion, this study presents evidence that inadequate asthma self-management skills are common and that inadequate health literacy is associated with worse asthma medication knowledge and MDI technique. However, inadequate health literacy was not associated with difficulty learning or retaining instructions about discharge regimen and proper MDI technique. Our results suggest the need for a systematic assessment of patient comprehension of discharge instructions and medications at hospital discharge. In patients with deficiencies in asthma self-management, we recommended a tailored, teach-to-goal, asthma education program.

Conflict of Interest Statement: None of the authors have a financial relationship with a commercial entity that has an interest in the subject matter of this manuscript.

References

1. Adams K, Corrigan JM. Institute of Medicine. Priority areas for national action: transforming health care quality. Committee on Identifying Priority Areas for Quality Improvement. Washington, DC: National Academies Press; 2003.
2. Carmona RH. Surgeon General of the United States: health literacy in America: the role of health care professionals. Prepared remarks given at the American Medical Association House of Delegates Meeting, June 14th, 2003. Available from: http://www.surgeongeneral.gov/news/speeches/ama061403.htm (accessed February 19, 2004).
3. Ratzan SC, Parker RM. Introduction. In: Selden CR, Zorn M, Ratzan SC, Parker RM, editors. National Library of Medicine current bibliographies in medicine: health literacy. Vol. NLM. Bethesda, MD: National Institutes of Health, U.S. Department of Health and Human Services; 2000. Publication No. CBM 2000-1. Available from: http://www.nlm.nih.gov/pubs/cbm/hliteracy.html (accessed January 15, 2004).
4. Nielsen-Bohlman LT, Panzer AM, Hamlin B, Kindig DA. Institute of Medicine. Health literacy: a prescription to end confusion. Committee on Health Literacy, Board on Neuroscience and Behavioral Health. Washington, DC: National Academies Press; 2004.
5. DeWalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. *J Gen Intern Med* 2004;19:1228–1239.
6. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. *J Gen Intern Med* 2005;20:175–184.
7. Weiss BD, Blanchard JS, McGee DL, Hart G, Warren B, Burgoon M, Smith KJ. Illiteracy among Medicaid recipients and its relationship to health care costs. *J Health Care Poor Underserved* 1994;5:99–111.
8. Rudd RE, Moeykens BA, Colton TC. Health and literacy: a review of medical and public health literature. In: Comings J, Garners B, Smith C, editors. Health and literacy, 1st ed. Available from: http://www.hsph.harvard.edu/healthliteracy/litreview_final.pdf (accessed September 9, 2005).
9. Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, Ren J, Peel J. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. *Am J Public Health* 2002;92:1278–1283.
10. Gazmararian JA, Baker DW, Williams MV, Parker RM, Scott TL, Green DC, Fehrenbach SN, Ren J, Koplan JP. Health literacy among Medicare enrollees in a managed care organization. *JAMA* 1999;281:545–551.
11. 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;25:337–344.
12. Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, Palacios J, Sullivan GD, Bindman AB. Association of health literacy with diabetes outcomes. *JAMA* 2002;288:475–482.
13. Mannino DM, Homa DM, Akinbami LJ, Moorman JE, Gwynn C, Redd SC. Surveillance for asthma: United States, 1980–1999. *MMWR Surveill Summ* 2002;51:1–13.
14. Williams MV, Baker DW, Honig EG, Lee TM, Nowlan A. Inadequate literacy is a barrier to asthma knowledge and self-care. *Chest* 1998;114:1008–1015.
15. Gazmararian JA, Williams MV, Peel J, Baker DW. Health literacy and knowledge of chronic disease. *Patient Educ Couns* 2003;51:267–275.
16. Williams SG, Schmidt DK, Redd SC, Storms W. Key clinical activities for quality asthma care. Recommendations of the National Asthma Education and Prevention Program. *MMWR Recomm Rep* 2003;52:1–8.
17. Berkman ND, DeWalt DA, Pignone MP, Sheridan SL, Lohr KN, Lux L, Sutton SF, Swinson T, Bonito AJ. Literacy and health outcomes. Evidence Report/Technology Assessment No. 87 (prepared by RTI International–University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016). January. Rockville, MD: Agency for Healthcare Research and Quality; 2004. AHRQ Publication No. 04-E007-2.
18. Boudreaux ED, Emond SD, Clark S, Camargo CA Jr. Acute asthma

among adults presenting to the emergency department: the role of race/ethnicity and socioeconomic status. *Chest* 2003;124:803–812.
19. Grant EN, Lyttle CS, Weiss KB. The relation of socioeconomic factors and racial/ethnic differences in US asthma mortality. *Am J Public Health* 2000;90:1923–1925.
20. Kirsch I, Jungeblut A, Jenkins L, Kolstad A. Adult literacy in America: a first look at the findings of the national adult literacy survey. Washington, DC: National Center for Education Statistics, U.S. Department of Education; 1993.
21. Krishnan JA, Riekert KA, McCoy JV, Stewart DY, Schmidt S, Chanmugam A, Hill P, Rand CS. Corticosteroid use after hospital discharge among high-risk adults with asthma. *Am J Respir Crit Care Med* 2004;170: 1281–1285.
22. Paasche-Orlow MK, Brancati FL, Rand CS, Krishnan JA. Education of patients with asthma and low literacy. *J Gen Intern Med* 2003; 18(Suppl):S227.
23. Nurss JR, Parker RM, Williams MV, Baker DW. Test of functional literacy in adults, 2nd ed. Snow Camp, NC: Peppercorn Books and Press; 2001.
24. Nurss JR, Parker RM, Williams MV, Baker DW. Test of functional literacy in adults. Greensboro, NC: Peppercorn Books and Press; 1995. Available from: http://www.peppercornbooks.com/test_of_functional_health_literacy_in_adults.asp (accessed February 24, 2004).
25. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. *Patient Educ Couns* 1999;38:33–42.
26. Ho J, Bender BG, Gavin LA, O'Connor SL, Wamboldt MZ, Wamboldt FS. Relations among asthma knowledge, treatment adherence, and outcome. *J Allergy Clin Immunol* 2003;111:498–502.
27. Barry PW, O'Callaghan C. Inhalational drug delivery from seven different spacer devices. *Thorax* 1996;51:835–840.
28. Lee H, Evans HE. Evaluation of inhalation aids of metered dose inhalers in asthmatic children. *Chest* 1987;91:366–369.
29. Interiano B, Guntupalli KK. Metered-dose inhalers: do health care providers know what to teach? *Arch Intern Med* 1993;153:81–85.
30. Schillinger D, Piette J, Grumbach K, Wang F, Wilson C, Daher C, Leong-Grotz K, Castro C, Bindman AB. Closing the loop: physician communication with diabetic patients who have low health literacy. *Arch Intern Med* 2003;163:83–90.
31. Juniper EF, O'Byrne PM, Guyatt GH, Ferrie PJ, King DR. Development and validation of a questionnaire to measure asthma control. *Eur Respir J* 1999;14:902–907.
32. Julius SM, Sherman JM, Hendeles L. Accuracy of three electronic monitors for metered-dose inhalers. *Chest* 2002;121:871–876.
33. Simmons MS, Nides MA, Kleerup EC, Chapman KR, Milgrom H, Rand CS, Spector SL, Tashkin DP. Validation of the Doser, a new device for monitoring metered-dose inhaler use. *J Allergy Clin Immunol* 1998; 102:409–413.
34. George CF, Peveler RC, Heiliger S, Thompson C. Compliance with tricyclic antidepressants: the value of four different methods of assessment. *Br J Clin Pharmacol* 2000;50:166–171.
35. Liu H, Golin CE, Miller LG, Hays RD, Beck CK, Sanandaji S, Christian J, Maldonado T, Duran D, Kaplan AH, *et al.* A comparison study of multiple measures of adherence to HIV protease inhibitors. *Ann Intern Med* 2001;134:968–977.
36. Cramer JA, Mattson RH, Prevey ML, Scheyer RD, Ouellette VL. How often is medication taken as prescribed? A novel assessment technique. *JAMA* 1989;261:3273–3277.
37. Harrell FE Jr, Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep* 1985;69:1071–1077.
38. Eaton ML, Holloway RL. Patient comprehension of written drug information. *Am J Hosp Pharm* 1980;37:240–243.
39. Meade CD, Byrd JC, Lee M. Improving patient comprehension of literature on smoking. *Am J Public Health* 1989;79:1411–1412.
40. Meade CD, McKinney WP, Barnas GP. Educating patients with limited literacy skills: the effectiveness of printed and videotaped materials about colon cancer. *Am J Public Health* 1994;84:119–121.
41. Michielutte R, Bahnson J, Dignan MB, Schroeder EM. The use of illustrations and narrative text style to improve readability of a health education brochure. *J Cancer Educ* 1992;7:251–260.
42. Murphy PW, Chesson AL, Walker L, Arnold CL, Chesson LM. Comparing the effectiveness of video and written material for improving knowledge among sleep disorders clinic patients with limited literacy skills. *South Med J* 2000;93:297–304.
43. Wydra EW. The effectiveness of a self-care management interactive multimedia module. *Oncol Nurs Forum* 2001;28:1399–1407.
44. Hayes KS. Randomized trial of geragogy-based medication instruction in the emergency department. *Nurs Res* 1998;47:211–218.
45. Davis TC, Berkel HJ, Arnold CL, Nandy I, Jackson RH, Murphy PW. Intervention to increase mammography utilization in a public hospital. *J Gen Intern Med* 1998;13:230–233.
46. Howard-Pitney B, Winkleby MA, Albright CL, Bruce B, Fortmann SP. The Stanford Nutrition Action Program: a dietary fat intervention for low-literacy adults. *Am J Public Health* 1997;87:1971–1976.
47. Hartman TJ, McCarthy PR, Park RJ, Schuster E, Kushi LH. Results of a community-based low-literacy nutrition education program. *J Community Health* 1997;22:325–341.
48. DeWalt DA, Pignone M, Malone RM, Bryant B, Felix K, Corr K, Kosnar MC, Rawls C, Rogers N, Rothman R, *et al.* Randomized controlled trial of a primary care-based heart failure disease management program for patients with low literacy. *J Gen Intern Med* 2004;19(Suppl): S203.
49. Gans KM, Lovell HJ, Fortunet R, Lasater TM. Low-literacy audio intervention for lowering fat intake. *J Nutr Educ* 1998;30:410B–410B1.
50. Bill-Harvey D, Rippey R, Abeles M, Donald MJ, Downing D, Ingenito F, Pfeiffer CA. Outcome of an osteoarthritis education program for low-literacy patients taught by indigenous instructors. *Patient Educ Couns* 1989;13:133–142.
51. Bhat BR, Friedman S, Adimoolam S, Schneider AT, Chiaramonte LT. Study of social, educational, environmental and cultural aspects of childhood asthma in clinic and private patients in the city of New York. *Ann Allergy* 1978;41:89–92.
52. George M, Freedman TG, Norfleet AL, Feldman HI, Apter AJ. Qualitative research-enhanced understanding of patients' beliefs: results of focus groups with low-income, urban, African American adults with asthma. *J Allergy Clin Immunol* 2003;111:967–973.
53. Huovinen E, Kaprio J, Laitinen LA, Koskenvuo M. Social predictors of adult asthma: a co-twin case-control study. *Thorax* 2001;56:234–236.
54. Adams RJ, Smith BJ, Ruffin RE. Factors associated with hospital admissions and repeat emergency department visits for adults with asthma. *Thorax* 2000;55:566–573.
55. Radeos MS, Leak LV, Lugo BP, Hanrahan JP, Clark S, Camargo CA Jr. Risk factors for lack of asthma self-management knowledge among ED patients not on inhaled steroids. *Am J Emerg Med* 2001;19:253–259.
56. Diette GB, Krishnan JA, Dominici F, Haponik E, Skinner EA, Steinwachs D, Wu AW. Asthma in older patients: factors associated with hospitalization. *Arch Intern Med* 2002;162:1123–1132.
57. Silverman RA, Boudreaux ED, Woodruff PG, Clark S, Camargo CA Jr. Cigarette smoking among asthmatic adults presenting to 64 emergency departments. *Chest* 2003;123:1472–1479.
58. Eisner MD, Yelin EH, Katz PP, Shiboski SC, Henke J, Blanc PD. Predictors of cigarette smoking and smoking cessation among adults with asthma. *Am J Public Health* 2000;90:1307–1311.
59. Weiss KB, Sullivan SD, Lyttle CS. Trends in the cost of illness for asthma in the United States, 1985–1994. *J Allergy Clin Immunol* 2000;106:493–499.
60. Flora GS, Sharma AM, Sharma OP. Asthma mortality in a metropolitan county hospital, a 38-year study. *Allergy Proc* 1991;12:169–179.
61. Miller TP, Greenberger PA, Patterson R. The diagnosis of potentially fatal asthma in hospitalized adults: patient characteristics and increased severity of asthma. *Chest* 1992;102:515–518.
62. Crane J, Pearce N, Burgess C, Woodman K, Robson B, Beasley R. Markers of risk of asthma death or readmission in the 12 months following a hospital admission for asthma. *Int J Epidemiol* 1992;21: 737–744.
63. McCarren M, McDermott MF, Zalenski RJ, Jovanovic B, Marder D, Murphy DG, Kampe LM, Misiewicz VM, Rydman RJ. Prediction of relapse within eight weeks after an acute asthma exacerbation in adults. *J Clin Epidemiol* 1998;51:107–118.
64. Thomas SD, Whitman S. Asthma hospitalizations and mortality in Chicago: an epidemiologic overview. *Chest* 1999;116:135S–141S.
65. Ray NF, Thamer M, Fadillioglu B, Gergen PJ. Race, income, urbanicity, and asthma hospitalization in California: a small area analysis. *Chest* 1998;113:1277–1284.
66. De Palo VA, Mayo PH, Friedman P, Rosen MJ. Demographic influences on asthma hospital admission rates in New York City. *Chest* 1994;106: 447–451.
67. Evans R III, LeBailly S, Gordon KK, Sawyer A, Christoffel KK, Pearce B. Restructuring asthma care in a hospital setting to improve outcomes. *Chest* 1999;116:210S–216S.
68. George MR, O'Dowd LC, Martin I, Lindell KO, Whitney F, Jones M, Ramondo T, Walsh L, Grissinger J, Hansen-Flaschen J, *et al.* A comprehensive educational program improves clinical outcome measures in inner-city patients with asthma. *Arch Intern Med* 1999;159:1710–1716.

0145-6008/05/2903-0389$03.00/0
Alcoholism: Clinical and Experimental Research

Vol. 29, No. 3
March 2005

Substance Abuse Treatment and Hospitalization among a Cohort of HIV-Infected Individuals with Alcohol Problems

Anita Palepu, Nicholas J. Horton, Nicole Tibbetts, Seville Meli, and Jeffrey H. Samet

Background: We examined the association of substance abuse treatment services on hospitalization among participants in the HIV-Alcohol Longitudinal Cohort (HIV-ALC) study of HIV-infected individuals with a history of alcohol problems.

Methods: A standardized questionnaire that inquired about demographics, substance use, use of substance abuse treatment services, and hospitalization was administered to 349 HIV-ALC participants. We defined substance abuse treatment services as any of the following in the past 6 months: 12 weeks in a half-way house or residential facility, 12 visits to a substance abuse counselor or mental health professional, or participation in any methadone maintenance program.

Results: Almost one third of this cohort were hospitalized in the past 6 months. Substance abuse treatment was not significantly associated with hospitalization [adjusted odds ratio (AOR) 1.0; 95% confidence interval (CI) 0.7–1.5), whereas homelessness (AOR 2.3; 95% CI 1.5–3.6), injection drug use (AOR 1.7; 95% CI 1.0–2.7), severity of alcohol dependence (AOR 1.02; 95% CI 1.00–1.05), CD4 cell count (AOR 0.999; 95% CI 0.998–1.00), and HIV RNA (AOR 1.1; 95% CI 1.0–1.2) were independently associated with increased odds of hospitalization over time.

Conclusions: Engagement in substance abuse treatment was not associated with a decrease in hospital use by HIV-infected individuals with a history of alcohol problems. The period of substance abuse treatment may present an opportunity to address health care utilization patterns of HIV-infected individuals.

Key Words: Health Services Utilization, Substance Use, Alcohol Use, HIV Disease/AIDS, Hospitalizations.

LESS THAN OPTIMAL patterns of health care utilization have been observed among certain HIV-infected individuals (Shapiro et al., 1999). In particular, HIV-infected drug users have frequent hospitalization (Laine et al., 2001; Palepu et al., 2001; Stein, 1994) with high associated costs (Palepu et al., 2001; Stein and Sobota, 2001). Direct complications of illicit drug use (Palepu et al., 2001; Stein and Sobota, 2001) as well as indirect consequences such as a reduced ability to adhere to antiretroviral therapy (Lucas et al., 2001) account for a substantial portion of the health care utilization by HIV-infected injection drug users. Patients with alcohol problems have also been found to have less than optimal patterns of health care utilization (Hunkeler et al., 2001; Saitz et al., 2000; Weintraub et al., 2001). In addition, it is not uncommon to encounter alcohol problems among HIV-infected individuals (Conigliaro et al., 2003; Lefevre et al., 1995; Petry, 1999; Samet et al., 2004). Laine et al. (2001) used New York State Medicaid files to examine the relationship of outpatient medical and substance abuse care with subsequent hospitalization rates among drug users. Subsequent hospitalization was reduced for both HIV-infected and HIV-negative drug users who had regular medical and drug abuse care. Little is known, however, about the relationship of substance abuse treatment services and the hospital use of HIV-infected individuals with alcohol problems. As substance abuse treatment services can also provide an opportunity to address one's behaviors and medical needs beyond the use of substances

From the Centre for Health Evaluation and Outcome Sciences (AP), St. Paul's Hospital, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; Department of Mathematics (NJH), Smith College, Northampton, Massachusetts; DM-STAT Inc. (NT), Medford, Massachusetts; and Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Departments of Medicine (SM, JHS) and Social and Behavioral Sciences (JHS), Boston University Schools of Medicine and Public Health, Boston, Massachusetts.

Received for publication April 9, 2004; accepted November 20, 2004.

Support for this study came from the following grants from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) of the National Institutes of Health: 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, US Public Health Service Grant M01 RR00533. AP is supported by the Canadian Institute for Health Research through a New Investigator Award and a Senior Scholar Award from the Michael Smith Foundation for Health Services.

Reprint requests: Anita Palepu, MD, MPH, FRCPC, Division of General Internal Medicine, University of British Columbia, St. Paul's Hospital, 620B-1081 Burrard Street, Vancouver, BC V6Z 1 Y6, Canada; Fax: 604-806-8005; E-mail: anita@hivnet.ubc.ca.

DOI: 10.1097/01.ALC.0000156101.84780.45

that affect one's overall health, we hypothesized that engagement in substance abuse treatment would demonstrate a protective association with hospitalization. We examined this among participants in the HIV-Alcohol Longitudinal Cohort (HIV-ALC) study, composed of HIV-infected individuals with a history of alcohol problems.

MATERIALS AND METHODS

Study Population

Patients who had HIV infection and a history of alcohol problems were identified by explicit eligibility criteria. All potential participants who gave two or more positive responses to the CAGE questionnaire (Buchsbaum et al., 1991; Ewing, 1984; Fiellin et al., 2000), a screening test for lifetime alcohol problems (sensitivity ~80%, specificity ~90%), were eligible. In addition, patients who were recruited from the Boston Medical Center HIV Diagnostic Evaluation Unit (Samet et al., 1995) and did not meet CAGE criteria were eligible when one of two attending physicians made a specific diagnosis of lifetime alcohol abuse or dependence. Thus, individuals with a history of lifetime alcohol problems, despite not being detected by the CAGE, were detected by the clinical interview and recruited. Other entry criteria included the following: fluency in English or Spanish, Mini-Mental State Examination score ≥21 (Folstein et al., 1975), and no plans to move from the Boston area in the next 2 years. As chronic alcohol use is associated with cognitive impairment, we used the Mini-Mental State Examination cutoff of 21 to exclude individuals in whom such impairment may preclude obtaining informed consent, an accurate and complete interviewer-administered questionnaire, or a follow-up interview. The Institutional Review Boards of Boston Medical Center and Beth Israel Deaconess Medical Center approved this study.

From July 1997 through July 2001, recruitment of participants occurred by multiple methods and from several sites: Boston Medical Center HIV Diagnostic Evaluation Unit, 56%; posted fliers, 17%; Boston Medical Center Primary Care Clinic, 13%; respite facility for homeless people, 5%; methadone clinic, 4%; subject referrals, 4%; and Beth Israel Deaconess Medical Center, 2%. The majority of participants were recruited from medical settings that addressed HIV-related issues. The eligibility criteria of a history of alcohol problems was determined by the CAGE questionnaire in 313 (90%) of 349 participants and based on clinical assessment in 36 (10%) of 349 participants. Diagnostic interviews for alcohol problems in a sample of these participants ($n = 238$) revealed a lifetime history of alcohol dependence [76% (181 of 238)] or abuse [15% (37 of 238)] in >90% (Samet et al., 2004).

Data Collection

After obtaining informed consent, a research associate interviewed participants using a standardized instrument to ascertain baseline information including the following: demographics, HIV risk behaviors, alcohol severity, use of substance abuse treatment services, medications used, and health care utilization in the preceding 6 months. For the Spanish interview, standardized 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. We attempted to obtain CD4 cell counts and HIV RNA (viral load) levels on all participants. Laboratory tests performed within 6 months of the interview as part of clinical care were recorded. When not available through routine clinical care, blood samples were obtained and tested for CD4 cell count and HIV RNA using the Boston Medical Center Clinical Laboratory. Participants were followed every 6 months for up to seven observations, and follow-up ended July 2001.

Outcome and Primary Independent Variables

Hospital admission in the previous 6-month period was our outcome of interest. We defined substance abuse treatment services as any of the following in the past 6 months: 12 weeks in a half-way house or residential facility, 12 visits to a substance abuse counselor or mental health professional, or participation in any methadone maintenance program (Brands et al., 2002; Farre et al., 2002).

Other Independent Variables

Other specific variables assessed included sex, age, race/ethnicity (black, white, or other, which comprised mostly Hispanic participants), injection drug use in the previous 6 months, abstinence from alcohol in the past 30 days, severity of alcohol dependence as measured by the Alcohol Dependence Scale (Ross et al., 1990), incarceration in the previous 6 months, CD4 cell count, HIV RNA, and homelessness. Homelessness was defined as having spent at least 1 night either on the street or in a shelter in the 6 months before the interview (Kertesz et al., 2003). We also assessed depressive symptoms as measured by the 20-item Centers for Epidemiologic Studies Depression Scale, in which ≥16 denotes severe depressive symptoms (Andresen et al., 1994). We also assessed physical functioning and mental health functioning by the physical component score and mental component score derived from the Short Form-36 (Ware, 1993, 1994). Medical diagnoses were read to the participants, and they indicated whether they had been hospitalized for these reasons in the previous 6 months. We examined the distribution of these medical diagnoses among those who reported hospitalization at baseline to assess whether the admissions were potentially HIV related or consequences of substance abuse. Measurement of HIV RNA was performed using branched-chain DNA techniques (Pachl et al., 1995).

Analysis

In the bivariable analysis, categorical variables were compared using χ^2 test, and continuous variables were compared using the Student's t test. Multivariable longitudinal logistic regression models were constructed to examine the association of substance abuse treatment on hospitalization over time. Because serial measures on the same individuals were considered, generalized estimating equations were used to adjust for correlation between these measures over time using an independence working correlation matrix (Liang and Zeger, 1986; Zeger and Liang, 1986). The substance abuse treatment variable was dichotomous. We adjusted for potential confounding factors by including in the model sex, race, age, severity of alcohol dependence, homelessness, recent drug injection, incarceration in the past 6 months, current use of highly active antiretroviral therapy, abstinence from alcohol in the past 30 days, baseline CD4 cell count, and HIV RNA. These were factors that we hypothesized would affect the odds of hospitalization. All analyses were carried out using SAS/STAT version 8.2 (SAS Institute, Cary, NC).

RESULTS

Baseline participant characteristics are described in Table 1: 79% were men, two thirds were ethnic minorities, mean age was 41 years, and 29% were homeless. The most common HIV risk factor was injection drug use (59%), with men having sex with men and heterosexual sex each stated by 20% of participants. Of the injection drug users, 23% had injected drugs in the previous 6 months. In the past 30 days, 24% of participants reported using alcohol and heroin or cocaine, 18% used alcohol alone, and 5% used heroin or cocaine alone; 12% were enrolled in methadone maintenance. Among the participants who drank alcohol in the past 30 days, the average daily alcohol consumption was 6.4 drinks. One third (118 of 349) of the participants were engaged in substance abuse treatment at the initial observation, and the specific treatment services and average exposure are presented in Table 1. More than half (59%) of

Table 1. Characteristics of Participants in the HIV-ALC Cohort, HIV-Infected Individuals with a History of Alcohol Problems ($N = 349$)

Characteristics	*n* (%)
Mean age [years (SD)]	40.6 (7)
Female	73 (21)
Ethnicity	
Black	154 (44)
White	116 (33)
Other	79 (23)
Homeless	101 (29)
Jail[a]	102 (29)
Mean days in jail (SD)	90 (66)
Drug and alcohol use in the past 30 days	
Alcohol and heroin or cocaine	84 (24)
Heroin or cocaine only	17 (5)
Alcohol only	64 (18)
None	184 (53)
Drank to intoxication >3×/week for >1 year	333 (96)
Drank alcohol in the past 30 days	148 (42)
Alcohol consumption [mean drinks/day (SD)]	6.4 (15)
Drank to intoxication in the past 30 days	122 (35)
No. of times [mean (SD)]	9.9 (9)
Inject drug[a]	81 (23)
Mean ADS score (SD)	12.4 (10)
Substance abuse treatment[b]	118 (34)
Residential facility (≥90 days)	26 (8)
Mean days (SD)	132.4 (32.4)
Day treatment program (≥30 days)	8 (2)
Methadone maintenance[a]	40 (11.5)
Substance abuse counselor (≥12 visits)[a]	56 (16)
Mean no. of visits (SD)	22.4 (7.2)
Mental health treatment (≥12 visits)[a]	40 (12)
Mean no. of visits (SD)	24.6 (14.6)
Mean CD4 cell count (SD)	401 (278)
Mean HIV RNA ($\log_{10}$)	2.7 (1.9)
Receiving HAART	249 (59)
Hospitalization[a]	107 (30)
Psychiatric medications[a]	
Antianxiety	52 (15)
Antidepressant	158 (45)
Antipsychotic	53 (15)
CNS stimulant or miscellaneous	17 (5)
Sedative/hypnotic	14 (4)

ADS, Alcohol Dependence Scale; HAART, highly active antiretroviral treatment.

[a] In the past 6 months.

[b] Substance abuse treatment refers to any of the following in the past 6 months: 12 weeks in a half-way house or residential facility, 12 visits to a substance abuse counselor or mental health professional, or participation in any methadone maintenance program.

the participants reported that they had received antiretroviral therapy. Among the participants who were not engaged in substance abuse treatment at the initial observation, 70 entered substance abuse treatment during the study period. In terms of primary care, 92% (323 of 349) saw a physician two or more times in the preceding 6 months. Of all study participants at the initial observation ($n = 349$), the number of participants who reported being on psychiatric medications in the previous 6 months were as follows: antianxiety ($n = 52$; 15%), antidepressant ($n = 158$; 45%), antipsychotic ($n = 53$; 15%), central nervous system stimulant or miscellaneous ($n = 17$; 5%), and sedative/hypnotic ($n = 14$; 4%). In this research study, the participants were followed every 6 months for up to seven occasions, and the median number of observations per subject was three. Time of recruitment was the strongest predictor of number of follow-up observations because of the design of the investigation (study participants were recruited over a 4-year period, and all follow-up interviews ceased in July 2001).

Table 2. Baseline Bivariate Associations of Participants Who Were Hospitalized in the Preceding 6 Months

Characteristics	Yes ($N = 106$)	No ($N = 243$)	*p* value
Mean age (SD)	42.1 (7.7)	39.9 (7.1)	0.01
Female	24 (23)	49 (20)	0.60
Ethnicity			
Black	59 (56)	95 (39)	0.01
White	31 (29)	85 (35)	
Other	16 (15)	63 (26)	
Mean PCS (SD)	39 (11)	47 (11)	0.0001
Mean MCS (SD)	37 (15)	39 (16)	0.31
Depressive symptoms ≥16[a]	79 (75)	153 (63)	0.03
Abstinent from EtOH	68 (64)	133 (55)	0.10
Homeless	39 (37)	62 (26)	0.03
Jail[b]	24 (23)	78 (32)	0.07
Inject drugs[b]	28 (26)	53 (22)	0.35
Mean ADS score[c] (SD)	11.7 (11.4)	6.6 (9.1)	0.0001
Health insurance[d]	105 (99)	230 (99)	0.94
Substance abuse treatment	45 (42)	73 (30)	0.02
Receiving HAART	62 (58)	143 (59)	0.95
Self-help attendance	71 (67)	147 (61)	0.27
Mean CD4 cell count (SD)[e]	368 (322)	414 (256)	0.21
Mean HIV RNA ($\log_{10}$)[e]	3.06 (1.9)	2.59 (1.9)	0.04

PCS, physical component scale; MCS, mental component scale; EtOH, ethanol.

[a] Centers for Epidemiologic Studies Depression Scale score ≥16 indicating severe depressive symptoms.

[b] In past 6 months.

[c] Possible range of ADS score: 1–44.

[d] Fully insured versus partly or none.

[e] Data are missing for 11 participants.

The characteristics of hospitalized participants are shown in Table 2. They were more likely to be black, be homeless, and have higher alcohol dependence scores. Hospitalized participants were also more likely to have received substance abuse treatment in the past 6 months (42 vs. 30%; $p = 0.02$). The results of the multivariable logistic regression of factors associated with hospitalization are presented in Table 3. Substance abuse treatment was not significantly associated with hospitalization [adjusted odds ratio (AOR) 1.0; 95% confidence interval (CI) 0.7–1.5], whereas homelessness (AOR 2.3; 95% CI 1.5–3.6), injection drug use (AOR 1.7; 95% CI 1.0–2.7), severity of alcohol dependence (AOR 1.02; 95% CI 1.00–1.05), CD4 cell count (AOR 0.999; 95% CI 0.998–1.00), and HIV RNA (AOR 1.1; 95% CI 1.0–1.2), were independently associated with increased odds of hospitalization over time. The frequency of episodic illnesses in the past 6 months among those who reported hospitalization is shown in Table 4. The conditions most frequently cited were gastrointestinal-related conditions (peptic disease, abdominal pain, and hematemesis) followed by AIDS opportunistic infections (*Pneumocystis carinii*, esophageal candidiasis, and tuberculosis), traumatic injuries, and infections. Overall, a majority of the episodic illnesses were not HIV related.

DISCUSSION

In this sample of HIV-infected individuals with alcohol problems, engagement in substance abuse treatment ser-

Table 3. Multivariable Logistic Regression of Factors Associated with Hospital Admission[a]

Factor	AOR	95% CI
Substance abuse treatment[b]	1.0	0.7–1.5
Female	1.5	0.9–2.5
Race[c]		
Black	1.8	0.9–3.2
White	1.0	0.6–1.9
ADS	1.02	1.01–1.05
Homeless[d]	2.3	1.5–3.6
Injection drug use[d]	1.7	1.02–2.7
Jail[d]	0.9	0.4–1.1
Abstinent from EtOH	1.3	0.9–1.9
Age (per year)	1.02	0.99–1.06
HAART	1.3	0.8–1.9
CD4 cell count[e]	0.999	0.998–1.00
HIV RNA ($\log_{10}$)[f]	1.11	1.0–1.2

[a] Using generalized estimating equations and controlling for time.
[b] Having at least 12 weeks in a half-way house or residential facility, 12 visits to a substance abuse counselor or mental health professional, or participation in any methadone maintenance program in the previous 6 months.
[c] Reference group is other.
[d] In the past 6 months.
[e] $p = 0.04$.
[f] $p = 0.03$.

Table 4. Self-Reported Episodic Medical Illnesses in the Previous 6 Months of Participants Who Have HIV and Alcohol Problems and Have Been Hospitalized ($N = 107$)

Episodic medical illnesses	n[a]
Peptic ulcer disease, abdominal pain, and hematemesis	41
AIDS opportunistic infection[b]	33
Traumatic injuries[c]	28
Bacterial pneumonia	20
Low back pain >3 months	19
Skin infection (cellulitis or abscess)	17
Hepatitis or jaundice	14
Seizures	13
Drug or alcohol overdose	11
Pancreatitis	7
Septic arthritis or osteomyelitis	6
Chest pain secondary to cocaine use	4
Endocarditis	3
Venous thromboembolism	2

[a] Categories are not mutually exclusive.
[b] *Pneumocystis carinii*, esophageal candidiasis, and tuberculosis.
[c] Gunshot or stab wound.

vices was not associated with hospitalization compared with participants who did not use these services. Many participants who were hospitalized had episodic medical illnesses that may have been related to ongoing substance abuse. In contrast, Laine et al. (2001) found that among 11,566 HIV-infected drug users, the AOR for hospitalization was lowest among those with both regular medical and drug abuse care (0.76) followed by those with regular medical care alone (0.82) and regular drug abuse care alone (0.85) compared with those with neither. They used administrative data to derive their regular drug abuse care variable and stringently defined it as at least three weekly claims per month for drug abuse treatment over a 6-month period. In our cohort, only one third were engaged in substance abuse services and on average at a lower intensity than the criteria used by Laine et al.

The association of substance abuse and health care utilization has been documented in numerous studies (Alexandre et al., 2001; French et al., 2000; Knowlton et al., 2001; McGeary and French, 2000; Palepu et al., 2001). In addition, the substantial health care utilization of HIV-infected individuals has been well documented (Bozzette et al., 1998; Gifford et al., 2000; Shapiro et al., 1999; Smith et al., 2000). In the HIV Cost and Service Utilization Study, ~20% of the patients reported being hospitalized in the previous 6 months (Bozzette et al., 1998), which is lower than participants in our study (44%). The prevalence of alcohol use among HIV-infected individuals (Samet et al., 2004) and its potential negative impact on behavioral and clinical issues increase the importance of understanding the predictors of health care use in a population of HIV-infected individuals with alcohol problems (Conigliaro et al., 2003).

The high use of primary care in this cohort, a result of the strategy by which recruitment occurred, allows for an examination of an effect of substance abuse treatment in addition to primary medical care for decreasing utilization of hospital resources. The absence of any association of substance abuse treatment on hospitalization was not what we had hypothesized and not consistent with the work of Laine et al. (2001). Several reasons may account for the absence of this effect. The effect of substance abuse treatment on hospital utilization in HIV-infected patients with a history of alcohol problems may be mild, if at all present. Alternatively, the effect is potentially substantial, but services as defined in this study may not have been sufficiently intense or adequately integrated with their medical care to have an association with hospital utilization. Special efforts may be required to appreciate the potential synergistic effect of primary medical care and substance abuse treatment services on utilization of hospital services. Despite having primary care (Samet et al., 2001), these patients may not be able to attend all set appointments because of competing priorities as a result of substance abuse (French et al., 2000; Gifford et al., 2000) and present later in the course of an illness, thus requiring hospitalization (Aday, 1994; French et al., 2000; Palepu et al., 2001; Samet et al., 1998; Weber et al., 2000). Substance abuse treatment programs can stress the critical nature of ongoing medical care for HIV-infected patients. A focus on dimensions of medical care that can include optimal medication use, safer behaviors regarding sex, and drug-related HIV transmission may result in fewer medical complications and fewer hospitalizations. Similarly, primary care providers can facilitate their patients' entry into substance abuse treatment (Strathdee et al., 1999) and play a crucial role in the long-term management of patients in recovery from alcohol or other drug problems (Friedmann et al., 1998; Stein and Friedmann, 2001). It is possible that only after the establishment of such integration of substance abuse and primary care services that optimization of the utilization of hospital services will be achieved.

Our study has several limitations. Most study participants were already engaged in primary care as that setting was a

major site for recruitment. This selection bias may overestimate the degree of health care utilization among HIV-infected urban individuals with alcohol problems as they are already engaged in the medical care system. The hospital utilization and reported episodic illnesses (Table 4) was by self-report and not validated in our study; however, studies have found self-reported health care use to be a valid measure at 4-month (Weissman et al., 1996) and 6-month recall (Palepu et al., 1999; Solomon et al., 1993) among HIV-infected individuals, including drug injectors; of note, the validity is highest for self-reported hospital use compared with ambulatory visits. Given the concurrent illicit drug use, we cannot ascertain the independent contribution that alcohol problems had on hospital admissions. However, polysubstance use among HIV-infected individuals is not uncommon given that many people contract HIV via injection drug use, and our sample is likely representative of HIV-infected individuals with alcohol problems. The duration of exposure to substance abuse treatment may not have been sufficient to have an association on hospitalization patterns. It is possible that our study was underpowered to detect a small association. Finally, we do not have the specific dates of when the patients were hospitalized or when they received substance abuse treatment. It is possible that hospitalization may serve as a sentinel event that motivates HIV-infected individuals with alcohol problems to seek substance abuse treatment, or, conversely, patients in substance abuse treatment relapse and this results in hospitalization for episodic non–HIV-related medical illnesses.

In summary, we found that engagement in substance abuse treatment services was not associated with reduced hospitalizations among HIV-infected individuals with alcohol problems. The counterintuitive nature of this finding suggests the need to better understand the phenomenon. Does providing substance abuse treatment for HIV-infected individuals with a history of alcohol problems not have an impact on their utilization of hospital resources, or, rather, does the present configuration of service delivery not take advantage of an opportunity to reduce hospitalizations in this high-risk population? This opportunity to provide higher quality and more efficient health care to this marginalized population seems too substantial not to examine further.

ACKNOWLEDGEMENTS

We appreciate the contributions of the clinical staff of the HIV Diagnostic Evaluation Unit at Boston Medical Center, including Colleen LaBelle, RN, and Jennifer Doyle, as well as assistance provided by researchers on the project: Jacqueline Savetsky, MPH, Jessica Garcia, Laura Weinstein, Sharon Phillips, Vera Doljanskaia, MPH, Kathleen Doyle, and Katie Schoolwerth. We are also grateful for the assistance by Karen Traister and Carly Briden. We likewise appreciate the contributions of Howard Libman, MD, and the clinical staff at the Beth Israel Deaconess Medical Center.

REFERENCES

Aday L (1994) Health status of vulnerable populations. Annu Rev Public Health 15:487–509.

Alexandre PK, Roebuck MC, French MT, Chitwood DD, McCoy CB (2001) Problem drinking, health services utilization, and the cost of medical care. Recent Dev Alcohol 15:285–298.

Andresen EM, Malmgren JA, Carter WB, Patrick DL (1994) Screening for depression in well older adults—evaluation of a short-form of the CES-D. Am J Prev Med 10:77–84.

Bozzette SA, Berry SH, Duan N, Frankel MR, Leibowitz AA, Lefkowitz D, Emmons CA, Senterfitt JW, Berk ML, Morton SC, Shapiro MF (1998) The care of HIV-infected adults in the United States. HIV Cost and Services Utilization Study Consortium. N Engl J Med 339:1897–1904.

Brands B, Blake J, Marsh D (2002) Changing patient characteristics with increased methadone maintenance availability. Drug Alcohol Depend 66:11–20.

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.

Conigliaro J, Gordon AJ, McGinnis KA, Rabeneck L, Justice AC (2003) How harmful is hazardous alcohol use and abuse in HIV infection: do health care providers know who is at risk? J Acquir Immune Defic Syndr 33:521–525.

Ewing JA (1984) Detecting alcoholism. The CAGE questionnaire. JAMA 252:1905–1907.

Farre M, Mas A, Torrens M, Moreno V, Cami J (2002) Retention rate and illicit opioid use during methadone maintenance interventions: a meta-analysis. Drug Alcohol Depend 65:283–290.

Fiellin DA, Reid C, O'Connor PG (2000) Screening for alcohol problems in primary care—a systematic review. Arch Intern Med 160:1977–1989.

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.

French MT, McGeary KA, Chitwood DD, McCoy CB (2000) Chronic illicit drug use, health services utilization and the cost of medical care. Soc Sci Med 50:1703–1713.

Friedmann PD, Saitz R, Samet JH (1998) Management of adults recovering from alcohol or other drug problems: relapse prevention in primary care. JAMA 279:1227–1231.

Gifford AL, Collins R, Timberlake D, Schuster MA, Shapiro MF, Bozzette SA, Kanouse DE (2000) Propensity of HIV patients to seek urgent and emergent care. HIV Cost and Services Utilization Study Consortium. J Gen Intern Med 15:833–840.

Hunkeler EM, Hung YY, Rice DP, Weisner C, Hu TW (2001) Alcohol consumption patterns and health care costs in an HMO. Drug Alcohol Depend 64:181–190.

Kertesz SG, Horton NJ, Friedmann PD, Saitz R, Samet JH (2003) Slowing the revolving door: stabilization programs reduce homeless persons' substance use after detoxification. J Subst Abuse Treat 24:197–207.

Knowlton AR, Hoover DR, Chung SE, Celentano DD, Vlahov D, Latkin CA (2001) Access to medical care and service utilization among injection drug users with HIV/AIDS. Drug Alcohol Depend 64:55–62.

Laine C, Hauck WW, Gourevitch MN, Rothman J, Cohen C, Turner BJ (2001) Regular outpatient medical and drug abuse care and subsequent hospitalization of persons who use illicit drugs. JAMA 285:2355–2362.

Lefevre F, O'Leary B, Moran M, Mossar M, Yarnold PR, Martin GJ, Glassroth J (1995) Alcohol consumption among HIV-infected patients. J Gen Intern Med 10:458–460.

Liang KY, Zeger SY (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22.

Lucas GM, Cheever LW, Chaisson RE, Moore RD (2001) Detrimental effects of continued illicit drug use on the treatment of HIV-1 infection. J Acquir Immune Defic Syndr 27:251–259.

McGeary KA, French MT (2000) Illicit drug use and emergency room utilization. Health Serv Res 35:153–169.

Pachl C, Todd JA, Kern DG, Sheridan PJ, Fong SJ, Stempien M, Hoo B, Besemer D, Yeghiazarian T, Irvine B, et al. (1995) Rapid and precise quantification of HIV-1 RNA in plasma using a branched DNA signal amplification assay. J Acquir Immune Defic Syndr 8:446–454.

Palepu A, Strathdee S, Hogg R, Anis A, Rae S, Cornelisse P, Patrick DM, O'Shaughnessy MV, Schechter MT (1999) The social determinants of emergency department and hospital use by injection drug users in Canada. J Urban Health 76:409–418.

Palepu A, Tyndall MW, Leon H, Muller J, O'Shaughnessy MV, Schechter MT, Anis AH (2001) Hospital utilization and costs in a cohort of injection drug users. Can Med Assoc J 165:415–420.

Petry NM (1999) Alcohol use in HIV patients: what we don't know may hurt us. Int J STD AIDS 10:561–570.

Ross HE, Gavin DR, Skinner HA (1990) Diagnostic validity of the MAST and the alcohol dependence scale in the assessment of DSM-III alcohol disorders. J Stud Alcohol 51:506–513.

Saitz R, Ghali WA, Moskowitz MA (2000) The impact of leaving against medical advice on hospital resource utilization. J Gen Intern Med 15:103–107.

Samet JH, Freedberg KA, Stein MD, Lewis R, Savetsky J, Sullivan L, Levenson SM, Hingson R (1998) Trillion virion delay: time from testing positive for HIV to presentation for primary care. Arch Intern Med 158:734–740.

Samet JH, Friedmann P, Saitz R (2001) Benefits of linking primary medical care and substance abuse services: patient, provider, and societal perspectives. Arch Intern Med 161:85–91.

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 (2004) Detecting alcohol problems in HIV-infected patients: use of the CAGE questionnaire. AIDS Res Hum Retroviruses 20:151–155.

Shapiro MF, Morton SC, McCaffrey DF, Senterfitt JW, Fleishman JA, Perlman JF, Athey LA, Keesey JW, Goldman DP, Berry SH, Bozzette SA (1999) Variations in the care of HIV-infected adults in the United States: results from the HIV Cost and Services Utilization Study. JAMA 281:2305–2315.

Smith MY, Rapkin BD, Winkel G, Springer C, Chhabra R, Feldman IS (2000) Housing status and health care service utilization among low-income persons with HIV/AIDS. J Gen Intern Med 15:731–738.

Solomon L, Astemborski J, Warren D, Munoz A, Cohn S, Vlahov D, Nelson KE (1993) Differences in risk factors for human immunodeficiency virus type 1 seroconversion among male and female intravenous drug users. Am J Epidemiol 137:892–898.

Stein MD (1994) Injected-drug use: complications and costs in the care of hospitalized HIV-infected patients. J Acquir Immune Defic Syndr 7:469–473.

Stein MD, Friedmann PD (2001) Generalist physicians and addiction care: from turfing to sharing the turf. JAMA 286:1764–1765.

Stein MD, Sobota M (2001) Injection drug users: hospital care and charges. Drug Alcohol Depend 64:117–120.

Strathdee SA, Celentano DD, Shah N, Lyles C, Stamboli VA, Macalino G, Nelson K, Vlahov D (1999) Needle-exchange attendance and health care utilization promote entry into detoxification. J Urban Health 76: 448–460.

Ware JE (1993) *SF-36 Health Survey. Manual and Interpretation Guide*. The Health Institute, Boston.

Ware JE (1994) *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. The Health Institute, Boston.

Weber AE, Yip B, O'Shaughnessy MV, Montaner JS, Hogg RS (2000) Determinants of hospital admission among HIV-positive people in British Columbia. Can Med Assoc J 162:783–786.

Weintraub E, Dixon L, Delahanty J, Schwartz R, Johnson J, Cohen A, Klecz M (2001) Reason for medical hospitalization among adult alcohol and drug abusers. Am J Addict 10:167–177.

Weissman JS, Levin K, Chasan-Taber S, Massagli MP, Seage GR 3rd, Scampini L (1996) The validity of self-reported health-care utilization by AIDS patients. AIDS 10:775–783.

Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42:121–130.

ELSEVIER

Journal of Substance Abuse Treatment 28 (2005) 3–9

Journal of Substance Abuse Treatment

Regular article

Substance abuse treatment and risk behaviors among HIV-infected persons with alcohol problems

Anita Palepu, (M.D.), (M.P.H.)[a,*], Anita Raj, (Ph.D.)[b], Nicholas J. Horton, (Sc.D.)[c], Nicole Tibbetts, (B.A.)[d], Seville Meli, (M.P.H.)[e], Jeffrey H. Samet, (M.D.), (M.A.), (M.P.H.)[b,e]

[a]*Centre for Health Evaluation and Outcome Sciences, St. Paul's Hospital, Department of Medicine, University of British Columbia, Vancouver, BC, Canada*
[b]*Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA*
[c]*Department of Mathematics, Smith College, Northampton, MA, USA*
[d]*DM-STAT Inc., Medford, MA, USA*
[e]*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, USA*

Received 18 February 2004; received in revised form 19 July 2004; accepted 24 September 2004

Abstract

We examined the association of substance abuse treatment with sexual and drug use risk behaviors among 349 HIV-infected persons with a history of alcohol problems using a standardized questionnaire regarding sexual and drug use risk behaviors, demographics, substance use, and use of substance abuse treatment.We defined substance abuse treatment services as any of the following in the past 6 months: 12 weeks in a half-way house or residential facility; 12 visits to a substance abuse counselor or mental health professional; day treatment for at least 30 days; or participation in any methadone maintenance program. Our three outcome variables of high-risk behavior were the Risk Assessment Battery sex-risk and drug-risk scores and high-risk sex behavior which included any of the following: inconsistent condom use; having more than one sexual partner; and exchanging sex for money or drugs. Although sexual risk was high (51%) in our HIV-infected cohort, engagement in substance abuse treatment was not independently associated with lower frequency of any of our measures of high- risk behaviors. Although the opportunity exists to address HIV risk behaviors in the setting of substance abuse treatment, effective institutionalization of this challenging behavior change effort has not yet been realized.

Keywords: HIV; sex risk; drug risk behaviors; alcohol; substance use; substance abuse treatment

1. Introduction

Several studies have reported that drug and alcohol use is associated with high-risk sexual and drug use behaviors (Fitterling, Matens, Scotti, & Allen, 1993; Malow, Devieux, Jennings, Lucenko, & Kalichman, 2001; Rees, Saitz, Horton, & Samet, 2001; Stein et al., 2000). People with heavy alcohol use tend to engage in riskier behaviors, such as sex with multiple partners, unprotected vaginal and anal intercourse, and injection drug use (Bagnall, Plant, & Warwick, 1990; Halpern-Felsher, Millstein, & Ellen, 1996; Leigh, Temple, & Trocki, 1994). The relationship between alcohol and sexual risk taking is complex, and may be explained in a number of ways. Alcohol use may influence high-risk behaviors by affecting judgment and increasing disinhibition. Alcohol use may be a marker of a risk-taking personality; people with heavy alcohol use may also be more likely to engage in high-risk sexual behaviors (Leigh et al., 1994). Thus, one critical approach to slow the spread of HIV among persons with alcohol problems is to change risky sexual behaviors.

Active drug and alcohol use is associated with increased sexual and drug using risk behaviors (Battjes, Pickens, &

* Corresponding author. 620B-1081 Burrard Street, Vancouver, BC, Canada, V6Z 1Y6. Tel.: +1 604 682 2344, local 63194; fax: +1 604 806 8005.

E-mail address: anita@hivnet.ubc.ca (A. Palepu).

0740-5472/05/$ – see front matter
doi:10.1016/j.jsat.2004.09.002

Brown, 1995; Booth, Kwiatkowski, & Chitwood, 2000; Rees et al., 2001; Stein et al., 2000). Certain harm reduction strategies have reduced these risk behaviors, such as enhancing methadone maintenance with an HIV harm reduction program targeting HIV-infected injection drug users. The authors found increases in both harm reduction and health promotion behaviors (Margolin, Avants, Warburton, Hawkins, & Shi, 2003).

Numerous studies, including systematic reviews, have examined the effect of various types of HIV prevention interventions and substance abuse treatment on high-risk sex and drug use behaviors (Avins et al., 1997; Gibson, McCusker, & Chesney, 1998; Gossop, Marsden, Stewart, & Treacy, 2002; Margolin et al., 2003; Prendergast, Podus, Chang, & Urada, 2002; Semaan et al., 2002; Sorensen & Copeland, 2000; Woods et al., 1999). These studies found that that these interventions can result in reduced sexual risk behaviors among drug users (Semaan et al., 2002). In the context of substance abuse treatment programs which were predominantly methadone maintenace therapy, one meta-analysis of such programs in North America, found reductions in risky sexual behavior and injection practices (Prendergast et al., 2002). Similarly, a study of 753 drug users enrolled in either methadone maintenance therapy or residential treatment settings in the United Kingdom found a reduction of risky behaviors at one-year follow-up (Gossop et al., 2002).

These findings demonstrate the utility of substance abuse treatment as a means of HIV prevention with persons with substance use problems, but few studies to date have assessed whether such intervention is associated with HIV-related risk reduction in HIV-infected persons with alcohol problems. High-risk behaviors in this population have serious public health implications with regards to HIV transmission and have been a particular focus of recent U.S. Centers for Disease Control and Prevention (CDC) prevention efforts. The effectiveness of prevention utilizing this opportunity, substance abuse treatment, in the life of a substance-abusing HIV-infected person is important to assess. Therefore, we hypothesized that engagement in substance abuse treatment services would reduce sexual and drug use risk behaviors among participants in the HIV-Alcohol Longitudinal Cohort (HIV-ALC) study.

2. Materials and methods

2.1. Study design and population

We analyzed data from a prospective cohort of HIV-infected patients with a history of alcohol problems. Patients who were HIV-infected and had a history of alcohol problems were identified by explicit eligibility criteria. All potential subjects who gave two or more positive responses to the CAGE questionnaire (Buchsbaum, Buchanan, Centor, Schnoll, & Lawton, 1991; Ewing, 1984; Fiellin, Reid, & O'Connor, 2000), a screening test for lifetime alcohol problems (sensitivity ~80%, specificity ~90%), were eligible. In addition, those patients recruited from the Boston Medical Center HIV Diagnostic Evaluation Unit (DEU; Samet et al., 1995) who did not meet CAGE criteria, were eligible if one of two attending physicians made a specific diagnosis of alcohol abuse or dependence. Thus subjects with alcohol problems, despite not being detected by the CAGE questionnaire, were detected by the clinical interview and recruited. Still, within the DEU site, most subjects were recruited based on CAGE criteria. Other entry criteria included the following: fluency in English or Spanish; Mini-Mental State Examination (MMSE) score greater or equal to 21 (Folstein, Folstein, & McHugh, 1975); and no plans to move from the Boston area in the next two years. As chronic alcohol use is associated with cognitive impairment, we used the MMSE cut-off of 21 to exclude subjects in whom such impairment may preclude obtaining informed consent, an accurate and complete interviewer-administered questionnaire, or a follow-up interview (Smith, Saitz, Horton, & Samet, 2003). The Institutional Review Boards of Boston Medical Center and Beth Israel Deaconess Medical Center approved this study.

From July 1997 through July 2001, recruitment of subjects occurred by multiple methods and from several sources: (a) Boston Medical Center HIV Diagnostic Evaluation Unit (56%); (b) posted flyers (17%); (c) Boston Medical Center Primary Care Clinic (13%); (d) respite facility for homeless persons (5%); (e) methadone clinic (4%); subject referrals (4%); and (f) Beth Israel Deaconess Medical Center (2%). The majority of subjects were recruited from medical settings that addressed HIV-related issues. The eligibility criteria of a history of alcohol problems was determined by the CAGE questionnaire in 313/349 (90%) of subjects, and based on clinical assessment in 36/349 (10%) of subjects. Diagnostic interviews for alcohol problems in a sample of the recruited subjects ($N = 141$) revealed in the overwhelming majority a lifetime history of alcohol dependence 80% (113/141) or abuse 15% (21/141; Samet, Phillips, Horton, Traphagen, & Freedberg, 2004). One hundred and fifty-one subjects in the cohort participated in a randomized controlled trial of an antiretroviral adherence intervention; appropriate adjustments were made to the analysis to account for the trial (Samet et al., 2002).

2.2. Data collection

After obtaining informed consent, a research associate or study investigator interviewed subjects using a standardized instrument to ascertain baseline information including the following: demographics, HIV risk behaviors, alcohol severity, use of substance abuse treatment services, and health care utilization in the preceding 6 months. For the Spanish interview standardized 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. We attempted to obtain CD4 cell counts and HIV RNA (viral load) levels on all subjects. Laboratory tests performed within 6 months of the interview as part of clinical care were recorded. If not available through routine clinical care, blood samples were obtained and tested for CD4 cell count and HIV RNA using the Boston Medical Center Clinical Laboratory. Subjects were followed every 6 months for up to seven observations; follow-up ended July 2001.

2.3. Outcome variables

Estimates of sexual risk and drug risk behaviors over the previous 6 months were derived from the Risk Assessment Battery (RAB; Navaline et al., 1994). This instrument sums the scores for individual items, for totals of 35 (sex risk) and 25 (drug risk). Items on each scale ask about HIV risk behavior in the past 6 months. Examples of the sex-risk questions are "with how many men/women have you had sex?" and "how often did you use condoms when you had sex?" Examples of the drug-risk questions include "have you shared needles or works?" and "how often have you been to a shooting gallery?" We also created a variable for high risk sexual behavior in the past 6 months which included any of the following: inconsistent condom use (i.e. not using a condom with every sexual encounter); having more than one sexual partner; and exchanging sex for money or drugs.

2.4. Primary independent variable

We defined substance abuse treatment services as any of the following in the past 6 months: at least 12 weeks in a half-way house or residential facility; at least 12 visits to a substance abuse counselor or mental health professional; day treatment for at least 30 days; or participation in any methadone maintenance program (Brands, Blake, & Marsh, 2002; Farre, Mas, Torrens, Moreno, & Cami, 2002).

2.5. Other independent variables

Other specific variables assessed included: (1) age; (2) gender; (3) depressive symptoms as measured by the 20-item Centers for Epidemiologic Studies Depression Scale (Andresen, Malmgren, Carter, & Patrick,1994); (4) self-reported use of cocaine and heroin in the past 30 days; (5) homelessness which was defined as having spent at least one night either on the street or in a shelter in the 6 months prior to the interview (Kertesz, Horton, Friedmann, Saitz, & Samet, 2003); and (6) currently being on antiretroviral therapy at the time of an interview. We used the National Institute on Alcohol Abuse and Alcoholism (NIAAA) guideline for at-risk drinking: greater than 14 drinks per week (or more than four drinks per day) for men, and greater than seven drinks per week (or more than three drinks per day) for women (NIAAA, 1995). Alcohol consumption below these levels was considered not at-risk in our study.

2.6. Analysis

We used multivariable longitudinal linear regression models to examine the association of substance abuse treatment on the following outcomes: sexual and drug use risk behavior as measured by the RAB sex risk score and the RAB drug risk score. A multivariable longtitudinal logistic regression model was constructed to examine the association of substance abuse treatment on high risk sexual behavior. Since serial measures on the same individuals were considered for these analyses, generalized estimating equations were used to adjust for correlation between these measures over time using a working independence correlation matrix and empirical variance estimator (Liang & Zeger, 1986; Zeger & Liang, 1986). The substance abuse treatment variable was dichotomous. For the two models examining sexual risk behavior, we adjusted for age, gender, race, cocaine use, heroin use, at-risk drinking, homelessness, depressive symptoms, and current receipt of antiretroviral therapy. In the drug use risk (i.e., RAB drug score) multivariable model we only included subjects who reported injecting drugs in the previous 6 months at baseline. In all models, all of the predictor variables except for gender, race and age were allowed to vary with time. All

Table 1
Bivariate associations of characteristics of HIV-infected subjects with alcohol problems and High-risk sexual risk behavior at baseline ($N = 349$)

	High-risk sexual behavior		
Characteristics	Yes	No	*p*-value
N	179	170	
Mean Age in years [*SD*]	39.5 [7.2]	41.7 [7.3]	0.006
Female (%)	41 (23)	32 (19)	0.35
Ethnicity (%)			
Black	83 (46)	71 (42)	0.36
White	61 (34)	55 (32)	
Other	35 (20)	44 (26)	
Homeless (%)[a]	52 (29)	49 (29)	0.96
Cocaine (%)[b]	61 (34)	23 (14)	<0.001
Heroin (%)[b]	25 (14)	12 (7)	0.04
At-risk alcohol (%)[c]	70 (39)	41 (24)	0.003
S.A. treatment (%)[d]	64 (36)	54 (32)	0.43
Mean CESD [SD][e]	22.2 [12.6]	22.7 [13.5]	0.69
Antiretroviral receipt[a] (%)	105 (59)	100 (59)	0.98

[a] In the past 6 months.
[b] In the past 30 days.
[c] At-risk alcohol consumption refers to greater than 14 drinks per week (or more than 4 drinks per day) for men, and greater than 7 drinks per week (or more than 3 drinks per day) for women.
[d] Substance abuse treatment: Having at least 12 weeks in a half-way house or residential facility; 12 visits to a substance abuse counselor or mental health professional; day treatment for at least 30 days; or participation in any methadone maintenance in the previous 6 months.
[e] Measure of depressive symptoms where a CESD score ≥16 indicate depressive symptoms.

Table 2
Bivariate associations of characteristics of HIV-infected subjects with alcohol problems and RAB sex-risk and RAB drug-risk scores at baseline

	RAB sex-risk ($N = 349$)		RAB drug-risk ($N = 122$)	
Characteristics	Mean (SD)	p-value	Mean (SD)	p-value
Sex				
Female	3.7 (3.3)	0.52	8.2 (7.0)	0.82
Male	3.4 (3.1)		8.5 (6.1)	
Ethnicity				
Black	3.8 (3.2)	0.09	8.3 (5.7)	0.20
White	3.5 (2.9)	0.30	7.1 (5.6)	0.03
Other (ref)	3.0 (3.1)		10.7 (7.5)	
Homeless[a]				
Yes	3.9 (3.7)	0.11	10.3 (6.1)	0.03
No	3.3 (2.8)		7.1 (6.2)	
Cocaine[b]				
Yes	4.1 (3.3)	<0.001	7.7 (5.8)	0.45
No	2.7 (2.9)		8.8 (6.6)	
Heroin[b]				
Yes	3.3 (2.6)	0.73	7.5 (6.1)	0.30
No	3.5 (3.2)		9.0 (6.5)	
At-risk alcohol[c]				
Yes	4.1 (3.3)	0.007	7.6 (5.5)	0.46
No	3.2 (2.9)		8.8 (6.7)	
S.A. treatment[d]				
Yes	3.3 (3.1)	0.48	8.0 (6.3)	0.45
No	3.6 (3.1)		9.0 (6.4)	
Antiretroviral Receipt[a]				
Yes	3.4 (2.9)	0.64	8.2 (6.7)	0.75
No	3.6 (3.3)		8.7 (6.0)	
Pearson Correlation Coefficients				
Age	−0.18	0.0007	0.002	0.98
CESD[e]	0.003	0.95	0.11	0.18

[a] In the past 6 months.
[b] In the past 30 days.
[c] At-risk alcohol consumption refers to greater than 14 drinks per week (or more than 4 drinks per day) for men, and greater than 7 drinks per week (or more than 3 drinks per day) for women.
[d] Substance abuse treatment: Having at least 12 weeks in a half-way house or residential facility; 12 visits to a substance abuse counselor or mental health professional; day treatment for at least 30 days; or participation in any methadone maintenance in the previous 6 months.
[e] Measure of depressive symptoms where a CESD score ≥16 indicate depressive symptoms.

analyses were carried out using SAS version 8.2 (SAS Institute Inc., Carey, NC).

3. Results

The 349 subjects had the following baseline characteristics: 79% were men, two thirds were ethnic minorities, mean age was 41 years, and 29% were homeless. The most common HIV risk factor was injection drug use (59%); with men having sex with men (19%) and heterosexual sex (22%). Of the injection drug users, 35% had injected drugs in the previous 6 months. Of those injecting drugs, 51 of 122 (42%) reported sharing needles or works. In the past 30 days, 24% of subjects (84/349) reported using alcohol and heroin or cocaine, 18% used alcohol alone and 5% used heroin or cocaine alone; 12% were enrolled in a methadone maintenance program. The average daily alcohol consumption of those drinking in the past 30 days was 6.4 drinks. Over one third (118/349) of the subjects were engaged in substance abuse treatment at the initial observation. Among the subjects who were not engaged in substance abuse treatment at the initial observation, 70 subjects entered substance abuse treatment during the study period. In this research study, the subjects were followed every 6 months for up to seven occasions and the median number of observations per subject was three. The distribution of interviews (observations) conducted per subject was as follows: 111 subjects had one interview; 40, two interviews; 48, three interviews; 44, four; 39, five; 47, six; and 20 completed seven interviews. Because study subjects were recruited over a four-year period, and all follow-ups ceased at the end of recruitment, time of recruitment was the major factor affecting the number of follow-up observations in this study (Ehrenstein, Horton, & Samet, 2004).

Subjects' characteristics associated with high-risk sexual risk behavior at baseline are presented in Table 1. A significantly higher proportion of subjects who were younger, more likely to use cocaine and heroin, and consume at-risk quantities of alcohol reported high-risk sexual behavior. Table 2 depicts the RAB sex-risk and drug-risk scores

Table 3
Multivariable linear and logistic regression models for the factors associated with RAB sex-risk and high-risk sexual behavior[a]

Factor	RAB sex-risk Adjusted Beta Estimate (95% CI)	High-risk sexual behavior Adjusted OR (95% CI)
S.A. treatment[b]	−0.0006 (−0.47, 0.47)	1.25 (0.87, 1.81)
Female	0.38 (−0.29, 1.05)	1.18 (0.72, 1.94)
Age (per year)	−0.08 (−0.12, −0.05)	0.95 (0.92, 0.98)
Race		
Black	0.04 (−0.60, 0.67)	1.09 (0.65, 1.84)
White	0.73 (−0.01, 1.47)	1.51 (0.87, 2.62)
Cocaine[c]	1.79 (1.19, 2.40)	2.33 (1.45, 3.74)
Heroin[c]	−0.88 (−1.51, −0.25)	0.92 (0.52, 1.62)
At-risk alcohol[d]	0.27 (−0.22, 0.75)	1.28 (0.86, 1.89)
Homelessness[e]	0.34 (−0.15, 0.84)	1.10 (0.73, 1.66)
CESD (per unit)[f]	−0.008 (−0.03, 0.009)	0.99 (0.98, 1.00)
Antiretroviral receipt[g]	0.21 (−0.21, 0.64)	1.05 (0.74, 1.49)

[a] Using generalized estimating equations and controlling for time; 1111 observations used in the RAB sex-risk model and 1114 observations used in the High-risk sexual behavior model.
[b] S.A. Treatment: Having at least 12 weeks in a halfway house or residential facility; 12 visits to a substance abuse counselor or mental health professional; day treatment for at least 30 days; or participation in any methadone maintenance in the previous 6 months.
[c] Use in the past 30 days.
[d] At-risk alcohol consumption refers to greater than 14 drinks per week (or more than 4 drinks per day) for men, and greater than 7 drinks per week (or more than 3 drinks per day) for women.
[e] In the past six months.
[f] A measure of depressive symptoms (CESD score ≥16 indicates depressive symptoms).
[g] Receipt of antiretroviral therapy at any time point.

Table 4
Multivariable linear regression model for the factors associated with RAB drug-risk[a]

Factor	RAB drug-risk Adjusted Beta Estimate (95% CI)
S.A. treatment[b]	0.20 (−1.17, 1.57)
Female	0.68 (−1.57, 2.94)
Age (per year)	−0.04 (−0.16, 0.09)
Race	
Black	−1.83 (−3.59, −0.07)
White	0.63 (−2.37, 1.11)
Cocaine[c]	1.28 (−0.25, 2.81)
Heroin[c]	0.15 (−1.18, 1.47)
At-risk alcohol[d]	−0.41 (−2.05, 1.24)
Homelessness[e]	2.23 (0.57, 3.89)
CESD (per unit)[f]	0.03 (−0.03, 0.09)
Antiretroviral receipt[g]	0.86 (−2.43, 0.70)

[a] Using generalized estimating equations and controlling for time; 227 observations used in the RAB drug-risk model.
[b] S.A. Treatment: Having at least 12 weeks in a halfway house or residential facility; 12 visits to a substance abuse counselor or mental health professional; day treatment for at least 30 days; or participation in any methadone maintenance in the previous 6 months.
[c] Use in the past 30 days.
[d] At-risk alcohol consumption refers to greater than 14 drinks per week (or more than 4 drinks per day) for men, and greater than 7 drinks per week (or more than 3 drinks per day) for women.
[e] In the past six months.
[f] A measure of depressive symptoms (CESD score ≥16 indicates depressive symptoms).
[g] Receipt of antiretroviral therapy at any time point.

by subject characteristics. HIV-infected persons who were younger, who used cocaine and consumed at-risk quantities of alcohol had significantly higher RAB sex-risk scores. Higher RAB drug-risk scores were significantly associated with being white and homeless.

Substance abuse treatment was not associated with less risky sexual behavior as measured by the two study outcomes, RAB sex-risk, or high-risk sexual behavior. The multivariable models of the factors independently associated with the RAB sex-risk score and high-risk sexual behavior are shown in Table 3. For both models, younger age and cocaine use were independently associated with measures of sexual risk behavior. In the RAB sex-risk model, heroin use was associated with a significantly lower score: less sexual risk.

Substance abuse treatment was not associated with the less risky drug use behavior as measured by RAB drug-risk score. The factors associated with the RAB drug-risk score are presented in Table 4. Black race was associated with a lower score and homelessness was associated with a higher RAB drug-risk score.

4. Discussion

Addressing the behaviors of HIV-infected persons has only recently attracted the main focus of prevention researchers and public health practitioners. This strategic approach has been adopted by the CDC (Jaffe & Janssen, 2003). By giving particular effort to promoting safe behaviors in the HIV-infected, in addition to the ongoing efforts to reduce HIV risks in the general population, the goal of limiting the spread of HIV infection may be better achieved. This study identified a group of HIV-infected persons with particularly high reports of recent HIV sexual and drug use risk, namely HIV-infected persons with alcohol problems (Ehrenstein et al., 2004). In this cohort, over half reported sexual or drug use risk behaviors for HIV infection in the past 6 months. Notably, 51% of this same at-risk population received substantial substance abuse treatment over the study period. Thus, it would be particularly appropriate for this population to receive effective HIV risk reduction while receiving care in a substance abuse treatment program. Our investigation found that unfortunately, in this cohort, engagement in substance abuse treatment was not associated with lower sexual or drug use risk behaviors. These findings are in stark contrast to previous research indicating that substance abuse treatment promotes HIV risk reduction behaviors in client populations (Avins et al., 1997; Gibson et al., 1998; Gossop et al., 2002; Margolin et al., 2003; Prendergast et al., 2002; Sorensen & Copeland, 2000; Woods et al., 1999).

Conflicting findings between our study and previous work on the specific finding that exposure to substance abuse treatment is not beneficial in terms of HIV risk reduction is noteworthy and merits explanation. This discrepancy may be a consequence of the fact that some of the previous studies were mainly examining substance abusers without regard to their HIV infection status. It is possible that education about the risk of HIV transmission when one is not already infected is more effective than when one is infected. In the latter situation the major perceived benefit of risk reduction is not one's own health, but the health of one's sex or drug partners. It may be that the education regarding high-risk behaviors need to better tailored to HIV-infected persons with substance use problems. For example, Margolin et al. (2003) studied 90 HIV-infected persons on methadone and tested a targeted HIV harm reduction program using a randomized controlled trial. They found reductions in both illicit opioid use and risk behaviors, as well as an increased uptake of antiretroviral therapy.

Another potential explanation for the discrepant findings between this study and previous work is the fact that, unlike in our study, many of the specific substance abuse treatment programs examined in previous work were aware that there was ongoing monitoring for clinical outcomes. In contrast, our research assessed drug treatment programs broadly, as they are practiced in the "real world." Although Massachusetts, the state in which the study was conducted, mandates HIV education as a part of substance abuse treatment delivery (Klein Walker, 1997) the effort expended to reduce HIV risks in these settings may simply be insufficient to promote HIV risk reduction. These findings suggest a need

for ongoing monitoring of efforts to address HIV risk behavior among the clients of substance abuse treatment programs. It is possible that periodic assessments of HIV risk behaviors could be a marker of quality of substance abuse treatment used in the accreditation process of substance abuse treatment programs.

Although we saw no relationship between substance use treatment exposure and HIV risk behaviors, consistent with other studies (Booth et al., 2000; Fuller et al., 2002; Latkin, Mandell, & Vlahov, 1996; Logan & Leukefeld, 2000; Rasch et al., 2000; Roy et al., 2003; Semaan et al., 1998; Tortu et al., 2000), we did find that younger age and cocaine use were associated with higher sexual risk behaviors, suggesting the need to target intervention efforts to these groups in particular. Notably, our findings demonstrated lower rates of HIV-related drug use risk among Blacks; these findings are surprising as previous study indicates higher injection drug use rates (Turner, Miller, & Moses, 1989) and related HIV risk (Kottiri, Friedman, Neaigus, Curtis, & Des Jarlais, 2002; CDC, 2002) among Blacks as compared with Whites. Discrepancy in findings may be attributable to the unique aspects of our sub-population of substance using HIV-infected persons (i.e., history of alcohol problems) and demonstrates the need for further focus on this population as risk and risk factors for this group may not be similar to other groups.

Our study has limitations. Although our definition of substance abuse treatment services is not as stringent as that used by Laine et al. (2001), we think it has face validity. Although 36% of this cohort were categorized as receiving substance abuse treatment services at a reasonable level of exposure, it is possible that higher levels of exposure may be required to demonstrate an effect. A strength of this study was the ability to follow a cohort of HIV-infected subjects over time, thus assessing subjects at times when substance abuse treatment was both present and absent.

In summary, we found that being engaged in substance abuse treatment was not associated with reduced sexual and drug use risk behaviors among HIV-infected persons with alcohol problems. Further research is needed to assess whether more intensive substance abuse treatment or integrated substance abuse treatment and HIV risk reduction intervention can effectively meet the needs of this population. Nonetheless, given that the majority of our participants reported recent involvement with substance abuse treatment programs, using this existing health care site for these high risk individuals may provide an important venue to reach this population for provision of effective HIV risk reduction programs.

Acknowledgments

The authors appreciate the contributions of the clinical staff of the HIV Diagnostic Evaluation Unit at Boston Medical Center, including Colleen LaBelle, RN, as well as assistance provided by researcher associates on the project. The authors likewise appreciate the contributions of Howard Libman, M.D., and the clinical staff 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) of the NIH: 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. Dr Anita Palepu is supported by the Canadian Institute for Health Research through a New Investigator Award and a Michael Smith Foundation for Health Research Senior Scholar Award.

References

Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults—Evaluation of a short-form of the CES-D. *American Journal of Preventive Medicine*, *10*, 77–84.

Avins, A. L., Lindan, C. P., Woods, W. J., Hudes, E. S., Boscarino, J. A., Kay, J., Clark, W., & Scully, S. B. (1997). Changes in HIV-related behaviors among heterosexual alcoholics following addiction treatment. *Drug and Alcohol Dependence*, *44*, 47–55.

Bagnall, G., Plant, M., & Warwick, W. (1990). Alcohol, drugs and AIDS-related risks: results from a prospective study. *AIDS Care*, *2*, 309–317.

Battjes, R. J., Pickens, R. W., & Brown Jr., L. S. (1995). HIV infection and AIDS risk behaviors among injecting drug users entering methadone treatment: an update. *Journal of Acquired Immune Deficiency Syndromes*, *10*, 90–96.

Booth, R. E., Kwiatkowski, C. F., & Chitwood, D. D. (2000). Sex related HIV risk behaviors: differential risks among injection drug users, crack smokers, and injection drug users who smoke crack. *Drug & Alcohol Dependence*, *58*, 219–226.

Brands, B., Blake, J., & Marsh, D. (2002). Changing patient characteristics with increased methadone maintenance availability. *Drug and Alcohol Dependence*, *66*, 11–20.

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.

Ehrenstein, V., Horton, N. J., & Samet, J. H. (2004). Inconsistent condom use among HIV-infected patients with alcohol problems. *Drug and Alcohol Dependence*, *73*, 159–166.

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

Farre, M., Mas, A., Torrens, M., Moreno, V., & Cami, J. (2002). Retention rate and illicit opioid use during methadone maintenance interventions: a meta-analysis. *Drug and Alcohol Dependence*, *65*, 283–290.

Fiellin, D. A., Reid, C., & O'Connor, P. G. (2000). Screening for alcohol problems in primary care—A systematic review. *Archives of Internal Medicine*, *160*, 1977–1989.

Fitterling, J. M., Matens, P. B., Scotti, J. R., & Allen, J. S. (1993). AIDS risk behaviors and knowledge among heterosexual alcoholics and non-injecting drug users. *Addiction*, *88*, 1257–1265.

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.

Fuller, C. M., Vlahov, D., Ompad, D. C., Shah, N., Arria, A., & Strathdee, S. A. (2002). High-risk behaviors associated with transition from illicit non-injection to injection drug use among adolescent and young adult drug users: a case-control study. *Drug and Alcohol Dependence*, *66*, 189–198.

Gibson, D. R., McCusker, J., & Chesney, M. (1998). Effectiveness of psychosocial interventions in preventing HIV risk behavior in injecting drug users. *AIDS*, *12*, 919–929.

Gossop, M., Marsden, J., Stewart, D., & Treacy, S. (2002). Reduced injection risk and sexual risk behaviours after drug misuse treatment: results from the National Treatment Outcome Research Study. *AIDS Care*, *14*, 77–93.

Halpern-Felsher, B. L., Millstein, S. G., & Ellen, J. M. (1996). Relationship of alcohol use and risky sexual behavior: a review and analysis of findings. *Journal of Adolescent Health*, *19*, 331–336.

Jaffe, H. W., & Janssen, R. S. (2003). Incorporating HIV prevention into the medical care of persons living with HIV. *MMWR*, *52* (RR12), 1–24.

Kertesz, S. G., Horton, N. J., Friedmann, P. D., Saitz, R., & Samet, J. H. (2003). Slowing the revolving door: stabilization programs reduce homeless persons' substance use after detoxification. *Journal of Substance Abuse Treatment*, *24*, 197–207.

Klein Walker, D. (1997). *HIV/AIDS Policy Guidelines For Substance Abuse Treatment Programs*. Boston: Commonwealth of Massachusetts Department of Public Health. Available at: http://www.state.ma.us/dph/bsas/licensing/hiv_aids.htm#1. Accessed on January 21, 2004.

Kottiri, B. J., Friedman, S. R., Neaigus, A., Curtis, R., & Des Jarlais, D. C. (2002). Risk networks and racial/ethnic differences in the prevalence of HIV infection among injection drug users. *Journal of Acquired Immune Deficiency Syndromes*, *30*, 95–104.

Laine, C., Hauck, W. W., Gourevitch, M. N., Rothman, J., Cohen, C., & Turner, B. J. (2001). Regular outpatient medical and drug abuse care and subsequent hospitalization of persons who use illicit drugs. *Journal of the American Medical Association*, *285*, 2355–2362.

Latkin, C. A., Mandell, W., & Vlahov, D. (1996). The relationship between risk networks' patterns of crack cocaine and alcohol consumption and HIV-related sexual behaviors among adult injection drug users: a prospective study. *Drug & Alcohol Dependence*, *42*, 175–181.

Leigh, B. C., Temple, M. T., & Trocki, K. F. (1994). The relationship of alcohol use to sexual activity in a U.S. national sample. *Social Science & Medicine*, *39*, 1527–1535.

Liang, K. Y., & Zeger, S. Y. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22.

Logan, T. K., & Leukefeld, C. (2000). Sexual and drug use behaviors among female crack users: a multi-site sample. *Drug & Alcohol Dependence*, *58*, 237–245.

Malow, R. M., Devieux, J. G., Jennings, T., Lucenko, B. A., & Kalichman, S. C. (2001). Substance-abusing adolescents at varying levels of HIV risk: Psychosocial characteristics, drug use, and sexual behavior. *Journal of Substance Abuse*, *13*, 103–117.

Margolin, A., Avants, S. K., Warburton, L. A., Hawkins, K. A., & Shi, J. (2003). A randomized clinical trial of a manual-guided risk reduction intervention for HIV-positive injection drug users. *Health Psychology*, *22*, 223–228.

Navaline, H. A., Snider, E. C., Petro, C. J., Tobin, D., Metzger, D., Alterman, A. I., & Woody, G. E. (1994). Preparations for AIDS vaccine trials. An automated version of the Risk Assessment Battery (RAB): enhancing the assessment of risk behaviors. *AIDS Research and Human Retroviruses*, *10* (Suppl 2), S281–S283.

National Institute on Alcohol Abuse and Alcoholism. (1995). *The physician's guide to helping patients with alcohol problems*. Washington, DC: National Institutes of Health.

Prendergast, M. L., Podus, D., Chang, E., & Urada, D. (2002). The effectiveness of drug abuse treatment: a meta-analysis of comparison group studies. *Drug and Alcohol Dependence*, *67*, 53–72.

Centers for Disease Control and Prevention. (2002). *HIV/AIDS Surveillance Report*. Accessed January 21, 2004 at http://www.cdc.gov/hiv/stats/hasrlink.htm.

Rasch, R. F., Weisen, C. A., MacDonald, B., Wechsberg, W. M., Perritt, R., & Dennis, M. L. (2000). Patterns of HIV risk and alcohol use among African-American crack abusers. *Drug & Alcohol Dependence*, *58*, 259–266.

Rees, V., Saitz, R., Horton, N. J., & Samet, J. (2001). Association of alcohol consumption with HIV sex- and drug-risk behaviors among drug users. *Journal of Substance Abuse Treayment*, *21*, 129–134.

Roy, E., Haley, N., Leclerc, P., Cedras, L., Blais, L., & Boivin, J. F. (2003). Drug injection among street youths in Montreal: predictors of initiation. *Journal of Urban Health*, *80*, 92–105.

Samet, J. H., Horton, N. J., Dukes, K., Meli, S., Tripp, T., Labelle, C., & Freedberg, K. A. (2002). *Randomized controlled trial of a multidimensional intervention to enhance adherence to antiretroviral therapy in HIV-infected patients with a history of alcohol problems.* Abstract presented at the XIV International AIDS Conference, Barcelona.

Samet, J. H., Libman, H., LaBelle, C., Steger, K., Lewis, R., Craven, D. E., & Freedberg, K. A. (1995). A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. *Archives of Internal Medicine*, *155*, 1629–1633.

Samet, J. H., Phillips, S. J., Horton, N. J., 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.

Semaan, S., Des Jarlais, D. C., Sogolow, E., Johnson, W. D., Hedges, L. V., Ramirez, G., Flores, S. A., Norman, L., Sweat, M. D., & Needle, R. (2002). A meta-analysis of the effect of HIV prevention interventions on the sex behaviors of drug users in the United States. *Journal of Acquired Immune Deficiency Syndromes*, *30*, S73–S93.

Semaan, S., Kotranski, L., Collier, K., Lauby, J., Halbert, J., & Feighan, K. (1998). Temporal trends in HIV risk behaviors of out-of-treatment injection drug users and injection drug users who smoke crack. *Journal of Acquired Immune Deficiency Syndromes*, *19*, 274–281.

Smith, K. L., Saitz, R., Horton, N. J., & Samet, J. H. (2003). Use of the mini-mental state examination in clinical alcohol research studies. *Alcoholism: Clinical and Experimental Research*, *27*, 28A.

Sorensen, J. L., & Copeland, A. L. (2000). Drug abuse treatment as an HIV prevention strategy: a review. *Drug and Alcohol Dependence*, *59*, 17–31.

Stein, M. D., Hanna, L., Natarajan, R., Clarke, J., Marisi, M., Sobota, M., & Rich, J. (2000). Alcohol use patterns predict high-risk HIV behaviors among active injection drug users. *Journal of Substance Abuse Treatment*, *18*, 359–363.

Tortu, S., Beardsley, M., Deren, S., Williams, M., McCoy, H. V., Stark, M., Estrada, A., & Goldstein, M. (2000). HIV infection and patterns of risk among women drug injectors and crack users in low and high seroprevalence sites. *AIDS Care*, *12*, 65–76.

Turner, C. F., Miller, H. G., & Moses, L. E. (1989). *AIDS, sexual behavior and intravenous drug use*. Washington, DC: National Academy Press.

Woods, W. J., Guydish, J. R., Sorensen, J. L., Coutts, A., Bostrom, A., & Acampora, A. (1999). Changes in HIV-related risk behaviors following drug abuse treatment. *AIDS*, *13*, 2151–2155.

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121–130.

Nicotine & Tobacco Research Volume 7, Number 3 (May 2005) 1–6

Taylor & Francis
Taylor & Francis Group

Smoking policies in U.S. outpatient drug treatment facilities

Kimber P. Richter, Won S. Choi, Daniel P. Alford

[Received 24 June 2004; accepted 19 November 2004]

Most drug treatment patients smoke cigarettes, and some facilities are beginning to help patients quit. Facility smoking policies can help or hinder this effort. The present study describes smoking policies in outpatient drug treatment. It is a secondary analysis of a survey on smoking cessation treatment in outpatient methadone maintenance facilities in the United States. One clinic leader (a medical director, head nurse, or clinic director) from each of the 697 U.S. facilities was invited to participate in the study. Main outcome measures included whether clinics had a written smoking policy as well as the types of indoor and outdoor policies in place for patients and staff. A total of 408 (59%) of U.S. clinics responded. Most clinics (73%) had a written smoking policy for patients, and more (82%) had written policies for staff. Over 90% banned indoor smoking by staff and patients. Few totally banned outdoor smoking. Approximately half in some way restricted where patients (48%) and staff (55%) smoke outdoors. Compared with clinics that did not treat nicotine dependence, significantly more clinics that treated nicotine dependence had written policies on smoking and restricted outdoor smoking for patients and staff. Likewise, many public clinics and those affiliated with hospitals had outdoor smoking restrictions for patients and staff. Drug treatment facilities routinely ban alcohol use and drug dealing, but only one in 10 banned smoking on the grounds. Outpatient facilities should restrict or ban outdoor tobacco use in order to remain consistent with their mission and avoid sabotaging clinic efforts to treat, and patient and staff efforts to stop, smoking.

Introduction

The vast majority of patients (80%–90%) in treatment for drug abuse smoke cigarettes (Sullivan & Covey, 2002). Most clinics do not treat nicotine dependence routinely, and many patients in recovery die from tobacco-related illnesses. Among 405 persons who initially entered drug treatment in 1964–1965, the death rate of persons who continued to smoke was four times that of nonsmokers 20 years later (Hser, McCarthy, & Anglin, 1994). Likewise, among 845 persons admitted for addiction treatment, the observed mortality 20 years later was 48% versus an expected 19%; tobacco accounted for 51% of these deaths (Hurt et al., 1996).

Kimber P. Richter, Ph.D., M.P.H., Won S. Choi, Ph.D., M.P.H., University of Kansas Medical School and Kansas Cancer Institute, Kansas City; Daniel P. Alford, MD, MPH., Boston University School of Medicine and Boston Medical center, Boston

Correspondence: Kimber P. Richter, Ph.D., M.P.H., Department of Preventive Medicine and Public Health, Mail-Stop 1008, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160, USA. Tel: +1 (913)-588-2718; Fax: +1 (913)-588-2780; E-mail: krichter@kumc.edu

Some drug treatment facilities are beginning to treat smoking among their patients, and smoking policy reform is an important part of the process (Fogg & Borody, 2001). Rationales for smoking restrictions in drug treatment facilities parallel reasons for limiting smoking in other health care facilities (Centers for Disease Control and Prevention, 1989; U.S. Department of Health and Human Services, 2000): Environmental tobacco smoke is harmful, treatment facilities—like any other worksite—should ensure clean indoor air for employees, and permitting smoking undermines staff advice to quit as well as quit attempts by staff and patients. Stopping smoking will reduce disability, improve quality of life, and lengthen the lives of patients (McCarthy, Zhou, Hser, & Collins, 2002; Peto et al., 2000). Stopping smoking will probably improve illicit drug use outcomes; several studies have shown that attempts to quit smoking do not worsen other drug use outcomes (Burling, Marshall, & Seidner, 1991; Campbell, Wander, Stark, & Holbert, 1995; Richter, Catley, McCool, Hall, & Ahluwalia, in press; Shoptaw, Jarvik, Ling, & Rawson, 1996) and that

ISSN 1462-2203 print/ISSN 1469-994X online
DOI: 10.1080/

abstinence from smoking is related to lower rates of illicit drug use among drug treatment patients (Shoptaw et al., 1996, 2002).

Smoking policies can help or hinder efforts to incorporate nicotine dependence treatment into drug treatment. Some drug treatment facilities provide a variety of formal and informal services for nicotine dependence (Currie, Nesbitt, Wood, & Lawson, 2003; Richter, Choi, McCool, Harris, & Ahluwalia, 2004). Several developments suggest that more facilities are beginning to do so. Knapp and colleagues surveyed Minnesota chemical dependency treatment facility directors in 1988 and again in 1990; the percentage of facilities that treated tobacco dependence increased from 10% at baseline to 18% at follow-up, and those prohibiting smoking increased from 11% to 27% (Knapp, Rosheim, Meister, & Kottke, 1993). In 1996, the Massachusetts Bureau of Substance Abuse Services (BSAS) required that all facilities receiving BSAS funds be smoke-free; this policy also provided training on nicotine addiction and treatment (Bureau of Substance Abuse Services, 1996). In 1997, facilities licensure standards in New Jersey required addictions treatment programs to address tobacco in their policies and treatment offerings (Hoffman et al., 1997). More recently, several national associations for the treatment of drug addiction have issued policy statements supportive of treating nicotine dependence among persons with other drug addictions (American Society of Addiction Medicine, 2001; NAADAC, 2002).

Lax smoking policies could send a mixed message to counselors (who are charged with treating smoking) as well as to the patients they serve. For example, a counselor might provide brief advice to a patient to quit smoking, only to witness that patient walk through a gauntlet of smokers outside the facility on the way to his car. Such a situation could undermine counselor motivation to address smoking with other patients. Conversely, restrictive smoking policies in facilities could support counselor advice to quit smoking and support counselor attempts to quit smoking themselves. In this way, policies could either speed or slow adoption of consistent treatment practices within individual facilities and across treatment systems.

We do not know the prevalence or types of smoking policies in drug treatment facilities. Multiple studies describe the impact of indoor smoking bans in solitary units and focus on residential or inpatient facilities (El-Guebaly, Cathcart, Currie, Brown, & Gloster, 2002). Their findings apply to a minority of facilities because only 26%–31% of U.S. drug treatment facilities offer residential rehabilitation or hospital inpatient services (Substance Abuse and Mental Health Services Administration, 2002). Smoking policies probably are different for outpatient facilities because patients have more autonomy and spend less time at the treatment unit. Understanding the prevalence of clinics with smoking policies, the types of policies in place, and the factors that influence those policies may help policy makers develop strategies for disseminating optimal policies. Strong policies may help protect the health of nonsmoking patients and staff as well as support the diffusion of nicotine dependence treatment in drug treatment.

The present study describes smoking policies in outpatient methadone maintenance facilities. This study was based on a national survey of methadone clinics that sought to establish the prevalence and type of nicotine dependence services offered. This is the first study, to our knowledge, that examines smoking policies in a national sample of outpatient drug treatment facilities, which serve the vast majority of persons presenting for drug treatment.

Method

Clinic leaders from all outpatient methadone maintenance facilities in the United States were invited to participate. The 20-min survey was collected by phone, fax, or mail from fall 2001 through spring 2002. One respondent—the medical director, clinic director, head nurse, or supervising counselor—completed the survey for each clinic. Details of survey development, methods, and main findings are available elsewhere (Richter et al., 2004).

Survey items collected information on respondent demographics, clinic and staff characteristics, and smoking cessation services provided. The present study focused on whether clinics had any written policy on smoking and what their indoor and outdoor policies were. Questions were asked separately regarding policies for staff and patients. Questions (and response categories) were derived from a regional survey of drug treatment facilities (Hahn, Warnick, & Plemmons, 1999) and included the following: "Does your program have a written smoking policy for your employees/patients that designates where and when they can smoke?" (yes, no); "Which of the following statements best describes your program's current indoor smoking rules for staff/patients?" (anywhere indoors, designated rooms or areas only, prohibited from smoking indoors); "Which of the following statements best describes your program's current outdoor smoking rules for staff/patients?" (anywhere outside, in designated outdoor areas only, prohibited from smoking outside on facility grounds).

Our survey included measures of characteristics that might be related to clinic smoking policies. Hospital size and whether hospitals are in tobacco states are two factors associated with the prevalence and restrictiveness of smoking policies found in

hospitals (Emmons & Biener, 1993; Longo et al., 1998). Main study findings from our methadone facility survey were that clinic ownership (nonprofit, for profit, or public) was related to whether the facility provided tobacco treatment services (Richter et al., 2004); perhaps ownership as well as whether the facility provides services are related to clinic tobacco policies. Other characteristics also might be linked to tobacco policy, including the type of organization with which methadone clinics were affiliated (mental hospital, hospital, or freestanding) and the clinic's type of accreditation. Our survey included measures derived from hospital-based studies (Emmons & Biener, 1993; Longo et al., 1998) and national surveys of methadone services (D'Aunno, Folz-Murphy, & Lin, 1999). We hypothesized that these characteristics, along with whether the clinic offered nicotine dependence services, would be associated with the prevalence and type of methadone clinic tobacco policies.

We summarized clinic responses to tobacco policy questions using frequencies and percentages. We used chi-square statistics to examine whether organizational characteristics were related to tobacco policies.

Results

Out of 697 outpatient methadone maintenance clinics in the United States, 408 (59%) participated in the survey (Richter et al., 2004). Most participants (72%; $n=293$) completed surveys by phone, the remainder (28%; $n=115$) by mail or fax. Clinics had on average 229 patients; their ownership was a mix of private for-profit (40), private nonprofit (37), and public (23); and an estimated 22% of their staff and 71% of their patients smoked cigarettes. We compared data on our participating clinics to data from the National Survey of Substance Abuse Treatment Services, which collects data from 100% of drug treatment facilities in the United States. Participating clinics were similar in size, region, and ownership to nonresponding clinics (Richter et al., 2004). Below, we describe smoking policies across all clinics and examine differences in the prevalence of policies according to clinic type.

Written smoking policies

Most clinics (73%) had a written policy on smoking for patients, and a higher percentage (82%) had written policies for staff. Table 1 displays the organizational characteristics of clinics that did and did not have written smoking policies. A small proportion of clinics with no accreditation had written smoking policies for their patients. A large proportion of clinics that provided nicotine dependence treatment to their patients had written smoking policies for both patients and staff.

Indoor smoking policies

Some clinics restricted patient (2%) and staff (2.7%) smoking to designated areas indoors. Most banned indoor smoking completely—95.8% banned patients and 94.8% banned staff from smoking anywhere indoors.

Outdoor smoking policies

Clinics were less restrictive in their outdoor smoking policies. Few completely banned outdoor smoking: 10.7% banned patients and 8.4% banned staff from smoking anywhere outside. Many clinics restricted in some way where patients (47.8%) and staff (55.2%) could smoke outdoors. A sizable minority of clinics allowed patients (40.8%) and staff (35.2%) to smoke anywhere outdoors.

Table 1 displays the organizational characteristics of clinics that banned or restricted outdoor smoking. Outdoor smoking policies tended to be more restrictive for staff than patients. Compared with clinics that did not treat nicotine dependence, significantly more clinics that treated nicotine dependence had written policies on smoking and banned or restricted outdoor smoking for patients and staff. A significantly higher proportion of public clinics, compared with private for-profits and nonprofits, banned or restricted outdoor smoking for patients and staff. Significantly more JCAHO-accredited clinics, compared with those accredited by the Commission on Accreditation of Rehabilitation Facilities (CARF) and other organizations, banned or restricted outdoor smoking for patients. A significantly higher proportion of clinics attached to hospitals, compared with those attached to mental health organizations or freestanding clinics, banned or restricted outdoor smoking for both patients and staff.

Discussion

Most outpatient methadone maintenance clinics banned indoor smoking, and many restricted where patients and staff could smoke outdoors. Whether clinics had a written policy and whether they restricted outdoor smoking was consistently related to whether the clinic offered nicotine dependence treatment. Other characteristics, including accreditation, ownership, and organizational affiliation, also were associated with having written policies or outdoor smoking restrictions. Being located in a tobacco state was not associated with policies or restrictions.

Some caution should be exercised in considering the findings from this study. First, 41% of the group did not respond. This could mean those that did not have policies or were not interested in them did

Table 1. Smoking policies and outdoor smoking restrictions at U.S. methadone maintenance facilities[a,b]

Organizational characteristic (*n*)	Smoking policy for patients		Smoking policy for staff	
	Have written smoking policy	Policy restricts or bans outdoor smoking[c]	Have written smoking policy	Policy restricts or bans outdoor smoking[c]
Type of ownership				
Private for profit (159)	69	50	80	57
Private nonprofit (149)	75	63	57	67
Public (91)	76	70	80	76
p value	.47	<.00	.21	.01
Type of accreditation[d]				
JCAHO (160)	72.5	67.7	85.2	71.8
CARF (45)	71.1	56.8	82.2	60.0
Other (140)[e]	79.3	53.6	80.0	60.6
None (57)	56.1	48.2	74.1	55.4
p value	.01	.02	.30	.07
Type of organization				
Hospital (70)	77	70	89	76
Mental health (37)	73	58	78	60
Freestanding (245)	72	54	80	59
Other (50)	68	70	80	77
p value	.71	.03	.40	.02
Clinic size (patients)				
<150 (116)	71.6	59.1	81.7	60.5
150–300 (149)	69.1	52.4	76.5	60.3
>300 (137)	76.6	65.9	86.9	72.1
p value	.355	.067	.079	.069
Provides nicotine dependence treatment[f]				
Yes (170)	80.0	66.3	87.6	70.6
No (232)	66.8	53.5	77.2	59.7
p value	.004	.010	.008	.025
In a tobacco state[g]				
Yes (95)	73.7	59.1	80.8	61.7
No (307)	72.0	58.8	81.8	65.2
p value	.746	.957	.843	.542

Note. [a]All variable values are reported as percentages.
[b]The *p* values are based on chi-square test for differences between groups; 408 clinics participated in the survey, of these 401 had complete data and are included in the table.
[c]This column includes clinics that have any form of outdoor smoking policy: either restrictions or a total ban.
[d]The exact wording of this question is "Which of the following agencies has accredited or licensed your methadone unit? (JCAHO, CARF, Other, None)" (JCAHO=Joint Commission on the Accreditation of Health Care Organizations; CARF=Commission on Accreditation of Rehabilitation Facilities). At the time the survey was conducted, some were monitored by the U.S. Food and Drug Administration and had not yet been accredited. These clinics answered "None."
[e]Many respondents chose the "Other" category and wrote in descriptions of their facilities. These included clinics affiliated with outpatient health clinics, medical schools, other forms of outpatient drug treatment, and municipal/county facilities.
[f]This item is a composite score including clinics that (a) provided individual or group counseling or (b) provided any form of nicotine replacement therapy to at least one patient in the past 30 days.
[g]Eleven states growing at least 10 million pounds of tobacco: North Carolina, Kentucky, Tennessee, South Carolina, Virginia, Georgia, Pennsylvania, Ohio, Florida, Indiana, Maryland.

not participate. Second, one individual was asked to respond from each facility—that individual's responses reflect the degree of familiarity he or she has with day-to-day operations and also his or her own personal biases. Third, findings may not generalize to other treatment facilities such as therapeutic communities, chemical-free outpatient treatment, and self-help groups such as Alcoholics Anonymous. Fourth, the relationships observed in this cross-sectional survey cannot be assumed to be causal. Finally, it is unknown whether the written policies were implemented or enforced because this information was not available.

New guidelines for methadone maintenance have switched oversight from the U.S. Food and Drug Administration to accreditation by JCAHO, CARF, and other health and rehabilitative organizations (Center for Substance Abuse Treatment, 2002). Our survey was conducted during this transition time, and many clinics had not undergone accreditation. Our survey found that unaccredited clinics were less likely than accredited clinics to have written smoking policies. Either (a) accreditation is making clinics adopt written smoking policies or (b) clinics that had smoking policies became accredited first. Future evaluations of the accreditation process should examine whether accreditation led to policy formation and whether these new written smoking policies were translated into practice.

Hospitals and drug treatment facilities may influence each others' tobacco policies. JCAHO accredits hospitals and requires them to adopt strict tobacco policies for indoor smoking. Many hospitals exceed these standards by instituting restrictive

outdoor smoking policies (Longo et al., 1998). This may be why most methadone clinics with JCAHO affiliation or those sited at hospitals had restrictive outdoor smoking policies. Interestingly, the relationship appeared to work both ways. Studies on tobacco policy in hospitals found that hospitals with chemical dependency units had more permissive indoor and outdoor smoking policies than did hospitals without inpatient drug treatment (Emmons & Biener, 1993; Longo et al., 1998).

Our data suggest that few patients or staff were exposed to tobacco smoke inside clinics. Outside clinics, the exposure to tobacco was probably substantial considering that the average clinic (with 229 patients and 71% smoking prevalence) treated 163 smokers—many of whom attended clinic daily.

These facilities appear to have a policy double-standard. To avoid sabotaging treatment efforts, alcohol use and drug dealing are routinely and explicitly banned from the grounds of drug treatment sites. We found that only one in 10 banned outdoor smoking.

Such a ban would be consistent with the organizational mission of these facilities. Most patients in drug treatment are interested in quitting smoking (Clemmey, Brooner, Chutuape, Kidorf, & Stitzer, 1997; Frosch, Shoptaw, Jarvik, Rawson, & Ling, 1998; Orleans & Hutchinson, 1993; Sees & Clark, 1993), staff are willing to help them do so (Knapp et al., 1993), and quitting smoking can improve other treatment outcomes (Shoptaw et al., 2002).

A number of implications flow from study findings. Private, for-profit clinics are the fastest growing sector in methadone maintenance treatment (Institute of Medicine, 1995). These facilities were the least likely to restrict outdoor smoking. They may not be subject to tobacco regulations that apply to clinics sited in hospitals or government buildings. Private for-profit clinics also are the type of facility least likely to provide nicotine dependence treatment (Richter et al., 2004). To address these discrepancies it may be necessary for methadone treatment regulations to explicitly require for-profit facilities to develop policies and treatments for nicotine dependence.

Facilities treating nicotine dependence were more likely to have restrictive outdoor smoking policies, compared with those not offering such treatment. Offering nicotine treatment could have caused facilities to adopt more stringent smoking policies; conversely, stringent policies could have encouraged facilities to treat smoking among their patients. A more in-depth study of the dynamics of the relationship between tobacco policies and nicotine dependence treatment might shed more light on how to help other clinics make the transition.

Because stopping smoking appears to facilitate recovery, and many patients want to make smoking cessation a treatment goal, outpatient drug treatment facilities should severely restrict or ban tobacco use on their premises. This change might, however, be accompanied by adverse consequences. For example, in a 1988 survey of 227 chemical dependency professionals located in Minnesota, Knapp and colleagues (1993) found that the most commonly cited barrier to a smoke-free facility was fear of losing patients. Smoking bans might prevent patients from succeeding in treatment, possibly related to drug relapse or attrition. Also, it is likely that many staff and patients smoke together outdoors, and informal but potentially therapeutic interactions might take place during these times. Future research should examine the impact of smoking bans on treatment outcomes and other outcomes important to facilities, including losing patients to other facilities and the potential loss of "smoke-break" contact between staff and patients.

These adverse consequences should be weighed against the potential benefits of banning smoking. Minnesota chemical dependency staff felt that a total smoking ban could benefit facilities by protecting staff and patient health, improving air quality, strengthening the message that smoking is an addiction, and enhancing efforts to treat tobacco (Knapp et al., 1993). If adopting smoking bans influences only a few clinics to begin to treat nicotine dependence, or helps only a small proportion of patients quit smoking, then the benefit to the drug treatment population could still be large because so many patients smoke and smoking is so deadly.

It may be that state- or federally mandated smoking bans would have the greatest impact and be the most welcome to treatment providers. Most states reimburse private and public facilities for providing drug treatment services. Reimbursement is contingent on adherence to guidelines and mandates. If smoking bans were mandatory, all facilities receiving reimbursement would have to comply. The playing field would be leveled and a major barrier—the fear of losing patient to facilities that permit smoking—would be removed. Total smoking bans at outpatient drug treatment facilities are ethically justifiable because patients are free to come and go at will and because these facilities exist to help people recover from drug dependence.

Acknowledgments

This study was funded by the Robert Wood Johnson Foundation Substance Abuse Policy Research Program (#042042) and the National Institute on Drug Abuse (DA00450).

References

American Society of Addiction Medicine. (2001). *Public policy of ASAM: Nicotine dependence and tobacco.* Retrieved October 19, 2004, from http://www.asam.org

Bureau of Substance Abuse Services. (1996). *Tobacco-free policy.* Retrieved October 19, 2004, from http://www.mass.gov/dph/bsas/licensing/notobaccoguide.htm

Burling, T. A., Marshall, G. D., & Seidner, A. L. (1991). Smoking cessation for substance abuse inpatients. *Journal of Substance Abuse, 3,* 269–276.

Campbell, B. K., Wander, N., Stark, M. J., & Holbert, T. (1995). Treating cigarette smoking in drug-abusing clients. *Journal of Substance Abuse Treatment, 12,* 89–94.

Center for Substance Abuse Treatment. (2002). *CSAT guidelines for the accreditation of opioid treatment programs.* Retrieved April 26, 2004, from http://dpt.samhsa.gov/guidelines.pdf

Centers for Disease Control and Prevention. (1989). *Reducing the health consequences of smoking: 25 years of progress. A report of the surgeon general* (DHHS Publication No. CDC88-406). Washington, DC.: U.S. Department of Health and Human Services, Office on Smoking and Health.

Clemmey, P., Brooner, R., Chutuape, M. A., Kidorf, M., & Stitzer, M. (1997). Smoking habits and attitudes in a methadone maintenance treatment population. *Drug and Alcohol Dependence, 44,* 123–132.

Currie, S. R., Nesbitt, K., Wood, C., & Lawson, A. (2003). Survey of smoking cessation services in Canadian addiction programs. *Journal of Substance Abuse Treatment, 24,* 59–65.

D'Aunno, T., Folz-Murphy, N., & Lin, X. (1999). Changes in methadone treatment practices: Results from a panel study, 1988–1995. *American Journal of Drug and Alcohol Abuse, 25,* 681–699.

El-Guebaly, N., Cathcart, J., Currie, S., Brown, D., & Gloster, S. (2002). Public health and therapeutic aspects of smoking bans in mental health and addiction settings. *Psychiatric Services, 53,* 1617–1622.

Emmons, K. M., & Biener, L. (1993). The impact of organizational characteristics on smoking policy restrictions in midwestern hospitals. *American Journal of Health Promotion, 8,* 43–49.

Fogg, B., & Borody, J. (2001, September). The impact of facility no smoking policies and the promotion of smoking cessation on alcohol and drug rehabilitation program outcomes. Canadian Centre on Substance Abuse Addictions Policy Working Group. Retrieved November 30, 2004, from http://www.ccsa.ca/pdf/ccsa-008967-2001.pdf

Frosch, D. L., Shoptaw, S., Jarvik, M. E., Rawson, R. A., & Ling, W. (1998). Interest in smoking cessation among methadone maintained outpatients. *Journal of Addictive Diseases, 17,* 9–19.

Hahn, E. J., Warnick, T. A., & Plemmons, S. (1999). Smoking cessation in drug treatment programs. *Journal of Addictive Diseases, 18,* 89–101.

Hoffman, A., , L., Kantor, B., Leech, T., Lindberg, D., Order-Connors, B., Schreiber, J., & Slade, J. (1997). *Drug-free is nicotine-free: Addressing tobacco in the treatment and prevention of other addictions.* New Brunswick, NJ: St. Peter's Medical Center.

Hser, Y. I., McCarthy, W. J., & Anglin, M. D. (1994). Tobacco use as a distal predictor of mortality among long-term narcotics addicts. *Preventive Medicine, 23,* 61–69.

Hurt, R. D., Offord, K. P., Croghan, I. T., Gomez-Dahl, L., Kottke, T. E., Morse, R. M., & Melton, L. J. 3rd. (1996). Mortality following inpatient addictions treatment. Role of tobacco use in a community-based cohort. *The Journal of the American Medical Association, 275,* 1097–1103.

Institute of Medicine. (1995). *Federal regulation of methadone treatment.* Washington, DC: National Academies Press.

Knapp, J. M., Rosheim, C. L., Meister, E. A., & Kottke, T. E. (1993). Managing tobacco dependence in chemical dependency treatment facilities: A survey of current attitudes and policies. *Journal of Addictive Diseases, 12,* 89–104.

Longo, D. R., Feldman, M. M., Kruse, R. L., Brownson, R. C., Petroski, G. F., & Hewett, J. E. (1998). Implementing smoking bans in American hospitals: Results of a national survey. *Tobacco Control, 7,* 47–55.

McCarthy, W. J., Zhou, Y., Hser, Y. I., & Collins, C. (2002). To smoke or not to smoke: Impact on disability, quality of life, and illicit drug use in baseline polydrug users. *Journal of Addictive Diseases, 21,* 35–54.

NAADAC. (2002). *Position statement: Nicotine dependence.* Retrieved October 19, 2004, from http://naadac.org/documents/display.php?documentid=36

Orleans, C. T., & Hutchinson, D. (1993). Tailoring nicotine addiction treatments for chemical dependency patients. *Journal of Substance Abuse Treatment, 10,* 197–208.

Peto, R., Darby, S., Deo, H., Silcocks, P., Whitley, E., & Doll, R. (2000). Smoking, smoking cessation, and lung cancer in the UK since 1950: Combination of national statistics with two case-control studies. *British Medical Journal, 321,* 323–329.

Richter, K. P., Catley, D., McCool, R. M., Hall, M., & Ahluwalia, J. S. (in press). Dual pharmacotherapy and motivational interviewing for tobacco dependence among drug treatment patients. *Journal of Addictive Diseases.*

Richter, K. P., Choi, W. S., McCool, R. M., Harris, K. J., & Ahluwalia, J. S. (2004). Smoking cessation services in U.S. methadone maintenance facilities. *Psychiatric Services, 55,* 1258–1264.

Sees, K. L., & Clark, H. W. (1993). When to begin smoking cessation in substance abusers. *Journal of Substance Abuse Treatment, 10,* 189–195.

Shoptaw, S., Jarvik, M. E., Ling, W., & Rawson, R. A. (1996). Contingency management for tobacco smoking in methadone-maintained opiate addicts. *Addictive Behaviors, 21,* 409–412.

Shoptaw, S., Rotheram-Fuller, E., Yang, X., Frosch, D., Nahom, D., Jarvik, M. E., Rawson, R. A., & Ling, W. (2002). Smoking cessation in methadone maintenance. *Addiction, 97,* 1317–1328; discussion, 1325.

Substance Abuse and Mental Health Services Administration. (2002). *National survey of substance abuse treatment services (N-SSATS): 2000.* [Data on substance abuse treatment facilities, DASIS series: S-16]. Rockville, MD: Author.

Sullivan, M. A., & Covey, L. S. (2002). Current perspectives on smoking cessation among substance abusers. *Current Psychiatry Reports, 4,* 388–396.

U.S. Department of Health and Human Services. (2000). Regulatory efforts. In *Reducing tobacco use: A report of the surgeon general—2000.* Atlanta, GA: Author, Public Health Service, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Automated Telephone Screening for Problem Drinking*

AMY RUBIN, PH.D.,† JEFFREY P. MIGNEAULT, PH.D.,† LISA MARKS, B.S.,† EDWARD GOLDSTEIN, B.E.E.,† KEVIN LUDENA, B.S.,† AND ROBERT H. FRIEDMAN, M.D.†

Addiction Services WEH 101, Department of Psychiatry, Massachusetts General Hospital, 16 Blossom Street, Boston, Massachusetts 02114

ABSTRACT. Objective: This study assessed test-retest reliability and criterion validity for an automated version of the Alcohol Use Disorders Identification Test (AUDIT), a screening tool for alcohol-related problems. Participants' willingness to use such a system to learn about and change their drinking behavior was also assessed. **Method:** Participants were 202 callers recruited through newspaper ads and flyers asking for volunteers concerned about their drinking and willing to help test a new method of screening and referral for alcohol problems. Participants were divided into two groups. The first group of subjects recruited received the Telephone-Linked Communications (TLC)-AUDIT twice, administered a week apart. The second group received the TLC-AUDIT once and a human-administered AUDIT once, also a week apart. **Results:** Test-retest reliability was assessed in 102 participants; the intraclass correlation of AUDIT scores between both administrations was .87; κ for nonproblem versus problem drinking (AUDIT score of 8 or above) was .89. The validity study compared the TLC-AUDIT scores of the next 100 participants to AUDIT questions administered by a human interviewer. The intraclass correlation was .94; κ was .75. Seventy-five percent of all participants who screened positive for problem drinking agreed they would "talk to a computer again to learn more about your drinking pattern and how to deal with it." **Conclusions:** Automated telephone technology can be used to administer the AUDIT instrument with high levels of reliability and validity. This technology could be used to deliver behavioral change interventions. (*J. Stud. Alcohol* **67:** 454-457, 2006)

ALTHOUGH MORE THAN 18% of the population meet criteria for an alcohol disorder during their lifetime (Grant, 1997) and another 20% are engaged in problem drinking (Grant and Dawson, 1997), many of these people do not seek help (National Institute on Alcohol Abuse and Alcoholism [NIAAA], 2000). Reasons for not seeking help include believing that people should do it on their own, fear that others will find out, fear of being labeled alcoholic, or not being aware of having a problem (Grant, 1997; Higgins-Biddle et al., 1997). To reach these individuals, emphasis has shifted to early detection and intervention in primary-care practices. However, few patients are screened or given care in these settings, in part because busy practitioners are not trained in screening and follow-up care (NIAAA, 2000).

One way to improve screening is through the use of technologies designed for that purpose. Automated or computerized telephone systems can provide a low cost way to screen for alcohol problems. People access these systems by using a device (the telephone) that is familiar, ubiquitous, accessible, and even portable. It can provide access for people with poor reading skills or who speak languages other than English. Research has shown that automated systems are perceived by callers as being more anonymous than therapists or other health professionals and that people tend to be more truthful when reporting on embarrassing or personal issues to a computer than compared with a human professional (Gerbert et al., 1999; Turner et al., 1998).

Computerized assessment for alcohol problems has been shown to be reliable (Bernadt et al., 1989) but has primarily targeted people coming into specialty substance use treatment or research programs (Bernadt et al., 1989; Mundt et al., 2002; Perrine et al., 1995). Automated telephone interviewing using interactive voice response technology has been used to track daily drinking in a research setting. These systems have shown good to excellent validity when compared with both objective and subjective measures, especially for the heaviest drinkers (Mundt et al., 2002; Perrine et al., 1995; Searles et al., 1995). More recently, automated telephone interviewing has been compared favorably to traditional paper methods for collecting data on drinking patterns and medication adherence (Kranzler et al., 2004) as well as alcohol-related expectancies (Collins et al., 2003).

In this study, we developed and tested the reliability, validity, and user acceptance of an automated telephone version of the Alcohol Use Disorders Identification Test (AUDIT), a widely accepted screening tool for problem drinking. In addition, callers who screened positive for

Received: April 18, 2005. Revision: January 10, 2006.

*This study was supported by a grant from the Alcohol Beverage Medical Research Foundation.

†Correspondence may be sent to Amy Rubin at the above address or via email at: arubin1@partners.org. She is also with Harvard Medical School, Boston, MA. Jeffrey P. Migneault, Lisa Marks, Edward Goldstein, Kevin Ludena, and Robert H. Friedman are with the Medical Information Systems Unit, Boston University School of Medicine, and the Boston Medical Center, Boston, MA.

problem drinking were asked for their reactions to the system and whether they would be willing to use it to change their drinking behavior.

Method

Design and development of the automated AUDIT

Telephone-Linked Communications (TLC), a set of computerized telephone technologies, was used to develop the program (Friedman et al., 1997). The TLC-AUDIT program included the following sections in order of presentation: (1) greeting, (2) purpose of TLC-AUDIT, (3) assurance of confidentiality, (4) AUDIT questions, (5) feedback and interpretation of AUDIT results to participants, and (6) closing and referral to the state hotline number for more information or treatment referral. For the purpose of this study, feedback and referral were general and standardized: Participants were told either they appeared to have or not have a drinking problem. All were told the screening was not definitive and were given the recommendation to call their health care provider or the state substance use information and referral line.

Development of TLC-AUDIT required a multistep process that involved mapping out the steps in the interview, writing conversational dialogue for each step that mirrored what a human interviewer would say, writing the computer programs that control the interviewer's statements and recognize the caller's verbal responses, creating the database for storing the users' responses, and having an actor record the interview scripts. Also, because this system used speech recognition, verbal responses that participants might give had to be anticipated so that the software could be "trained" to recognize these responses. This results in a more natural conversation than was possible with older technology, in which answers were "communicated" by pushing buttons on the telephone.

Reviews of the program were conducted in three rounds: by the research team, by 13 alcohol counselors recruited for this purpose, and by a sample of 50 callers recruited to pilot test the program. These 50 callers were anonymously recruited by placing flyers around the medical campus. The responses of the 50 callers were examined for the caller's ability to understand and respond to the system and for the ability of the speech recognition software to understand the caller. Their names or other information were not recorded, and they were not compensated. After each round of review, interview dialogue was modified, and computer software was reprogrammed.

Reliability and validity testing

Participants. Most participants were recruited through ads placed in two metropolitan area newspapers. In an attempt to reach all segments of the community, ads were also placed in nine small local neighborhood papers and one student paper. Flyers were posted on campus, ads were placed on an online bulletin board, and information was faxed to community health centers. Recruitment for the study was completed in 8 months: the first 4 months for the reliability study and the second 4 months for the validity study. The ads invited individuals concerned about their drinking and who were willing to test a new method of screening and referral for alcohol problems to call a toll-free number. The ads stated that participants would receive feedback on their drinking patterns and referral for further information and treatment, if required, as well as a $10 gift certificate to McDonald's or a local supermarket chain.

Respondents were contacted within 1 week of their initial call and screened for the following eligibility criteria: (1) able to use a telephone unassisted, (2) able to speak English, and (3) be 18 years of age or older. The research interviewer explained that, as part of the study, all of their conversations with the computer or with the interviewer would be recorded. The interviewer then asked permission to start tape recording the call. If permission was given, the informed consent was read and discussed, and participant consent was recorded. A copy of the consent form was mailed to participants at a later date.

Measure. The AUDIT is a 10-item alcohol-screening tool developed and used by the World Health Organization (Babor et al., 1992) in multinational trials of brief interventions. The AUDIT has been shown to be generalizable across cultural characteristics, demonstrating good sensitivity and specificity in many populations (Allen et al., 1997; Saunders et al., 1993). The AUDIT produces a total score from 0 to 40, with a score ≥ 8 considered positive for problem drinking (Babor et al., 1992; Cherpitel, 1995). For female participants, Item 3 was modified to read "5 or more drinks" instead of "6 or more drinks," based on currently accepted standards for alcohol consumption in women (NIAAA, 2000).

Procedure. In both studies, the AUDIT was given twice to each participant, with approximately 1 week between administrations (mean [SD] = 8.0 [2.4] days). The scoring and feedback sections of the AUDIT were given after the second administration only, primarily to avoid influencing their responses and to provide incentive for participants to perform the retest. Participants were also asked for additional information, including their willingness to use a similar system for changing their drinking. In the reliability study, participants received the TLC-AUDIT both times. In the validity study, the human-administered AUDIT was used as the "gold standard" comparison for the automated AUDIT. The order of presentation of the two versions of the AUDIT was randomized to control for order effects.

Analysis

Test-retest reliability and criterion validity were examined both for the total TLC-AUDIT score and for the dichotomized score of screening positive for problem drinking. Reliability for the total score was described through the intraclass correlation coefficient (ICC). An ICC above .80 is considered to represent strong agreement, and an ICC between .6 and .8 represents good agreement (Shrout and Fleiss, 1979). Reliability for a positive screen was described through kappa, which measures the agreement beyond chance between ratings on a categorical measure. A κ above .75 is generally considered excellent agreement, and a κ between .40 and .75 is considered moderate to good agreement (Fleiss, 1981). Validity of the TLC-AUDIT compared with a live interviewer was also described through ICC and kappa statistics.

Results

Participants were 37.4 [13.7] years old on average. Forty-nine percent were male, 49% were employed, 69% were single, 45% smoked cigarettes, and 8% were in substance-use treatment. A total of 282 people left messages on the toll-free line. Forty-six participants who left messages were unable to be reached. Seventeen people declined the study, and six subjects were found to be ineligible. Of the 213 subjects who enrolled and completed the initial AUDIT administration, nine participants were lost to follow-up, and two subjects withdrew from the study, leaving 202 participants: 102 in the reliability study and 100 in the validity study.

Test-retest reliability evaluation

On initial administration of the AUDIT, participants' scores ranged from 0 to 37, with a mean of 15.6 [9.7] and a 72.6% scoring in the problem drinking range. Scores on the second administration of the AUDIT ranged from 0 to 36, with a mean of 14.2 [9.2]; 70.6% of the participants scored in the problem drinking range. The ICC between both administrations of the TLC-AUDIT was .87. Kappa for the presence versus absence of problem drinking was .89.

Criterion validity evaluation

On initial administration of the AUDIT, participants' scores ranged from 0 to 40 with an overall mean of 15.5 [9.8]; 73% of the participants scored in the problem drinking range. The second administration of the AUDIT ranged from 0 to 36, with a mean of 13.8 [9.0]; 69% of the participants scored in the problem drinking range. The ICC for scores between the TLC and human administrations of the AUDIT was .94; κ for ratings of the presence or absence of problem drinking was .75. There were no significant differences in the classification of problem drinkers based on the order of method of administration of the AUDIT (e.g., TLC or human first). There was no difference in classification of subjects as problem drinkers by TLC versus human for the first administration of AUDIT ($\chi^2 = 0.87$, 1 df, $p = .35$) or for the second ($\chi^2 = 0.15$, 1 df, $p = .69$). The percentage agreement between human and TLC is shown in Table 1.

TABLE 1. Percentage of agreement between human and Telephone-Linked Communications (TLC) administration of the Alcohol Use Disorders Identification Test

	Human	
TLC	Problem	No problem
Problem	68%	2%
No problem	8%	22%

Potential acceptance of a TLC treatment program

Participants ($n = 148$) who scored in the problem drinking range on the TLC-AUDIT were asked to consider what their goals and methods for changing their drinking behavior might be. Sixty-five percent of the subjects indicated that they would prefer to drink moderately rather than abstain. They were also asked, "If you were to change your drinking habits, what type of help would you consider using?" Seven choices were listed, and they were asked to indicate whether they were willing to use each of them. Sixty percent of the participants stated that they would consider using a computer program for information and advice, and 78% of the participants stated they would consider using a face-to-face counselor. However, when asked, "If you could talk to the computer again to learn more about your drinking pattern and how to deal with it, would you do it?", 75% of the participants said yes. Of that 75% ($n = 111$), 83% were willing to make multiple calls, and 78% were willing to do some work in between calls, such as writing down how much they drank.

Discussion

The results of this study show that an automated telephone system with speech recognition capabilities can be used to administer the AUDIT with high levels of test-retest reliability and criterion validity. These results suggest that a system similar to this one could be offered to the public to provide low-cost anonymous screening for those with questions about their alcohol use. Inexpensive programs can be offered as anonymous learning tools by Public Health departments and Employee Assistance

Programs. When used in health care settings, TLC-AUDIT could also enable health care professionals to focus on those who screen positive as well as allow for screening efforts with much greater reach than is possible with traditional methods.

Another goal of this research was to assess the willingness of problem drinkers to use this technology as a self-help tool to change their behavior. Although these reports of willingness to utilize such a system may be biased somewhat by social desirability, they were encouraging. Reported behavioral intention of utilizing an automated telephone system for problem drinking treatment was high (75%). The authors are currently developing an automated intervention for problem drinkers based on Behavioral Self-Control Training (Miller and Munoz, 2005) that will test the willingness of problem drinkers to utilize this methodology as well as measure its efficacy.

The main limitation of this study is that independent verification of reports of alcohol use at the time of the interview was not collected. However, the AUDIT itself has been shown to measure drinking behavior accurately (Babor et al., 1992). It would also have been appropriate to compare human interviewer test-retest reliability with TLC reliability; however, we did not include that test in our design.

It was puzzling to find that on the second administration of the AUDIT in both studies (whether given by TLC or human administration), average AUDIT scores dropped. This drop, although not significant, may have lowered the kappa in the validity study.

Automated screening, education, and intervention programs for alcohol problems show great promise for improving access to health information, identifying affected individuals, and disseminating empirically based self-help programs. Automated programs can also provide accurate assessment and intervention with total fidelity to manualized instructions, thus simplifying interpretation of results by decreasing error variance. Altogether, computer-based telephone screening for problem drinking could increase the proportion of people who are willing to be screened and increase the probability of intervention and treatment.

Acknowledgment

The authors thank Nuance, Inc. (Burlington, MA) for donating speech recognition software and consultation services.

References

Allen, J.P., Litten, R.Z., Fertig, J.B., and Babor T. A review of research on the Alcohol Use Disorders Identification Test (AUDIT). Alcsm Clin. Exp. Res. **21:** 613-619, 1997.

Babor, T.F., de la Fuente, J.R., Saunders, J., and Grant, M. AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care (Revision), WHO Document No. WHO/PSA/92.4. Geneva, Switzerland: World Health Organization, 1992.

Bernadt, M.W., Daniels, O.J., Blizard, R.A., and Murray, R.M. Can a computer reliably elicit an alcohol history? Brit. J. Addict. **84:** 405-411, 1989.

Cherpitel, C.J. Analysis of cut points for screening instruments for alcohol problems in the emergency room. J. Stud. Alcohol **56:** 695-700, 1995.

Collins, R.L., Kashdan, T.B., and Gollnisch, G. The feasibility of using cellular phones to collect ecological momentary assessment data: Application to alcohol consumption. Exp. Clin. Psychopharmacol. **11:** 73-78, 2003.

Fleiss, J.L. Statistical Methods for Rates and Proportions, 2nd Edition, New York: John Wiley & Sons, 1981.

Friedman, R.H., Stollerman, J.E., Mahoney, D.M., and Rozenblyum, L. The virtual visit: Using telecommunications technology to take care of patients. J. Amer. Med. Informat. Assoc. **4:** 413-425, 1997.

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:** 104-111, 1999.

Grant, B.F. Barriers to alcoholism treatment: Reasons for not seeking treatment in a general population sample. J. Stud. Alcohol **58:** 365-371, 1997.

Grant, B.F. and Dawson, D.A. Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: Results from the National Longitudinal Alcohol Epidemiologic Survey. J. Subst. Abuse **9:** 103-110, 1997.

Higgins-Biddle, J.C., Babor, T.F., Mullahy, J., Daniels, J., and McRee, B. Alcohol screening and brief intervention: Where research meets practice. Conn. Med. **61:** 565-575, 1997.

Kranzler, H.R., Abu-Hasaballah, K., Tennen, H., Feinn, R., and Young, K. Using daily interactive voice response technology to measure drinking and related behaviors in a pharmacotherapy study. Alcsm Clin. Exp. Res. **28:** 1060-1064, 2004.

Miller, W.R. and Muñoz, R.F. Controlling Your Drinking: Tools to Make Moderation Work for You, New York: Guilford Press, 2005.

Mundt, J.C., Bohn, M.J., King, M., and Hartley, M.T. Automating standard alcohol use assessment instruments via interactive voice response technology. Alcsm Clin. Exp. Res. **26:** 207-211, 2002.

National Institute on Alcohol Abuse and Alcoholism. Tenth Special Report to the U.S. Congress on Alcohol and Health: Highlights from Current Research, NIH Publication No. 00-1583, Bethesda, MD: Department of Health and Human Services, 2000.

Perrine, M.W., Mundt, J.C., Searles, J.S., and Lester, L.S. Validation of daily self-reported alcohol consumption using interactive voice response (IVR) technology. J. Stud. Alcohol **56:** 487-490, 1995.

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 Consumption-II. Addiction **88:** 791-804, 1993.

Searles, J.S., Perrine, M.W., Mundt, J.C., and Helzer, J.E. Self-report of drinking using touch-tone telephone: Extending the limits of reliable daily contact. J. Stud. Alcohol **56:** 375-382, 1995.

Shrout, P.E. and Fleiss, J.L. Intraclass correlations: Uses in assessing rater reliability. Psychol. Bull. **86:** 420-428, 1979.

Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Pleck, J.H., and Sonenstein, F.L. Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. Science **280:** 867-873, 1998.

CLINICAL PRACTICE

Unhealthy Alcohol Use

Richard Saitz, M.D., M.P.H.

This Journal *feature begins with a case vignette highlighting a common clinical problem. Evidence supporting various strategies is then presented, followed by a review of formal guidelines, when they exist. The article ends with the author's clinical recommendations.*

A 32-year-old man has a three-month history of difficulty sleeping. On questioning, he mentions that he drinks four to six glasses of wine three to four times per week. How should his case be assessed and managed?

THE CLINICAL PROBLEM

From the Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine and Boston Medical Center; and the Youth Alcohol Prevention Center and the Department of Epidemiology, Boston University School of Public Health — both in Boston. Address reprint requests to Dr. Saitz at Boston Medical Center, 91 E. Concord St. #200, Boston, MA 02118, or at rsaitz@bu.edu.

N Engl J Med 2005;352:596-607.

Each year in the United States, 85,000 deaths, along with substantial disability from medical and psychiatric consequences, injuries, and "secondhand" effects (e.g., motor vehicle crashes), are attributed to the use of alcohol. The estimated annual costs that are attributable to alcohol use are $185 billion.[1,2] Unhealthy alcohol use covers a spectrum that is associated with varying degrees of risk to health (Table 1 and Fig. 1). The prevalence of unhealthy use is 7 to 20 percent or more among outpatients, 30 to 40 percent among patients in emergency departments, and 50 percent among patients with trauma.[11,12] Dependence (alcoholism) is best understood as a chronic disease, with peak onset by the age of 18.[13]

Moderate (i.e., less than risky) use of alcohol may be beneficial, but what constitutes "moderate" depends on age, sex, genetic characteristics, coexisting illnesses, and other factors. Observational studies indicate that for men under the age of 34 years and women under the age of 45 years, those who report no alcohol intake have the lowest mortality. Above these age cutoffs, weekly intakes of no more than five drinks for men or two drinks for women are associated with the lowest mortality.[14] The balance of harm (an increased risk of liver disease, motor vehicle crashes, hypertension, hemorrhagic stroke, and some cancers) and benefit (a reduced risk of ischemic heart disease and ischemic stroke) determines these amounts.

STRATEGIES AND EVIDENCE

IDENTIFICATION

Patients with unhealthy alcohol use often present either asymptomatically, with early-stage problems, or with problems that are not recognized as being alcohol-related. All adults should be screened with a validated survey instrument such as the CAGE questionnaire (where each of the letters in the acronym refers to one of the questions) or the Alcohol Use Disorders Identification Test (AUDIT)[11] (Table 2 and the Supplementary Appendix, available with the full text of this article at www.nejm.org). The CAGE questionnaire is brief but was designed primarily to detect dependence. The AUDIT questionnaire is long but detects the spectrum of unhealthy drinking. Asking questions about consumption (AUDIT questions 1 to 3, question 3 alone, or questions about per-occasion drinking) with or without use of the CAGE questionnaire is a less well validated approach that directly determines the degree of risky drinking.[3,15-17] There may be advantages (including increased truthfulness of patients and efficiency) to embedding

Table 1. Definitions of Unhealthy Alcohol Use.*

Category of Use	Prevalence	Definition and Features
	%	
Risky use	30	For women and persons >65 years of age, >7 standard drinks per week or >3 drinks per occasion; for men ≤65 years of age, >14 standard drinks per week or >4 drinks per occasion; there are no alcohol-related consequences, but the risk of future physical, psychological, or social harm increases with increasing levels of consumption; risks associated with exceeding the amounts per occasion that constitute "binge" drinking in the short term include injury and trauma; risks associated with exceeding weekly amounts in the long term include cirrhosis, cancer, and other chronic illnesses; "risky use" is sometimes used to refer to the spectrum of unhealthy use but usually excludes dependence; one third of patients in this category are at risk for dependence†
Problem drinking	Varies‡	Use of alcohol accompanied by alcohol-related consequences but not meeting ICD-10 or DSM-IV criteria; sometimes used to refer to the spectrum of unhealthy use but usually excludes dependence
Alcohol abuse, harmful use	5	In DSM-IV, recurrence of the following clinically significant impairments within 12 months: failure to fulfill major role obligations, use in hazardous situations, alcohol-related legal problems, or social or interpersonal problems caused or exacerbated by alcohol; in ICD-10, physical or mental health consequences only
Alcohol dependence, alcoholism	4	In DSM-IV, clinically significant impairment or distress in the presence of three or more of the following: tolerance; withdrawal; a great deal of time spent obtaining alcohol, using alcohol, or recovering from its effects; reducing or giving up important activities because of alcohol; drinking more or longer than intended; a persistent desire or unsuccessful efforts to cut down or control use; continued use despite having a physical or psychological problem caused or exacerbated by alcohol; in ICD-10, similar definition

* Data are from the Department of Health and Human Services,[3] Whitlock et al.,[4] the U.S. Preventive Services Task Force,[5] the World Health Organization,[6,7] the American Psychiatric Association,[8] and Grant et al.[9] ICD-10 denotes the *International Classification of Diseases,* 10th edition, and DSM-IV the *Diagnostic and Statistical Manual of Mental Disorders,* 4th edition.

† A standard drink is approximately 12 to 14 g of ethanol, which corresponds to 12 oz of beer, 5 oz of wine, or 1.5 oz of 80-proof liquor. The thresholds in the table do not apply to children, adolescents, or pregnant women; to persons taking medication that interacts with alcohol or engaging in activities that require attention, skill, or coordination (e.g., driving); or those with medical conditions that may be affected by alcohol (e.g., gastritis or hepatitis C). For all these groups, the healthiest choice is generally abstinence. The term "binge drinking" is sometimes used to mean heavy use that is prolonged (>1 day), with cessation of usual activities. It is also used to refer to consumption that exceeds the specified limits per occasion.

‡ Because the definition of problem drinking varies among studies, estimates of the prevalence also vary.

screening for alcohol use in interviews about other health issues, but stand-alone screening is the best-studied approach.[11]

The possibility of unhealthy alcohol use should be routinely considered in patients with hypertension (especially if the condition is difficult to treat), depression, insomnia, abnormal liver-enzyme levels, heartburn, anemia, thrombocytopenia, injury, or problems in social life or at work (e.g., missed work due to hangovers).[18] Approximately half of all cases of cirrhosis, nonischemic cardiomyopathy, pancreatitis, and cancers of the esophagus, larynx, and mouth are attributable to alcohol.[2]

ASSESSMENT AND DIAGNOSIS

Standardized interviews can diagnose alcohol abuse and dependence. Patients should be asked whether they have symptoms of alcohol-use disorders in order to determine the diagnosis, the severity of the problem, and the steps that should be taken to address it (Table 1). The assessment should identify common coexisting conditions and situations in which even a moderate amount of alcohol can be harmful, such as pregnancy; the use of medications that can interact with alcohol; the use of alcohol before situations that require attention, coordination, or skill (e.g., driving); a family history of alcoholism; and the presence of cirrhosis, depression, anxiety,[19] personality disorders (particularly antisocial and histrionic personality),[20] or other conditions that are potentially exacerbated by alcohol.[3]

INTERVENTION

Detoxification

Among patients who consume approximately 20 standard alcoholic drinks per day, symptomatic withdrawal is likely with abstinence[21]; however, reported consumption is an imperfect predictor of symptoms associated with withdrawal. Withdrawal can lead to seizures, delirium tremens, or death. However, most often it is mild and easily managed.

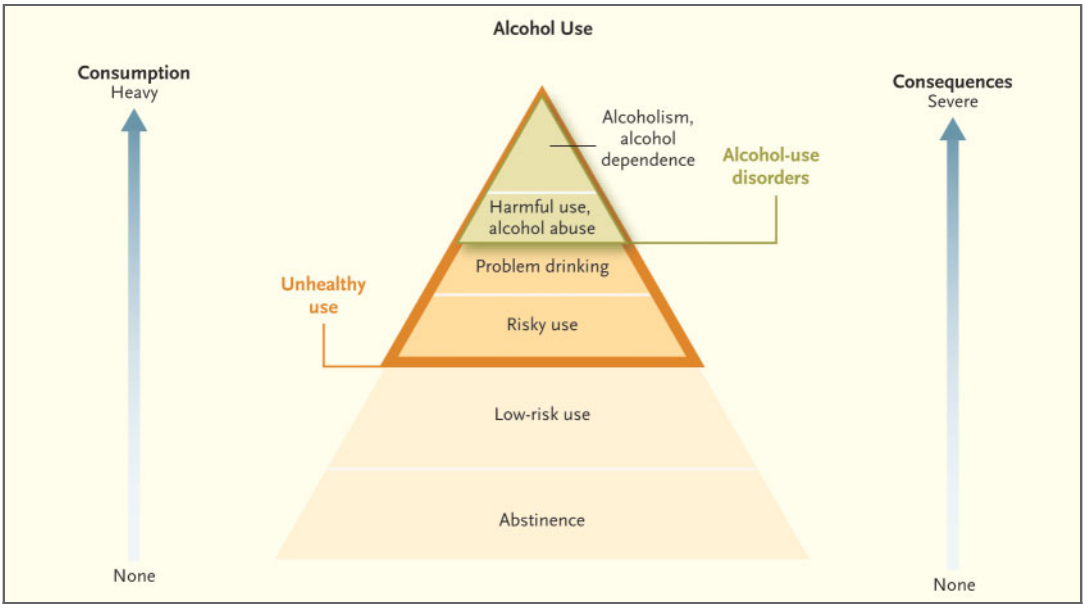


Figure 1. The Spectrum of Alcohol Use.

The spectrum of alcohol use extends from abstinence and low-risk use (the most common patterns of alcohol use) to risky use, problem drinking, harmful use and alcohol abuse, and the less common but more severe alcoholism and alcohol dependence.[10] Consumption and the severity of consequences increase from low-risk use through dependence. The areas of the pyramid reflect the approximate prevalence of each category. Clinicians and public health practitioners should be most concerned with the categories in the shaded upper portions of the pyramid (representing unhealthy alcohol use).

Benzodiazepines are the only medications proven to ameliorate symptoms and decrease the risk of seizures and delirium tremens; they are routinely indicated for patients with substantial symptoms of withdrawal and those at increased risk for complications (due to coexisting acute illnesses or a history of withdrawal seizures) (Table 3).[22] Ethanol should not be used to treat withdrawal.

Brief Intervention

"Brief intervention" generally refers to 10 to 15 minutes of counseling, with feedback about drinking, advice and goal setting, and follow-up contact (one or more discussions lasting 10 to 15 minutes with a clinician) (Table 4). Randomized trials in diverse settings (e.g., primary care facilities, emergency departments, hospitals, and colleges) have demonstrated that such brief interventions can decrease drinking and its consequences at six-month follow-up or later, with a reduction of 10.5 percent in the prevalence of risky drinking and a reduction in the intake of alcohol of three to nine drinks per week, as compared with no intervention.[4,26,27] Single five-minute contacts appear to be less effective. When such a strategy is used with patients who are not seeking treatment, efficacy is limited to those without alcohol dependence.[26]

One randomized trial compared the result of being given a booklet about general health topics (control group) with that of receiving a typical brief intervention (two discussions with a primary care physician, followed by two telephone calls from a nurse).[28] At one year, the brief intervention had led to greater reductions in self-reported drinking (from 19 to 12 drinks per week, vs. a reduction from 19 to 16 drinks per week in the control group) and in binges (from six to three binges, vs. a reduction from five to four binges per month among the controls). At three to four years, the intervention group was less likely to be engaged in risky drinking (prevalence, 23 percent, vs. 35 percent in the control group) and had spent fewer days in the hospital and had lower associated costs (a difference of $7,780 per patient) — all significant differences as compared with the control group. There were also fewer deaths in the intervention group (three, vs. seven among the controls), although this difference was not statistically significant.

Another study assessed the long-term effects of a brief intervention among middle-aged male drink-

ers who were selected on the basis of high serum levels of γ-glutamyltransferase. The intervention consisted of a monthly visit with a nurse and a quarterly visit with a physician for 18 to 48 months, including feedback regarding the importance of the patient's γ-glutamyltransferase levels and advice that the patient should restrict the use of alcohol. At the 16-year follow-up, alcohol-related mortality was lower in the group that received the intervention than in a group of patients who simply received a letter informing them of the results of the blood test and advising a 2-year follow-up (4 percent vs. 7 percent).[29]

Brief interventions should include counseling patients about setting a goal for a reduction in alcohol consumption and ways to achieve that goal (Table 4). Interventions may be effective regardless of a patient's readiness to change, but understanding the patient's perception of the problem and whether he or she is ready for change is considered to be important. Motivational-interviewing approaches (which emphasize empathic listening and the autonomy of patients in their own decision making and encourage people to identify their own reasons for change) have been shown to be more effective in reducing drinking than confrontational counseling (which imposes on the patient the clinician's view of the problem, minimizes the patient's perspective, and forces the patient to admit to having a problem).[30]

Treatment for Dependence

Data from observational and clinical studies indicate that with treatment for alcohol dependence (behavioral or pharmacologic), two thirds of patients have a reduction in the consequences of alcohol consumption (e.g., alcohol-related injury or job loss) and the amount of consumption (by more than 50 percent) after one year; one third of patients who are treated are either abstinent or drink moderately without consequences.[31] All patients with alcohol dependence should be offered treatment. Controlled studies that have compared the results of recommendations by physicians that patients cut down their alcohol consumption with those of recommendations that patients abstain did not find differences in drinking outcomes,[32] and no more than 11 percent of people with alcohol dependence achieved controlled drinking in the long term.[33] Patients with alcohol dependence who are not ready to begin treatment may still benefit from referral to a specialist for confirmation of the diagnosis and recommendations.

Counseling

Effective treatment for alcohol dependence can be provided in the outpatient setting. Patients who have little social support, who have environments that are not supportive of recovery, or who have complex coexisting medical or psychiatric illnesses may need to be removed from environments in which alcohol is likely to be used.[34]

Cognitive behavioral therapy, 12-step facilitation, and motivational-enhancement therapy (in weekly sessions) are effective treatments that are detailed in written guides for therapists.[35] Cognitive behavioral therapy emphasizes the learning of skills to cope with situations that precipitate heavy drinking.[36] Twelve-step facilitation emphasizes the concept of alcoholism as a disease and active involvement in Alcoholics Anonymous (AA).[37] Motivational-enhancement therapy is motivational interviewing as outlined in written guides.[38] A large clinical trial that randomly assigned patients with alcohol dependence to these treatments showed that they had similar efficacy. At the one-year follow-up, abstinence was reported on 85 percent of days in all three groups on average, as compared with 20 to 30 percent of days at the time the study began; at three years, two thirds of the patients were abstinent. In addition, in all groups the proportions of patients who had a relapse of heavy drinking, depression, alcohol-related problems, and other drug use were reduced, as were liver-enzyme levels.

Self-Help

Publications outlining self-help strategies to decrease drinking on the basis of the principles of cognitive behavioral therapy also have proven efficacy. In a randomized trial that compared the results of group or individual sessions designed to encourage self-control with the results of use of a book outlining the same principles, alcohol consumption was similarly reduced in the two groups at 12 months.[39] In another randomized trial, the consumption of alcohol above recommended limits was significantly less frequent at the six-month follow-up among drinkers who received a self-help manual, as compared with those who received a booklet with general information and advice (53 percent vs. 78 percent, respectively).[40]

Mutual Help

AA is a fellowship that provides support, at no charge, for people who want to stop drinking. This approach is appropriate for most persons with al-

Table 2. Screening Tests for Unhealthy Alcohol Use.*

Test or Question	Score
CAGE questionnaire	
Have you ever felt you should cut down on your drinking?	
Have people annoyed you by criticizing your drinking?	
Have you ever felt bad or guilty about your drinking?	
Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover (eye opener)?	
Alcohol Use Disorders Identification Test (AUDIT)	
The following questions are about your use of alcoholic beverages in the past year. Questions refer to standard drinks.†	
How often do you have a drink containing alcohol?	
Never	0
Monthly or less	1
2 to 4 times a month	2
2 to 3 times a week	3
4 or more times a week	4
How many drinks containing alcohol do you have on a typical day when you are drinking?	
1 or 2	0
3 or 4	1
5 or 6	2
7 to 9	3
10 or more	4
How often do you have 6 or more drinks on one occasion?	
Never	0
Less than monthly	1
Monthly	2
Weekly	3
Daily or almost daily	4
How often during the past year have you found that you were not able to stop drinking once you had started?	
Never	0
Less than monthly	1
Monthly	2
Weekly	3
Daily or almost daily	4
How often during the past year have you failed to do what was normally expected from you because of drinking?	
Never	0
Less than monthly	1
Monthly	2
Weekly	3
Daily or almost daily	4
How often during the past year have you needed a drink in the morning to get yourself going after a heavy drinking session the previous night?	
Never	0
Less than monthly	1
Monthly	2
Weekly	3
Daily or almost daily	4

coholism, except perhaps for those who have great difficulty with social interaction or for those with less severe dependence; however, even those with poor social skills may benefit from the alcohol-free social network.

Evidence for the effectiveness of AA comes primarily from observational studies of individual and group counseling based on 12-step principles[35,41] and of AA involvement.[42] Follow-up of military veterans revealed a higher frequency of abstinence at 12 months among those participating in 12-step programs than among those participating in programs with a cognitive behavioral orientation (26 percent vs. 19 percent).[41] Participation in AA (by attending meetings and having a sponsor) has been associated with increased rates of abstinence seven months after inpatient treatment, as compared with nonparticipation.[42] However, AA may be inferior to inpatient treatment. In a randomized trial comparing these two approaches among persons with alcohol-use disorders, hospitalization in the subsequent year was significantly less common among

Table 2. (Continued.)

Test or Question	Score
The following questions are about your use of alcoholic beverages in the past year. Questions refer to standard drinks.†	
How often during the past year have you had a feeling of guilt or remorse after drinking?	
Never	0
Less than monthly	1
Monthly	2
Weekly	3
Daily or almost daily	4
How often during the past year have you been unable to remember what happened the night before because you had been drinking?	
Never	0
Less than monthly	1
Monthly	2
Weekly	3
Daily or almost daily	4
Have you or someone else been injured as a result of your drinking?	
No	0
Yes, but not in the past year	2
Yes, during the past year	4
Has a relative, friend, or doctor or other health worker been concerned about your drinking or suggested you cut down?	
No	0
Yes, but not in the past year	2
Yes, during the past year	4
Screening question about per-occasion consumption	
For women: When was the last time you had more than 4 drinks in one day? For men: When was the last time you had more than 5 drinks in one day?	
Screening questions about consumption	
On average, how many days per week do you drink alcohol? On a typical day when you drink, how many drinks do you have? What is the maximum number of drinks you had on any given occasion during the past month?	

* Cutoff scores with reasonable sensitivity and specificity for unhealthy alcohol use are as follows: CAGE, one or two positive responses (sensitivity, 53 to 92 percent; specificity, 81 to 95 percent); AUDIT, score of 8 or more (sensitivity, 51 to 97 percent; specificity, 78 to 96 percent); AUDIT-C (first three questions, about consumption), score of 4 or more (sensitivity, 86 percent; specificity, 72 percent); AUDIT question 3 ("How often do you have 6 or more drinks on one occasion?"), score of 1 or more (sensitivity, 77 percent; specificity, 83 percent); screening question about per-occasion consumption, "in the past three months" (sensitivity, 62 to 86 percent; specificity, 86 to 93 percent) (see Supplementary Appendix). The CAGE and consumption screening questions can be used in combination; this seven-question test is considered positive if the results exceed either the cutoffs for "risky drinking" or there is an affirmative answer to any of the CAGE questions (sensitivity, 83 percent; specificity, 84 percent). Laboratory tests (e.g., levels of γ-glutamyltransferase [sensitivity, 65 percent] and carbohydrate-deficient transferrin [sensitivity, ≤60 percent]) are not more sensitive than are validated screening questionnaires and need to be followed by questions about alcohol use. As such, the tests have unknown incremental value. Questions regarding consumption and an additional interview are required to assess patients whose results on the screening tests are positive to identify the amounts and consequences of risky drinking.

† A standard drink is approximately 12 to 14 g of ethanol, which corresponds to 12 oz of beer, 5 oz of wine, or 1.5 oz of 80-proof liquor.

those who had been initially assigned to inpatient treatment than among those assigned to participate in AA (23 percent vs. 63 percent).[43]

AA involves a belief in a "higher power," a term that does not necessarily refer to a deity but rather to any power greater than oneself. AA supports the use of medications for alcohol dependence (as described below), but some members may disapprove of such a strategy. Meeting types vary (e.g., closed or open and with smoking permitted or not), and schedules are available locally (www.alcoholics-anonymous.org).[25]

Al-Anon, Alateen (for teenagers), and Adult Children of Alcoholics can help family and friends understand alcoholism and not feel responsible for the illness. In a study in which "concerned significant others" were randomly assigned to participate in various strategies to engage problem drinkers in treatment (one being an approach based on Al-Anon), all strategies led to improvements in the

Table 3. Pharmacotherapy for the Treatment of Alcohol Dependence.

Medication	Presumed Mechanism of Action	Dose	Side Effects	Comments
For detoxification or treatment of withdrawal				
Benzodiazepines (diazepam, chlordiazepoxide, lorazepam)	Decrease hyperautonomic state by facilitating inhibitory γ-aminobutyric acid receptor transmission, which is down-regulated by long-term exposure to alcohol	Diazepam, 10–20 mg; chlordiazepoxide, 50–100 mg; lorazepam, 2–4 mg every 1–2 hr until symptoms subside (e.g., CIWA-Ar score <8) for 24 hr*	Sedation	Administer every 1–2 hr until symptoms subside; no tapering necessary for long-acting drugs (e.g., diazepam); lorazepam preferable for elderly patients and those with hepatic synthetic dysfunction or at high risk for respiratory failure; if other short-acting benzodiazepines are used (e.g., oxazepam) or if there is concern that frequent reassessment will not occur, add a dose 4 times daily for 24 hr, followed by half a dose 4 times daily for 48 hr; reassessment of withdrawal symptoms is advisable 1–2 hr after every dose; daily assessments by a clinician are recommended for outpatients (with the patient, a responsible other person, or both assessing responses to each dose)
For treatment of alcohol dependence to prevent relapse				
Naltrexone (ReVia)	Acts as an opiate agonist; decreases heavy drinking by blocking endogenous opioids, a process that attenuates craving and the reinforcing effects of alcohol	Initial dose, 12.5 mg daily or 25 mg daily; therapeutic dose, 50 mg daily	Nausea, headache, dizziness, nervousness, fatigue, insomnia, vomiting, anxiety, somnolence, dry mouth, dyspepsia; elevated liver-enzyme levels (dose-related)	Check liver enzymes or symptoms periodically; contraindicated in patients with current opioid dependence or need for opioids; relatively contraindicated in patients with hepatitis (liver-enzyme levels 3 times the upper limit of the normal range) or cirrhosis
Acamprosate (Campral)†	Increases abstinence by stabilizing activity in the glutamate system, which is affected by long-term heavy consumption	666 mg 3 times a day	Diarrhea	Contraindicated in patients with renal insufficiency (creatinine clearance ≤30 ml/min); half a dose in those with creatinine clearance >30–50 ml/min
Disulfiram (Antabuse)	Blocks aldehyde dehydrogenase; blockade allows acetaldehyde to accumulate with alcohol consumption, causing unpleasant symptoms (e.g., flushing, headache, vomiting, dyspnea, confusion)	Initial dose, 250 mg daily; therapeutic dose, 500 mg daily	Idiosyncratic fulminant hepatitis, neuropathy (at doses >500 mg), psychosis, and symptoms that generally resolve on discontinuation of drug (headache, drowsiness, fatigue, rash, pruritus, dermatitis, garlicky taste in mouth)	Risk of complications: increased ethanol reaction in patients who have coronary artery disease, who are receiving treatment for hypertension, or who have esophageal varices; contraindicated in patients who have a limited capacity to understand consequences of alcohol use, who have allergies (to rubber [thiuram derivatives], cobalt, or nickel) or who are pregnant (fetal limb abnormalities reported); supervised dosing has best documented efficacy; check liver-enzyme levels or symptoms periodically; use a higher dose if no ethanol reaction at lower dose (testing for the reaction not necessary)

* CIWA-Ar denotes Clinical Institute Withdrawal Assessment for Alcohol, revised. The scale assesses 10 domains (nausea or vomiting; anxiety; tremor; sweating; auditory, visual, and tactile disturbances; headache; agitation; and clouding of sensorium) and assigns 0 to 7 points for each item except for the last item, which is assigned 0 to 4 points, with a total possible score of 67. This scale has been validated as a measure to assess the severity of alcohol withdrawal. Higher scores indicate a higher risk of complications; patients receiving scores of 8 or more should be treated.[22]

† Acamprosate was approved by the Food and Drug Administration in July 2004 and is now available in the United States.

Table 4. Brief Counseling and Referral.*

How to Advise or Refer Patients	Examples or Explanations
Elicit information about how the patient views the problem.	"What do you think about your drinking? Are you ready to make a change in your alcohol use? How confident are you that you could cut down if you wanted to?"
Express concern and provide clear advice regarding the ideal goal (abstinence or reduced consumption for those with nondependent alcohol use, achieved through brief counseling; abstinence for patients with alcohol dependence).†	"I am concerned about your drinking; my medical advice is that the healthiest choice for you is to cut down or abstain."
Provide specific feedback about alcohol consumption in comparison with population norms, and link existing problems to alcohol use when appropriate, to make information relevant to the patient.	"Ninety-three percent of adults drink less than the amounts you report drinking. You mentioned your heartburn is worse when you drink. Alcohol is probably causing your heartburn."
Express empathy, let the patient know you believe that change is possible, and acknowledge that it is the patient's responsibility to change.	"The fact you were able to quit before for a week tells me you can do it again. But it must be difficult. It is up to you to make these changes."
When the patient expresses interest or gives permission, provide information, including a menu of options, about how to change.	"Would you like information on how to cut down or abstain? Other people have found a range of options helpful, such as keeping a drinking diary, counseling, and mutual-help groups. What do you think about these?"
Anticipate and discuss situations in which the patient feels at risk for drinking excessively, and talk about strategies to avoid drinking excessively.	"What ways might help you avoid drinking excessively when you go out with friends who drink?" Have the patient keep a drinking diary (including the number of drinks consumed per day).
Schedule a follow-up session to assess drinking and changes in alcohol use.	"Please think about your drinking and the health risks we discussed; contact me if you decide you would like assistance in the future. Let's schedule a follow-up visit in a month to talk again." In the follow-up, review the drinking goal, the actual drinking history, and any consequences since the last visit. If the serum levels of γ-glutamyltransferase or carbohydrate-deficient transferrin were initially abnormal, monitor levels.
For patients who are not ready to change their alcohol use, advice about changing their habits or getting help is counterproductive because the patient will enumerate the reasons against change; avoid confrontation and argument.	
Elicit the patient's own reasons for drinking, reasons for not drinking, and concerns about changing.	"What do you like about drinking? What do you like to drink? What are some problems you have noticed when or after you drink? What would it be like not to drink?"
For patients with alcohol dependence, provide brief counseling with the goal of increasing motivation to change; the recommended change is abstinence and linkage with any or all known effective interventions (mutual-help groups, pharmacotherapy, and counseling).‡	Consider referral to a specialist (a physician who specializes in addiction medicine or an alcoholism-treatment provider) for evaluation and confirmation of the diagnosis, even if the patient is not ready to begin treatment.
Know local referral options, such as health plan referral services, public treatment resources, physicians, other counselors, employee-assistance programs, and national resources (in the United States, http://findtreatment.samhsa.gov); know what patients can expect when they seek assistance.§	Help the patient take the first step (e.g., make an appointment); follow up on treatment entry and engagement.
For patients in recovery, address plans for what to do in the event of relapse.¶	"What would you do if you felt your drinking was out of control?"

* Data are from the Department of Health and Human Services[3] and the U.S. Preventive Services Task Force.[5] This model includes a recommended structure for effective discussions about changing health behavior (elicit–provide–elicit).[23] The elements of brief interventions with proven efficacy include feedback, responsibility, advice, a menu of options, empathy, and support of self-efficacy.

† Patients may need additional assistance if their goal is not achieved. Patients who are pregnant or trying to conceive, who have a medical condition that would be worsened by drinking, or who are taking a medication that interacts with alcohol should be advised to abstain. Discussions about alcohol use with patients who report no current consequences of drinking are analogous to discussions about other risk factors (e.g., hypercholesterolemia and physical inactivity).

‡ Some generalist physicians who have expertise, availability, and adequate office support may choose to provide treatment rather than refer the patient to a specialist. Many patients will not be ready for referral. In such cases, a reasonable option would be brief counseling to help the patient abstain or, if the patient declines, to reduce consumption, with a follow-up session to assess progress. This is a reasonable option that provides information for both the patient and the physician about what intervention will be required.

§ Assistance that is commonly available by referral includes outpatient and inpatient detoxification, mutual-help groups (Alcoholics Anonymous and alternatives such as Self-Management and Recovery Training [SMART], Secular Organizations for Sobriety, Moderation Management, Rational Recovery, and Women for Sobriety [links available at www.mentalhelp.net/selfhelp]),[24] mutual help for relatives (Al-Anon, Alateen, and Adult Children of Alcoholics), outpatient counseling, inpatient treatment (including counseling, mutual help, and a sober environment for persons with coexisting illnesses or those for whom outpatient treatment is not successful), and sober living environments.

¶ More information on this topic is available in Friedmann et al.[25]

functioning of the significant others and in the quality of the relationship between the family member and the person with the drinking problem.[44]

Pharmacotherapy

Naltrexone, acamprosate, and disulfiram have reduced heavy drinking and increased abstinence in randomized trials of patients with alcohol dependence, with pharmacotherapy generally lasting 3 to 12 months. Information regarding mechanisms, dosing, and side effects is summarized in Table 3.[45,46] A meta-analysis showed that in placebo-controlled, randomized trials of a short duration (three months or less), naltrexone decreased the risk of a return to heavy drinking from 48 percent to 37 percent, and decreased drinking days by 4.5 percent; the proportion of patients who were abstinent was higher with naltrexone (35 percent, vs. 30 percent with placebo), but this finding was of borderline significance.[46] In one study,[47] even though the decrease in the proportion of patients who had a relapse with naltrexone was not significant (odds ratio, 0.75; 95 percent confidence interval, 0.53 to 1.08), the point estimate was consistent with those of other studies.[46] In addition, this study included a severely affected population that may have required more intensive therapy (male veterans with long-standing alcoholism, most not married and many disabled).

A meta-analysis of placebo-controlled trials lasting 3 to 24 months showed that acamprosate increased the proportion of patients who were abstinent (from 15 percent to 23 percent).[46] In a single-blind, 12-month study comparing naltrexone with acamprosate, the percentage of patients who reported no heavy drinking was higher with naltrexone than with acamprosate (41 percent vs. 17 percent). For the most recent six months, abstinence was reported by 54 percent and 27 percent, respectively, and percentages of days with heavy drinking were 33 percent and 53 percent, respectively.[46] Another trial comparing the combination of the drugs with either drug alone found the combination to be as safe and more effective.[48] Most efficacy studies of naltrexone and acamprosate have required detoxification first,[46] but two controlled trials found naltrexone to be effective even when patients were not abstinent before starting to take the medication.[46,49]

Controlled studies suggest that disulfiram can decrease the number of drinking days.[45] In small, controlled studies, administration of disulfiram under the supervision of another person improved abstinence as compared with unsupervised use.[50] In a six-month controlled trial (in which supervised administration of vitamin C was used as the control), supervised administration of disulfiram resulted in a greater increase in the number of abstinent days.[51] Abstinence is required before disulfiram therapy is started.

Counseling should be provided with pharmacotherapy, and primary care management is at least as effective as cognitive behavioral therapy when combined with pharmacotherapy. Primary care management, as tested in randomized trials, includes review of the patient's medical and alcohol-use history; development of a treatment plan with the patient; review of advice, medication issues, and goals for follow-up; referral to AA; and a follow-up session of 15 to 20 minutes every one to two weeks with a physician, nurse practitioner, or physician assistant to discuss adherence to the drug regimen, alcohol use, and any adverse effects of the drug regimen.[52]

Pharmacotherapy for Coexisting Psychiatric Conditions

Although a detailed review of the treatment of coexisting psychiatric illnesses is beyond the scope of this article, data from randomized trials suggest that pharmacotherapy with antidepressant or anxiolytic agents can decrease alcohol consumption. Increased time to a resumption of heavy drinking has been reported in a study of patients with coexisting anxiety who were treated with buspirone[53] and in a study of patients with a coexisting major depression who were treated with desipramine[54] or fluoxetine.[55] The selective serotonin-reuptake inhibitors citalopram (Celexa) and fluvoxamine (Luvox) have also been reported to increase the proportion of patients who are abstinent among those who do not have depression.[56]

AREAS OF UNCERTAINTY

Although screening for unhealthy alcohol use is routinely recommended, there are limited data that show improvements in clinical outcomes after implementation of screening. Despite good evidence to support brief intervention, some observers have questioned its effectiveness and value in practice.[27] Limited data suggest that brief interventions have benefits beyond decreased consumption and are

cost-effective.[4,26-29,57] Widespread implementation of brief intervention in clinical practice remains a challenge.

Promising strategies, such as additional brief counseling sessions for nondependent, unhealthy drinkers and treatment either with medications in doses as needed for craving[49,58] or with more than one medication, require study. The role of new medications for treating alcohol dependence — including ondansetron,[59] topiramate,[60] and depot preparations of naltrexone[61] — remains unclear. Data are limited to guide decisions regarding the type of therapy, the necessary duration and timing of treatments in relation to detoxification,[46,49] management in the context of other drug use, and the use of less sedating medications to manage withdrawal.

GUIDELINES

The U.S. Preventive Services Task Force recommends routine screening for unhealthy alcohol use with the use of the AUDIT or CAGE questionnaires in primary care settings. The group also recommends brief counseling interventions in primary care settings to reduce alcohol misuse and referral to specialty treatment for those with alcohol dependence.[5] The American Society of Addiction Medicine recommends the administration of benzodiazepines for the management of alcohol withdrawal and has published criteria for recommending specialty care.[22,34]

CONCLUSIONS AND RECOMMENDATIONS

Unhealthy alcohol use can and should be identified with the use of questions validated for this purpose (the AUDIT or CAGE questionnaires or validated questions about alcohol consumption). Asking questions in a matter-of-fact way in the context of the general health history can facilitate discussion of what can be a sensitive topic. For the patient who was described in the vignette, the consumption of alcohol — both per occasion and per week — poses health risks; his sleep disturbance may well be related to his drinking. The patient should be assessed for additional consequences (e.g., depression and hypertension) and symptoms of dependence. Brief counseling should be provided; the counselor should make explicit the relationship between drinking and health consequences, assess the patient's readiness to change, advise him to cut down on alcohol consumption (for nondependent use) or to abstain and obtain specialized treatment (for dependent use), negotiate a plan for reducing consumption, and follow up (at least once and as needed thereafter).

After detoxification, all patients with alcohol dependence should receive treatment from someone with expertise in the field. That treatment should include medication and counseling (on the basis of local availability but favoring a reproducible, tested approach), participation in AA, and weekly follow-up for a month with decreasing frequency thereafter to assess drinking, consequences, medication use, counseling, and participation in AA. Either naltrexone or acamprosate is first-line therapy; naltrexone is the better choice if the patient has not abstained from drinking for at least three to five days. Disulfiram is an alternative that works best when dosing is supervised.

Supported by grants from the National Institute on Alcohol Abuse and Alcoholism (R-01 AA12617, R-01 13216, R-01 13304, R-25 13822, and P60 AA013759), from the National Heart, Lung, and Blood Institute (K30 HL04124), and from the Center for Substance Abuse Treatment at the Substance Abuse and Mental Health Services Administration (04-3314093).

I am indebted to colleagues for discussing an umbrella term to encompass risky drinking and alcohol-use disorders and to Jeffrey H. Samet for critical review of the manuscript.

REFERENCES

1. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. JAMA 2004;291:1238-45.
2. Harwood HJ. Updating estimates of the economic costs of alcohol abuse in the United States: estimates, update methods, and data. Bethesda, Md.: National Institute on Alcohol Abuse and Alcoholism, 2000.
3. Helping patients with alcohol problems: a health practitioner's guide. Rockville, Md.: National Institute on Alcohol Abuse and Alcoholism, January 2003. (NIH publication no. 03-3769.)
4. Whitlock EPM, Polen MRM, Green CAP, Orleans TP, Klein JM. Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2004; 140:557-68.
5. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. Ann Intern Med 2004;140:554-6.
6. Lexicon of alcohol and drug terms published by the World Health Organization. (Accessed January 18, 2005, at http://www.who.int/substance_abuse/terminology/who_lexicon/en/.)
7. The ICD-10 classification of mental and behavioral disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization, 1992.
8. Diagnostic and statistical manual of mental disorders, 4th ed.: DSM-IV. Washington, D.C.: American Psychiatric Association, 1994.
9. Grant BF, Dawson DA, Stinson FS, Chou SP, Dufour MC, Pickering RP. The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991-1992 and 2001-2002. Drug Alcohol Depend 2004;74:223-34.
10. 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, D.C.: National Academy Press, 1990.
11. Fiellin DA, Reid MC, O'Connor PG. Screening for alcohol problems in primary care: a systematic review. Arch Intern Med 2000;160:1977-89.
12. D'Onofrio G, Bernstein E, Bernstein J, et al. Patients with alcohol problems in the emergency department, part 1: improving detection. Acad Emerg Med 1998;5:1200-9.
13. Li TK, Hewitt BG, Grant BF. Alcohol use disorders and mood disorders: a National Institute on Alcohol Abuse and Alcoholism perspective. Biol Psychiatry 2004;56:718-20.
14. White IR, Altmann DR, Nanchahal K. Mortality in England and Wales attributable to any drinking, drinking above sensible limits and drinking above lowest-risk level. Addiction 2004;99:749-56.
15. Bradley KA, Bush KR, Epler AJ, et al. Two brief alcohol-screening tests from the Alcohol Use Disorders Identification Test (AUDIT): validation in a female Veterans Affairs patient population. Arch Intern Med 2003;163:821-9.
16. Williams R, Vinson DC. Validation of a single screening question for problem drinking. J Fam Pract 2001;50:307-12.
17. 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.
18. Saitz R. Overview of medical and surgical complications. In: Graham AW, Schultz TK, Mayo-Smith MF, Ries RK, Wilford BB, eds. Principles of addiction medicine. 3rd ed. Chevy Chase, Md.: American Society of Addiction Medicine, 2003:1027-52.
19. 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;61:807-16.
20. Grant BF, Stinson FS, Dawson DA, Chou SP, Ruan WJ, Pickering RP. Co-occurrence of 12-month alcohol and drug use disorders and personality disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch Gen Psychiatry 2004;61:361-8.
21. Isbell H, Fraser HF, Wikler A, Belleville RE, Eisenman AJ. An experimental study of the etiology of rum fits and delirium tremens. Q J Stud Alcohol 1955;16:1-33.
22. Mayo-Smith MF. Pharmacological management of alcohol withdrawal: a meta-analysis and evidence-based practice guideline. JAMA 1997;278:144-51.
23. Rollnick S, Mason P, Butler C. Health behavior change: a guide for practitioners. Edinburgh: Churchill Livingstone, 1999.
24. Horvath AT. Alternative support groups. In: Lowinson JH, Ruiz P, Millman RB, Langrod JG, eds. Substance abuse: a comprehensive textbook. 3rd ed. New York: Lippincott Williams & Wilkins, 1997:390-6.
25. Friedmann PD, Saitz R, Samet JH. Management of adults recovering from alcohol or other drug problems: relapse prevention in primary care. JAMA 1998;279:1227-31.
26. 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.
27. 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.
28. Fleming MF, Mundt MP, French MT, Manwell LB, Stauffacher EA, Barry KL. Brief physician advice for problem drinkers: long-term efficacy and benefit-cost analysis. Alcohol Clin Exp Res 2002;26:36-43.
29. Kristenson H, Osterling A, Nilsson JA, Lindgarde F. Prevention of alcohol-related deaths in middle-aged heavy drinkers. Alcohol Clin Exp Res 2002;26:478-84.
30. Miller WR, Rollnick S. Motivational interviewing: preparing people for change. 2nd ed. New York: Guilford Press, 2002.
31. Miller WR, Walters ST, Bennett ME. How effective is alcoholism treatment in the United States? J Stud Alcohol 2001;62:211-20.
32. Sobell MB, Sobell LC. Controlled drinking after 25 years: how important was the great debate? Addiction 1995;90:1149-53.
33. Vaillant GE. A 60-year follow-up of alcoholic men. Addiction 2003;98:1043-51.
34. ASAM Patient Placement Criteria for the treatment of substance-related disorders (PPC-2R). 2nd ed. Chevy Chase, Md.: American Society of Addiction Medicine, 2001.
35. Project MATCH Research Group. Project MATCH secondary a priori hypotheses. Addiction 1997;92:1671-98.
36. Kadden R, Carroll KM, Donovan D, et al. Cognitive-behavioral coping skills therapy manual: a clinical research guide for therapists treating individuals with alcohol abuse and dependence. Project MATCH monograph series. Vol. 3. Rockville, Md.: National Institute on Alcohol Abuse and Alcoholism, 1995. (NIH publication no. 94-3724.)
37. Nowinski J, Baker S, Carroll KM. Twelve step facilitation therapy manual: a clinical research guide for therapists treating individuals with alcohol abuse and dependence. Project MATCH monograph series. Vol. 1. Rockville, Md.: National Institute on Alcohol Abuse and Alcoholism, 1995. (NIH publication no. 94-3722.)
38. 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. Vol. 2. Rockville, Md.: National Institute on Alcohol Abuse and Alcoholism, 1995. (NIH publication no. 94-3723.)
39. Miller WR, Taylor CA. Relative effectiveness of bibliotherapy, individual and group self-control training in the treatment of problem drinkers. Addict Behav 1980;5:13-24.
40. Heather N, Kissoon-Singh J, Fenton GW. Assisted natural recovery from alcohol problems: effects of a self-help manual with and without supplementary telephone contact. Br J Addict 1990;85:1177-85.
41. Ouimette PC, Finney JW, Moos RH. Twelve-step and cognitive-behavioral treatment for substance abuse: a comparison of treatment effectiveness. J Consult Clin Psychol 1997;65:230-40.
42. Montgomery HA, Miller WR, Tonigan JS. Does Alcoholics Anonymous involvement predict treatment outcome? J Subst Abuse Treat 1995;12:241-6.
43. Walsh DC, Hingson RW, Merrigan DM, et al. A randomized trial of treatment options for alcohol-abusing workers. N Engl J Med 1991;325:775-82.
44. Miller WR, Meyers RJ, Tonigan JS. Engaging the unmotivated in treatment for alcohol problems: a comparison of three strategies for intervention through family members. J Consult Clin Psychol 1999;67:688-97.
45. Garbutt JC, West SL, Carey TS, Lohr KN, Crews FT. Pharmacological treatment of alcohol dependence: a review of the evidence. JAMA 1999;281:1318-25.
46. 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;99:811-28.
47. Krystal JH, Cramer JA, Krol WF, Kirk GF, Rosenheck RA. Naltrexone in the treatment of alcohol dependence. N Engl J Med 2001;345:1734-9.
48. Kiefer F, Jahn H, Tarnaske T, et al. Comparing and combining naltrexone and acamprosate in relapse prevention of alcoholism: a double-blind, placebo-controlled study. Arch Gen Psychiatry 2003;60:92-9.
49. Heinala P, Alho H, Kiianmaa K, Lonnqvist J, Kuoppasalmi K, Sinclair JD. Targeted use of naltrexone without prior detoxification in the treatment of alcohol dependence: a factorial double-blind, placebo-controlled trial. J Clin Psychopharmacol 2001;21:287-92.
50. Azrin NH, Sisson RW, Meyers R, Godley M. Alcoholism treatment by disulfiram and community reinforcement therapy. J Behav Ther Exp Psychiatry 1982;13:105-12.
51. Chick J, Gough K, Falkowski W, et al. Disulfiram treatment of alcoholism. Br J Psychiatry 1992;161:84-9.
52. O'Malley SS, Rounsaville BJ, Farren C, et al. Initial and maintenance naltrexone treatment for alcohol dependence using primary care vs specialty care: a nested sequence of 3 randomized trials. Arch Intern Med 2003;163:1695-704.

53. Kranzler HR, Burleson JA, Del Boca FK, et al. Buspirone treatment of anxious alcoholics: a placebo-controlled trial. Arch Gen Psychiatry 1994;51:720-31.
54. Mason BJ, Kocsis JH, Ritvo EC, Cutler RB. A double-blind, placebo-controlled trial of desipramine for primary alcohol dependence stratified on the presence or absence of major depression. JAMA 1996;275:761-7.
55. Cornelius JR, Salloum IM, Ehler JG, et al. Fluoxetine in depressed alcoholics: a double-blind, placebo-controlled trial. Arch Gen Psychiatry 1997;54:700-5.
56. Angelone SM, Bellini L, Di Bella D, Catalano M. Effects of fluvoxamine and citalopram in maintaining abstinence in a sample of Italian detoxified alcoholics. Alcohol Alcohol 1998;33:151-6.
57. Wutzke SE, Shiell A, Gomel MK, Conigrave KM. Cost effectiveness of brief interventions for reducing alcohol consumption. Soc Sci Med 2001;52:863-70.
58. Kranzler HR, Armeli S, Tennen H, et al. Targeted naltrexone for early problem drinkers. J Clin Psychopharmacol 2003;23:294-304.
59. Johnson BA, Roache JD, Javors MA, et al. Ondansetron for reduction of drinking among biologically predisposed alcoholic patients: a randomized controlled trial. JAMA 2000;284:963-71.
60. Johnson BA, Ait-Daoud N, Bowden CL, et al. Oral topiramate for treatment of alcohol dependence: a randomised controlled trial. Lancet 2003;361:1677-85.
61. Kranzler HR, Wesson DR, Billot L. Naltrexone depot for treatment of alcohol dependence: a multicenter, randomized, placebo-controlled clinical trial. Alcohol Clin Exp Res 2004;28:1051-9.

Primary medical care and reductions in addiction severity: a prospective cohort study

Richard Saitz[1,3,4], Nicholas J. Horton[7], Mary Jo Larson[6], Michael Winter[4,5] & Jeffrey H. Samet[1,2]

Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston University School of Medicine,[1] Departments of Social and Behavioral Sciences[2] and Epidemiology,[3] Center to Prevent Alcohol Problems Among Young People,[4] Data Coordinating Center,[5] Boston University School of Public Health, Boston, MA, New England Research Institutes, Inc., Watertown, MA[6] and the Department of Mathematics, Smith College, Northampton, MA, USA[7]

Correspondence to:
Richard Saitz
Boston Medical Center
91 East Concord Street
Suite 200
Boston
MA 02118
USA
Tel: 617 414 7744
Fax: 617 414 4676
E-mail: rsaitz@bu.edu

Submitted 25 March 2004;
initial review completed 25 June 2004;
final version accepted 16 August 2004

ABSTRACT

Aims To assess whether receipt of primary medical care can lead to improved outcomes for adults with addictions.

Design We studied a prospective cohort of adults enrolled in a randomized trial to improve linkage with primary medical care.

Methods Subjects at a residential detoxification unit with alcohol, heroin or cocaine as a substance of choice, and no primary medical care were enrolled. Receipt of primary medical care was assessed over 2 years. Outcomes included (1) alcohol severity, (2) drug severity and (3) any substance use.

Findings For the 391 subjects, receipt of primary care (≥2 visits) was associated with a lower odds of drug use or alcohol intoxication (adjusted odds ratio (AOR) 0.45, 95% confidence interval (CI) 0.29–0.69, 2 d.f. χ^2 $P = 0.002$). For 248 subjects with alcohol as a substance of choice, alcohol severity was lower in those who received primary care [predicted mean Addiction Severity Index (ASI) alcohol scores for those reporting ≥2, 1 and 0 visits, respectively, 0.30, 0.26 and 0.34, $P = 0.04$]. For 300 subjects with heroin or cocaine as a substance of choice, drug severity was lower in those who received primary care (predicted mean ASI drug scores for those reporting ≥2, 1 and 0 visits, respectively, 0.13, 0.15 and 0.16, $P = 0.01$).

Conclusions Receipt of primary medical care is associated with improved addiction severity. These results support efforts to link patients with addictions to primary medical care services.

KEYWORDS Health services, primary care, severity of illness, substance abuse.

INTRODUCTION

Alcohol and drug abuse cost the United States $328 billion a year, more than heart disease or cancer [1–3]. Efficacious treatments for adults with addictions exist, yet many do not seek treatment [4,5]. Those seeking treatment may find barriers such as limited access and uncoordinated systems of care [6]. Thus, most patients who undergo detoxification do not link with addiction treatment to prevent relapse of this chronic illness [7].

There is great potential benefit to patients and providers for linking the addiction specialty treatment system with primary medical care [6]. Proven effective interventions for alcohol and drug problems can be delivered in primary care settings, such as screening, brief intervention and referral, relapse prevention and coordination of multiple specialty services (e.g. mental health, substance abuse, medical and social services) [6,8,9]; yet these systems of care typically remain unlinked and potential benefits do not accrue.

 doi:10.1111/j.1360-0443.2004.00916.x

Nevertheless, several studies have demonstrated that integrating primary medical care and addiction treatment realizes actual benefits, in particular for addiction outcomes [10,11]. While these studies of on-site care suggest benefit they are not conclusive, and furthermore they address a system of care that is currently not widely available, nor likely to be in the future given that it would require major changes in our health-care delivery systems.

Linkage of patients with addictions to primary care as it currently exists (in a distributive system of linkage/integration [12]) holds the promise of impact on addiction outcomes. Less impact might be expected from a distributive system than of a primary care intervention at an addiction treatment program, although distributive linkage, by using existing systems of care, would also probably be less costly. With tailored interventions, patients with addictions who do not have regular medical care can be linked successfully with primary care from the addiction treatment system [13]. Despite the promise of benefits, the clinical impact of this linkage model with primary care in the community remains largely unknown. As a clinical trial randomizing adults to receive or not receive primary care is unlikely to be conducted, we assessed the impact of receipt of primary medical care on addiction severity and substance use in a prospective cohort of subjects with no primary medical care.

METHODS

Subjects and design

Subjects were participants in a study of a multi-disciplinary assessment and brief motivational intervention to link adults with addictions in a residential detoxification unit, who had no primary care physician, with primary medical care. The details of the randomized trial have been published [13]. Briefly, after the acute symptoms of withdrawal had resolved, eligible subjects were enrolled and provided written informed consent. Eligible subjects were adults 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 established primary care relationships they planned to continue, significant dementia, specific plans to leave the Boston area that would prevent research participation, failure to provide contact information for tracking purposes, or pregnancy were excluded. The clinical trial intervention was associated significantly with increased primary care linkage [13]. However, a substantial proportion of intervention subjects did not link with primary care (31%); and a substantial proportion of control subjects did link (53%) during the first 12 months of follow-up. Thus, the clinical trial intention-to-treat analysis was not informative regarding the impact of primary medical care on clinical outcomes. This report is not the 'treatment received' analysis of that randomized trial; rather, this prospective study focuses on the impact of exposure to what was the outcome (i.e. receipt of primary medical care) of the clinical trial and takes advantage of the prospective data collection.

Subjects were interviewed at baseline during their detoxification stay and completed up to 4 bi-annual follow-up interviews over 2 years. The current study was restricted to a prospectively enrolled and followed cohort of these trial participants who completed at least one of four scheduled research follow-up encounters, 85% (400/468) of those enrolled in the trial and alive at first follow-up opportunity. There were no statistically significant differences in subject characteristics listed below between those who entered the trial and completed versus those who did not complete any research follow-up observations except for race. Of those with follow-up, 54% were white but of those lost, 34% were white. 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.

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] [14], substance problems [Inventory of Drug Use Consequences (InDUC-2R)] [15], readiness to change substance use [using the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES 8AOD)] [16], health-related quality of life [Short Form Health Survey (SF-36)] [17], self-report of attendance at mutual help groups such as Alcoholics Anonymous, and questions regarding primary medical care [13]. Except for demographics, all the assessments were repeated at follow-up interviews. At follow-up interviews, alcohol breath tests were performed to encourage truth telling [18].

Receipt of primary medical care was assessed during follow-up interviews using the following questions: 'Is there one particular doctor that you consider to be your regular personal doctor?'; 'Have you seen any doctors in the last 6 months (or since your last interview)?' If they did not report having a regular personal doctor but had seen a physician, they were asked: 'Would you call or go to one of these/this doctor(s) if you had a medical problem

that was not an emergency?'; 'Do you think one of these doctors could be your regular doctor?'. Subjects reporting either having or possibly having a regular personal doctor or that they would contact the doctor for non-emergent problems were asked 'What type of doctor is your regular personal/this doctor?'

While we could not assess directly the validity of self-report, we did compare self-report with administrative data sources. Computerized databases of patients seen for primary medical care at Boston Medical Center (BMC) or by Boston Health Care for the Homeless Program were queried for visits by study subjects during a 12-month period following study enrollment. This database included all visits to the two BMC-based primary care practices (>120 physicians and >70 000 visits per year), and all visits to primary health-care delivery sites for the homeless at BMC or in a city-wide network for the homeless. While subjects in the randomized trial intervention group were usually referred to care at BMC, all subjects in the cohort could pursue primary medical care anywhere. These administrative data were obtained for 95% of study subjects. Among subjects with any self-report data that were determined by administrative data to have linked, 81% (103/127) reported linkage (kappa = 0.41).

We also identified 100% of study subjects by means of substance abuse treatment utilization data obtained from the Treatment Management Information System, which is maintained by the Massachusetts Department of Public Health (DPH) Bureau of Substance Abuse Services (BSAS). These data were provided under an agreement between Boston Medical Center and the BSAS. This data set includes all episodes of substance abuse treatment utilization that occur at programs receiving state funding for addiction treatment, regardless of whether a particular treatment episode was paid for by the state. Treatment utilization was assessed for 6 months prior to study enrollment and 24 months afterwards.

Predictor variables

The main predictor variable was receipt of primary medical care. Receipt of primary care since the last research contact was defined as a visit to a primary care physician, nurse practitioner or physician assistant reported at a follow-up interview. For the visit to be defined as primary care, the subject had to report having a 'regular personal doctor', that they would call that doctor for a non-emergent issue or that they saw a doctor who 'could be their regular personal doctor'. In addition, that clinician had to be in a specialty that could be considered primary care, including obstetrics and gynecology, family medicine, pediatrics, adolescent medicine, internal medicine, AIDS doctor, asthma doctor, pulmonary doctor, cardiologist or a gastroenterologist. When the specialty was unknown to the subject or was a specialty other than those specifically queried, the physician's office was contacted to determine the specialty. Examples of specialties that were classified as non-primary care clinicians were podiatrists, emergency medicine physicians and psychiatrists. Because of its right-skewed distribution, we categorized primary care receipt for analyses as 0, 1 or 2 or more visits to avoid undue influence of outliers on the analysis. For a secondary sensitivity analysis to explore further dose–response, primary care receipt was categorized as a variable with eight levels (0, 1, 2, 3, 4, 5, 6 or 7 or more visits).

Additional predictor variables of interest included the following: demographics, homelessness (defined as one or more nights in a shelter or on the street in the preceding 6 months), addiction severity (the alcohol and drug ASI scores at study entry), attendance at mutual help meetings, the physical and mental component summary scores derived from the SF-36 [19] and the Taking Steps scale score from the SOCRATES. Substance abuse treatment utilization was a dichotomous variable of interest. This treatment variable represented any treatment, not known effective doses. The following BSAS services were considered treatment: transitional support services, recovery homes, therapeutic communities, supportive housing, residential treatment, family substance abuse shelters, day treatment, out-patient substance abuse counseling, methadone treatment, community-based case management, acupuncture, intensive out-patient treatment and postdetox recovery programs [20].

Outcome variables

Addiction severity was the primary outcome of interest. The alcohol and drug ASI composite scores (ranging from 0 to 1), and any drug use or alcohol use to intoxication (or more than three drinks in a day) (a dichotomous variable) were the main outcome variables. This latter variable was based on questions on substance use in the ASI, regarding the most recent 30 days.

A secondary analysis was conducted using substance-related problems as the dependent variable, as measured using the InDUC-2R covering a 6-month time frame (score range 0–135 [21], median of 73 observed previously for men entering drug abuse treatment [15]). Substance abuse treatment utilization (described above) was an additional outcome variable of interest, as a possible measure of impact of primary care linkage.

Subject characteristics

Of the 400 subjects eligible for this prospective cohort study, nine did not have complete data and were excluded. All subjects were interviewed at study entry; an additional 975 interviews occurred during the 2 years

after the initial interview. Of the nine subjects excluded, one was missing addiction outcome measures, five were missing addiction severity measures at study entry and three were missing homelessness, insurance and health-related quality of life data (as they completed abbreviated follow-up interviews that did not reassess these items). None were missing primary care receipt information.

Subject characteristics were similar regardless of drug of choice. Of the 391 subjects, 76% were male, mean age was 36, 50% were African American, 9% Hispanic and 60% had no health insurance; 57% of subjects with alcohol as a drug of choice were homeless, 43% of subjects with another drug as a substance of choice were homeless. The mean SF-36 Physical Component Summary (PCS) and Mental Component Summary (MCS) scores were 48 and 31, respectively (the mean MCS and PCS score for the US population is 50; 89% of adults with MCS scores of 30–34 screen positive for depression [19]). The mean ASI alcohol and drug scores were 0.47 and 0.26, respectively (these ASI scores are similar to those of individuals entering the public treatment system in Boston and more severe than those in a clinical addiction treatment sample in an HMO in California [22,23]). Subjects reporting alcohol as a drug of choice had a higher mean ASI alcohol score (0.66, SD 0.25); subjects reporting another drug as a substance of choice had a higher mean ASI drug score (0.31, SD 0.10). The mean Taking Steps scale score reflecting greater readiness to address addiction problems with higher scores was 36 (the possible range of this readiness assessment measure is 8–40 with a median of 33 for patients in alcoholism treatment [16]). Subject characteristics were similar across drug of choice groups with the exception of drug and alcohol ASI composite scores, as would be expected.

Analysis

The analysis for this paper was generated using SAS/STAT software, version 8.2 [24]. All study subjects were eligible for analyses of treatment utilization and substance use. Analyses of alcohol and drug addiction severity were restricted to subjects who reported alcohol or another drug as their substance of choice, respectively. Descriptive statistics were used to characterize the study sample (proportions, means and standard deviations). Reported *P*-values are two-tailed, and a *P*-value of less than 0.05 was considered statistically significant. Estimates of time to first report of linkage with primary medical care or substance abuse treatment, where linkage could occur at 6, 12, 18 or 24 months, were calculated using the Kaplan–Meier method. Subjects were censored after linkage or their last follow-up time.

We fitted longitudinal regression models adjusting for time and intrasubject correlations for all multi-variable analyses [25]. For dichotomous outcome variables we used generalized estimating equations (GEE) implemented in PROC GENMOD. For the continuous outcome variables we used generalized linear models for correlated data (GLMCD) implemented using PROC MIXED. Each subject contributed up to four observations of the outcome variables (from 6, 12, 18 and 24 months after study enrollment). An independent working correlation model and an empirical variance estimator were used for the GEE models, while an unstructured working covariance model was assumed for the GLMCD models. All analyses included covariates of clinical importance. Homelessness, health insurance, physical and mental health-related quality of life, readiness to change and addiction treatment, when included, were time-varying covariates. Other covariates were those assessed at study entry.

We fitted three multi-variable models to address the primary research question, one for each primary outcome. We also fitted one model for the outcome of substance abuse treatment. All analyses were adjusted for time, age, sex, race, randomized assignment in the clinical trial, homelessness, health insurance, mental and physical health-related quality of life, alcohol and drug severity at study entry and readiness to change. Analyses with addiction severity and substance use as the outcome variables were also adjusted for substance abuse treatment and use of mutual help groups. Analyses with substance abuse treatment and substance use as outcomes were adjusted for substance of choice.

Secondary confirmatory/sensitivity analyses addressed whether the results of the three main models and the substance abuse treatment model would differ when entering primary care as an eight-level variable. An additional analysis confirmed the results by testing the effect of primary care receipt on substance-related problems, adjusting for the same covariates as in the models with substance use as the outcome. To address the possibility of higher-order effects, we added the following seven interactions with receipt of primary medical care to the three main models: alcohol and drug severity at study entry, mutual help use, readiness to change, physical health-related quality of life, randomization assignment and substance abuse treatment utilization.

RESULTS

Of 391 subjects, 194 (49.6%) received a total of two or more primary care visits, 53 (13.5%) received a total of one visit and 144 (36.8%) received no primary medical care visits during the study period. Kaplan–Meier estimated proportions of subjects receiving a total of one or more primary care visits at 6, 12, 18 and 24 months after study entry were 31%, 44%, 58% and 70%, respectively.

The four corresponding proportions for receipt of a total of two or more primary care visits were 19%, 30%, 44% and 58%.

Receipt of primary medical care and addiction severity and substance abuse treatment

For subjects who reported alcohol as their first or second drug of choice, receipt of primary care was significantly associated with improved alcohol (ASI) severity ($F_{2,239} = 3.29$, $P = 0.04$) (predicted mean alcohol ASI for 0, 1, ≥2 visits, respectively, 0.34, 0.26, 0.30) in a multivariable analysis. Similarly, for subjects reporting heroin or cocaine as their first or second drug of choice, receipt of primary care was associated significantly with improved drug (ASI) severity ($F_{2,291} = 4.49$, $P = 0.01$) (predicted mean drug ASI for 0, 1, ≥2 visits, respectively, 0.16, 0.15, 0.13). For all 391 subjects, regardless of substance of choice, receipt of primary care was associated significantly with a decreased odds of drug use or use of alcohol to intoxication (during 30 days) ($\chi^2 = 12.90$, 2 d.f., $P = 0.0016$). During the 2 years of follow-up, 267 subjects (67.3%) received substance abuse treatment services (Kaplan–Meier estimates for each of the four successive 6-month periods 33.5%, 44.3%, 59.9%, 73.5%). In a multi-variable analysis, receipt of primary care was not significantly associated with receipt of substance abuse treatment (Table 1).

Confirmatory/sensitivity analyses

In multi-variable models entering the same covariates as the previously described models, an eight-level variable for receipt of primary care was associated with lower alcohol severity at a borderline level of significance (decrease in alcohol severity score for additional visit −0.009, 95% CI −0.020–0.001, $F_{1,239} = 2.88$, $P = 0.09$), was significantly associated with improved drug severity (decrease in drug severity for additional visit, −0.006, 95% CI −0.010 to −0.001, $F_{1,290} = 6.70$, $P = 0.01$), and was significantly associated with a reduced odds of drug use or drinking alcohol to intoxication (OR for one more visit 0.83, 95% CI 0.75–0.91, $\chi^2 = 15.76$, 1 d.f., $P < 0.0001$). This eight-level ordinal variable was not sig-

Table 1 Association between primary care visits and addiction outcomes in multi-variable analyses.

	Substance abuse treatment n = 391 Odds ratio, 95% CI	*Alcohol severity n = 248*	*Drug severity n = 300*	*30-day drug use or use of alcohol to intoxication n = 391 Odds ratio, 95% CI*
Primary care visits during 6 months		Predicted mean ASI score	Predicted mean ASI score	
0	–	0.34	0.16	–
1	1.08 (0.70–1.67)	0.26	0.15	0.91 (0.54–1.52)
≥ 2	1.04 (0.73–1.49)	0.30	0.13	0.45 (0.29–0.69)
P-value	$P = 0.94$	$P = 0.04$	$P = 0.01$	$P = 0.002$
		Parameter estimate (95% CI)	Parameter estimate (95% CI)	
Age (decade)	0.70 (0.58–0.86)	−0.0205 (−0.0502, 0.0132)	−0.0014 (−0.0161, 0.0134)	0.86 (0.65–1.15)
Homeless	1.52 (1.09–2.12)	0.0448 (0.0023, 0.0872)	0.0173 (−0.0002, 0.0349)	2.52 (1.72–3.69)
Health insurance	1.73 (1.21–2.47)	0.0392 (−0.0068, 0.0851)	0.0152 (−0.0022, 0.0325)	1.86 (1.28–2.72)
Mental health (MCS) (10 points)	0.88 (0.79–0.99)	−0.0636 (−0.0792, −0.0479)	−0.0319 (−0.0381, −0.0258)	0.67 (0.58–0.77)
Physical health (PCS) (10 points)	1.20 (1.02–1.40)	−0.0392 (−0.0590, −0.0194)	−0.0192 (−0.0273, −0.0112)	0.85 (0.70–1.02)
Baseline addiction severity, alcohol (0.10 points)	0.95 (0.90–1.01)	0.0176 (0.0076, 0.0276)	−0.0046 (−0.0077, −0.0015)	1.02 (0.94–1.09)
Baseline addiction severity, drug (0.10 points)	1.06 (0.93–1.20)	−0.0117 (−0.0283, 0.0049)	0.0234 (0.0136, 0.0331)	1.14 (0.98–1.34)
Taking Steps score (1 point)	1.05 (1.03–1.08)	−0.0088 (−0.0121, −0.0054)	−0.0026 (−0.0040, −0.0012)	0.86 (0.83–0.89)

For primary care visits, reference group for odds ratios is no primary care visits. All analyses are adjusted for the variables listed in the table and sex, race/ethnicity and randomized group. Time (1st, 2nd, 3rd or 4th interview) was a significant predictor of outcome only for the substance abuse treatment analysis ($\chi^2 = 29.53$, d.f. = 3, $P < 0.0001$). ASI and substance use outcome analyses also adjusted for substance abuse treatment and mutual help use (not significant in any analysis). Substance use and substance abuse treatment outcome analyses adjusted for substance of choice (not statistically significant).

nificantly associated with receipt of substance abuse treatment (OR for an additional visit 1.00, 95% CI, 0.93–1.08, $\chi^2 = < 0.01$, 1 d.f., $P = 0.97$).

For all 391 subjects, regardless of substance of choice, receipt of primary care (categorized as 0, 1 or ≥2 visits) was significantly associated with decreased substance-related problems (measured by the InDUC-2R, a measure of problems in past 6 months or since last research assessment if >6 months) ($F_{2,381} = 13.59$, $P < 0.0001$) (predicted mean score for 0, 1, ≥2 visits, respectively, 52, 51, 39).

To address the possibility of higher order effects, we tested seven interactions between receipt of primary care and alcohol and drug severity at study entry, mutual help use, the Taking Steps score, physical health-related quality of life, randomization group and substance abuse treatment utilization in models with alcohol and drug severity and drug use or alcohol intoxication outcomes (three outcomes, three models). Of these 21 interactions tested, only one was significant at $P < 0.05$, the interaction with physical health-related quality of life in the model predicting drug use severity (Table 2) ($F_{2,291} = 3.53$, $p = 0.03$). The association between primary care receipt and drug use severity varied by physical health-related quality of life. Predicted mean differences in drug severity were similar regardless of primary care receipt for those with better health (higher PCS). Primary care appeared to have a greater impact on drug severity in people with worse physical health status (lower PCS scores), than it did on people with better health status. At the observed mean PCS of 48, and at 2 standard errors above this mean (70), there was little difference in the decreases in ASI drug score attributable to receipt of primary care.

CONCLUSIONS

Receipt of primary medical care in a distributive model by adults with addictions who have not recently had such care is associated with reduced problems and severity of addictions over a 24-month period. This association does not appear to be mediated by exposure to substance abuse treatment, nor does it appear to be affected by addiction severity, mutual help use or readiness to change. As might be expected, primary care had a greater impact on addiction severity in patients with worse physical health. However, given the numerous interactions tested and the appearance of the finding for subjects with only cocaine or heroin as a substance of choice, these latter findings should be viewed as hypothesis generating. Analyses suggest some evidence for dose–response, both those that tested levels of primary care exposure and the ordered results from all other models (i.e. more effect from ≥2 visits versus 1, and 1 versus 0 visits). In the analysis of alcohol ASI results were not ordered according to primary care dose, but the one-visit group was substantially smaller than the 0- or ≥2-visit groups.

Our findings are consistent with prior studies. In a landmark randomized trial, substance abuse treatment patients receiving on-site medical, psychiatric, employment and family services had less opiate use and improved medical, employment, legal and psychiatric outcomes [26]. In another randomized trial patients with substance abuse-related medical conditions assigned to on-site primary care at an addiction treatment program were more likely to remain abstinent than patients in usual separate primary care [10]. In a retrospective cohort study patients with addictions treated at programs with on-site primary care had improved addiction outcomes when compared with patients treated in programs without such services [11]. Willenbring *et al*. tested another variation of on-site medical and addictions care integration [27]. In a randomized trial in a special alcohol clinic for veterans, the integrated care group was more likely to be abstinent than a usual care group [28]. Laine *et al*. found that patients with addictions who receive both regular addiction and medical (not necessarily primary) care are less likely to be hospitalized than those who received one or neither service [29]. Thus, medical care in general, and primary care specifically, appears to improve addiction outcomes.

Our study adds to this literature by suggesting that these results are also true for primary medical care as it is delivered in the community to a group of patients with addictions that include many without homes or

Table 2 Predicted mean differences in drug addiction severity by receipt of primary care for three values of physical health-related quality of life (sample mean 48, and 2 standard errors above and below).

	Predicted mean differences in drug addiction severity		
	Sample mean - 2 SE	*Sample mean*	*Sample mean + 2 SE*
Receipt of primary care	PCS = 26	PCS = 48	PCS = 70
No primary care visits	–0.08	–0.15	–0.21
1 visit	–0.19	–0.18	–0.19
≥ 2 visits	–0.16	–0.19	–0.23

health insurance. In fact, it is people who are socially disadvantaged who may be most likely to benefit from primary medical care. Laine *et al.*'s population, who received benefit from regular care, were insured by Medicaid (state health insurance for the poor or disabled) [29]. Gelberg *et al.* have noted previously that in a vulnerable population, more severe addictions and more severe homelessness did not deter access to needed care, and that having a regular source of care predicted better health outcomes [30]. The observation that substance abuse treatment did not appear to mediate the relationship between primary care and improved addiction severity is not surprising. Linkage between primary care and the addiction treatment system has been elusive [6]. Most patients with addictions do not seek specialty treatment [5] and substance abuse treatment has been notably not associated with linkage to primary medical care [31]. This absence of apparent contribution of substance abuse treatment to the observed benefit of primary care to addictions outcomes could be viewed as both a cup half empty and half full. Primary care may be contributing independently to improved addictions outcome. In addition, the opportunity to achieve even greater benefit may exist if mechanisms were instituted to link primary care more effectively with substance abuse treatment.

The principal limitation that should be considered in interpreting our results is that our data are from an observational study. As a result, the observed associations between receipt of primary care and addiction severity could be due to confounding. However, we adjusted for likely confounders of this relation. Furthermore, the fact that data for this observational study were collected prospectively in a study focused on primary care and addictions is a particular strength. One might also question whether primary care led to reduced addiction severity, or whether adults with improved addictions became health conscious and linked with primary care. Another similar possibility is that those with the most severe problems who 'hit rock bottom' began to take care of themselves, including a visit to primary care and abstinence or less use. Both explanations seem less plausible because analyses adjusted for substance abuse consequences and health status. Furthermore, measures of addiction severity referred to the recent past while measures of primary care receipt referred to the past 6 months or more, making the former more likely. The temporal association, the confirmation of a prior hypothesis, adjustment for important possible confounders and the consistency across outcomes suggest that receipt of primary care was associated with later improvements in addiction severity. Finally, some of the observed effects were relatively small (e.g. differences of 0.02 in drug ASI). Our goal was not to determine clinical significance but rather to identify an association. It is notable that an association between primary medical care and addiction severity was detectable at all, given the variability in subject characteristics and the many other contributors to addiction severity. In fact, some effects were quite substantial and clinically significant (e.g. >50% decrease in odds of substance use, large changes in InDUC problem score and alcohol ASI).

The important question, of whether or not primary medical care as it is currently delivered in the community can improve addiction severity, is best answered in a naturalistic study, and this highlights several important strengths of this study. First, the data collection in this cohort study was prospective. Secondly, two of the main purposes of the research assessments a priori were to assess primary medical care utilization and addiction severity in detail. Thirdly, multiple measures of addiction severity used in the study have been validated and results across these measures were consistent. Finally, receipt of primary care was corroborated by administrative data.

One could still argue that our findings are due to inadequately addressed confounding or that the outcome (i.e. addiction severity) leads to seeking primary medical care. Such a critical assessment might contend that the hypothesis, 'primary medical care leads to improved addiction severity', can only be answered definitively by a randomized trial of primary care. Addressing these concerns optimally is difficult, as such a study will probably not be conducted, for pragmatic and ethical reasons.

This study and previous studies support the contention that receipt of primary care is associated with improved outcomes for adults with addictions. Our study results and those of others cited herein may be the best type of evidence to bring to address the question. This evidence suggests strongly that patients with addictions should receive primary medical care in addition to addiction treatment. Efforts to link and/or integrate primary medical care with addiction specialty care are worthwhile. Patients with addictions can be linked with primary medical care [13]. In addition, actually providing services known to be efficacious and feasible in primary care settings that are not currently in widespread use could lead to even greater improvements for patients with addictions. How to make primary medical care truly accessible and actually used by adults with alcohol and other drug dependence remains a challenge for researchers and clinicians alike.

Acknowledgements

Preliminary results of this study were presented at the annual national meeting of the Association for Medical

Education and Research in Substance Abuse (AMERSA), November 2003, in Baltimore, MD. Data management was performed by the Data Coordinating Center, Boston University School of Public Health. The National Institute on Alcohol Abuse and Alcoholism (RO1-AA10870), and the National Institute on Drug Abuse (RO1-DA10019) provided support for the study. The Bureau of Substance Abuse Services, Department of Public Health, Commonwealth of Massachusetts, provided substance abuse treatment utilization data, and the Boston Health Care for the Homeless Program provided data on primary care utilization, as did Boston Medical Center. This work was supported in part by General Clinical Research Center grant MO1-RR00533 from the National Center for Research Resources.

References

1. Harwood, H. J., Fountain, D. & Livermore, G., eds (1998) *The Economic Costs of Alcohol and Drug Abuse in the United States—1992*. Washington, DC: National Institute on Alcohol Abuse and Alcoholism.
2. Anonymous (2003) *NHLBI Morbidity and Mortality Chartbook, 2000.* Bethesda, MD: National Heart, Lung, and Blood Institute.
3. Harwood, H. (2000) *Updating Estimates of the Economic Costs of Alcohol Abuse in the United States: Estimates, Update Methods and Data.* Bethesda, MD: The Lewin Group for the National Institute on Alcohol Abuse and Alcoholism.
4. Miller, W. R. & Wilbourne, P. L. (2002) Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. *Addiction*, **97**, 265–277.
5. Green-Hennessy, S. (2002) Factors associated with receipt of behavioral health services among persons with substance dependence. *Psychiatric Services*, **53**, 1592–1598.
6. Samet, J. H., Friedmann, P. & Saitz, R. (2001) Benefits of linking primary medical care and substance abuse services: patient, provider, and societal perspectives. *Archives of International Medicine*, **161**, 85–91.
7. Mark, T., Dilonardo, J., Chalk, M. & Coffey, R. (2002*) Substance Abuse Detoxification: Improvements Needed in Linkage to Treatment (SMA-02–3728).* Rockville, MD: Substance Abuse and Mental Health Services Administration.
8. Friedmann, P. D., Saitz, R. & Samet, J. H. (1998) Management of adults recovering from alcohol or other drug problems: relapse prevention in primary care. *JAMA*, **279**, 1227–1231.
9. Moyer, A., Finney, J. W., Swearingen, C. E. & 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.
10. Weisner, C., Mertens, J., Parthasarathy, S., Moore, C. & Lu, Y. (2001) Integrating primary medical care with addiction treatment: a randomized controlled trial. *JAMA*, **286**, 1715–1723.
11. Friedmann, P., Zhang, Z., Hendrickson, J., Stein, M. & Gerstein, D. (2003) Effects of primary medical care on addiction and medical severity in substance abuse treatment programs. *Journal of General Internal Medicine*, **18**, 1–8.
12. Samet, J. H., Saitz, R. & Larson, M. J. (1996) A case for enhanced linkage of substance abusers to primary medical care. *Substance Abuse*, **17**, 181–192.
13. Samet, J. H., Larson, M., 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 multidisciplinary health intervention in a detoxification unit. *Addiction*, **98**, 509–516.
14. 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. *Journal of Substance Abuse Treatment*, 9, 199–213.
15. Tonigan, J. & Miller, W. (2002) The inventory of drug use consequences (InDUC): test–retest stability and sensitivity to detect change. *Journal of Psychological Addictive Behaviors*, **16**, 165–168.
16. Miller, W. & Tonigan, J. (1996) Assessing drinkers' motivation for change: the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). *Journal of Psychological Addictive Behaviors*, **10**, 81–89.
17. Ware, J. E. (1993) *SF-36 Health Survey Manual and Interpretation Guide.* Boston, MA: The Health Institute.
18. Welte, J. W., Perry, P., Longabaugh, R. & Clifford, P. R. (1998) An outcome evaluation of a hospital-based early intervention program. *Addiction*, **93**, 573–581.
19. Ware, J. E. (1994) *SF-36 Physical and Mental Health Summary Scales: a User's Manual*. Boston, MA: The Health Institute.
20. Cellucci, A. P., Swift, J., O'Leary, W. D., Koh, H. K. & Walker, D. K. (2001) *Substance Abuse Prevention and Treatment Service Directory*. Boston, MA: Massachusetts Department of Public Health Bureau of Substance Abuse Services.
21. Miller, W. & Tonigan, J. (1995) *The Drinker Inventory of Consequences (DrInC). An Instrument for Assessing Adverse Consequences of Alcohol Abuse.* Test manual. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.
22. Leonhard, C., Mulvey, K., Gastfriend, D. R. & Shwartz, M. (2000) The Addiction Severity Index: a field study of internal consistency and validity. *Journal of Substance Abuse Treatment*, **18**, 129–135.
23. Weisner, C., McLellan, A. T. & Hunkeler, E. M. (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**, 103–109.
24. SAS Institute (1999–2001) *SAS/STAT software*, version 8.2, *SAS System for Windows*. Cary, NC: SAS Institute, Inc.
25. Cnaan, A., Laird, N. M. & Slasor, P. (1997) Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, **16**, 2349–2380.
26. McLellan, A. T., Arndt, I. O., Metzger, D. S., Woody, G. E. & O'Brien, C. P. (1993) The effects of psychosocial services in substance abuse treatment. *JAMA*, **269**, 1953–1959.
27. Willenbring, M. L., Olson, D. H. & Bielinski, J. (1995) Integrated out-patients treatment for medically ill alcoholic men: results from a quasi-experimental study. *Journal of Studies on Alcohol*, **56**, 337–343.
28. Willenbring, M. L. & Olson, D. H. (1999) A randomized trial of integrated out-patient treatment for medically ill alcoholic men. *Archives of Internal Medicine*, **159**, 1946–1952.
29. Laine, C., Hauck, W. W., Gourevitch, M. N., Rothman, J., Cohen, A. & Turner, B. J. (2001) Regular out-patient

medical and drug abuse care and subsequent hospitalization of persons who use illicit drugs. *JAMA*, **285**, 2355–2362.

30. Gelberg, L., Andersen, R. M. & Leake, B. D. (2000) The behavioral model for vulnerable populations: application to medical care use and outcomes for homeless people. *Health Services Research*, **34**, 1273–1302.
31. Saitz, R., Mulvey, K. P. & Samet, J. H. (1997) The substance-abusing patient and primary care: linkage via the addiction treatment system? *Substance Abuse*, **18**, 187–195.

Association for Medical Education and Research in Substance Abuse

Jeffrey H. Samet[1,2], Marc Galanter[3], Carly Bridden[1] & David C. Lewis[4]

Clinical Addiction Research and Education Unit (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center,[1] Department of Social and Behavioral Sciences, Boston University School of Public Health,[2] Department of Psychiatry, New York University School of Medicine,[3] and Center for Alcohol and Addiction Studies, Brown University, USA[4]

Correspondence to:
Jeffrey H. Samet
Professor of Medicine and Public Health
Boston University Schools of Medicine and Public Health
Chief, Section of General Internal Medicine
91 E. Concord Street
Suite 200
Boston
MA 02118
USA
Tel: (617) 414 7288
E-mail: jsamet@bu.edu

Submitted 9 May 2005;
initial review completed 24 May 2005;
final version accepted 24 May 2005

ABSTRACT

The Association for Medical Education and Research in Substance Abuse (AMERSA) is a multi-disciplinary organization committed to health professional faculty development in substance abuse. In 1976, members of the Career Teachers Training Program in Alcohol and Drug Abuse, a US federally funded multi-disciplinary faculty development program, formed AMERSA. The organization grew from 59 founding members, who were primarily medical school faculty, to over 300 health professionals from a spectrum of disciplines including physicians, nurses, social workers, dentists, allied health professionals, psychologists and other clinical educators who are responsible for advancing substance abuse education. AMERSA members promote substance abuse education among health professionals by developing curricula, promulgating relevant policy and training health professional faculty to become excellent teachers in this field. AMERSA influences public policy by offering standards for improving substance abuse education. The organization publishes a peer-reviewed, quarterly journal, *Substance Abuse*, which emphasizes research on the education and training of health professions and also includes original clinical and prevention research. Each year, the AMERSA National Conference brings together researchers and health professional educators to learn about scientific advances and exemplary teaching approaches. In the future, AMERSA will continue to pursue this mission of advancing and supporting health professional faculty who educate students and trainees to address substance abuse in patients and clients.

KEYWORDS Education, faculty development, professional organization, substance abuse.

INTRODUCTION

The Association for Medical Education and Research in Substance Abuse (AMERSA) is a multi-disciplinary organization committed to health professional faculty development in substance abuse (Table 1). It is the only multi-disciplinary national organization in the United States with this explicit educational mission. During its 29 years in existence, AMERSA has attracted health professionals including physicians, nurses, social workers, psychologists, public health practitioners, dentists, other allied health professionals and clinical educators from a broad spectrum of disciplines. Curriculum materials used in much of the addictions teaching for health professionals were developed by AMERSA members. Its members have been responsible for advancing an agenda in the

 doi:10.1111/j.1360-0443.2005.01228.x

Table 1 AMERSA's mission statement.

AMERSA, founded in 1976, is a multi-disciplinary organization of health-care professionals dedicated to improving education in the care of individuals with substance abuse problems. AMERSA's mission is to:

- Provide leadership and improve training for all health-care professionals in the management of problems related to alcohol, tobacco and other drugs
- Disseminate state-of-the-art scientific information about substance abuse education and research, through means such as the National AMERSA conference and the organization's journal, *Substance Abuse*
- Provide mentoring for health professionals interested in becoming teachers, clinicians and researchers in the field
- Promote cultural competence and inclusiveness among health-care professionals in their work with individuals affected by alcohol, tobacco and other drug problems
- Promote collaboration among multiple professions including, but not limited to, medicine, nursing, social work, psychology, dentistry, pharmacology and public health
- Build a national network of substance abuse experts who can advise local, national and international organizations on health professional substance abuse education through representation at national forums

United States focused on curriculum development in substance abuse health professional education [1–7].

FOUNDING OF AMERSA

Since its inception, AMERSA's goal has been to improve the substance abuse education of health professional trainees related to prevention, intervention and treatment of individuals and families. Advancing the knowledge and skills of faculty at academic professional schools has been seen as the most effective means. In 1976, members of the Career Teachers Training Program in Alcohol and Drug Abuse [8] formed AMERSA. The Career Teachers Program (1972–82), sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA), was one of the first multi-disciplinary health professional faculty development programs. Over the course of this program's existence, 59 career teachers, faculty in medical and public health schools, were challenged by the problems they encountered in pursuit of their goal: implementing curriculum changes to enhance substance abuse education within their own professional schools. Within the structure of the Careers Teachers Program, they were able to develop common strategies and support each other as they encountered common barriers to achieving this goal. Barriers included resistance from curriculum committees, faculties and deans who did not support inclusion of substance abuse issues in the curriculum. Despite the resistance, these substance abuse educator pioneers recognized the benefits of the support network of this national faculty development program. When it became clear that the effective federally funded Career Teachers Program was coming to an end, the faculty recipients of this support decided to broaden the group beyond the career teachers and form a new national organization. Thus, AMERSA was established with Marc Galanter MD, leading the organization as the first president. The founding members also initiated the AMERSA journal, *Substance Abuse*, which has continued to be published and grow since its inception. Members of the organization and AMERSA staff continue to work closely with NIDA, NIAAA and other federal agencies to carry on and strengthen the mission of the organization. Originally housed at the Brown University Center for Alcohol and Addiction Studies, AMERSA's national headquarters are now independently located in Providence, Rhode Island .

IMPACT ON FACULTY DEVELOPMENT IN SUBSTANCE ABUSE

As stated on the organization's website:

> AMERSA members from diverse departments at health professional schools have developed, implemented, and evaluated state-of-the-art curricula, educational programs, and faculty development programs. [Its] members have developed clinical and research measures for substance abuse services and professional education. They are actively engaged in research related to substance abuse education, clinical service, and prevention [9].

AMERSA has pursued the advancement of substance abuse education among health professionals by developing and promulgating appropriate policy and by supporting health professional faculty to become more knowledgeable and skillful about teaching in this field. AMERSA has been instrumental in setting educational standards for essential knowledge and skills required of primary care physicians and more recently a whole spectrum of health professions. The organization and its members have clearly articulated the rationale for inclusion of substance abuse health professional education [7,10].

In 1985, AMERSA sponsored a conference with the Betty Ford Center, NIAAA and NIDA to develop consensus on the knowledge, skills and desirable educational experiences necessary for primary care physicians in alcohol and drug abuse, the optimal roles and responsibilities of the involved organizations and the best strategies for implementation. The result of this landmark conference was a delineation of the subjects and necessary skills that should be taught, the role of medical schools and government, and the development of specialty-specific guidelines. This meeting was the forerunner of several subsequent US efforts in faculty development such as the US Health Services and Resources Administration (HRSA) Faculty Development Programs in alcohol and other drug abuse, targeting general internal medicine, family medicine and pediatrics [11], and several faculty development programs in the 1990s funded by the Center for Substance Abuse Prevention (CSAP), SAMHSA, targeting nursing, social work, medicine and public health. Most recently, US federal agencies supported AMERSA in the creation of the 2002 Strategic Plan for Interdisciplinary Faculty Development in substance abuse and the newest faculty development program, Project MAINSTREAM (*M*ulti*A*gency *IN*itiative for *S*ubstance abuse *T*Raining and *E*ducation for *A*Merica) [7]. Members have worked closely with the Center for Substance Abuse Treatment (CSAT) and HRSA in Project MAINSTREAM for continuing development of multidisciplinary addiction faculty. As part of Project MAINSTREAM, health professionals from a variety of disciplines attend the annual National AMERSA Conference. Because of its long-standing commitment to substance abuse training, AMERSA has been instrumental in establishing addiction training in medical, nursing and other health professional institutions nationally.

Thus, one of AMERSA's greatest contributions is the clear articulation of critical curriculum content including skills training. Through its consensus statements [12], AMERSA has advanced the concept that primary care clinical teams are in a critical position to detect and treat patients with substance abuse problems, yet they continue to struggle, due in part to lack of training. Therefore, a focus of the organization is to incorporate substance abuse clinical and research activities into mainstream clinical practice. AMERSA pursues its goal of setting educational standards by presenting a showcase of model programs at its annual national meeting and publishing educational research in its journal, *Substance Abuse*. Internationally, AMERSA was not alone in its early efforts to effect change in the education of medical professionals. Advances in training and curriculum design were taking place in many countries; prominent among these were Australia, England, Sweden and Canada [13–15].

IMPACT ON SUBSTANCE ABUSE EDUCATION POLICY

AMERSA affects public policy by offering standards that inform the federal government and others on how to improve substance abuse health professional education. AMERSA led the development of standards for a spectrum of generalist health professionals with multi-agency federal support. Its members developed a strategic plan for the nation, released at the National Press Club in 2002, addressing substance abuse health professional education [7]. The strategic plan includes recommendations to the US Department of Health and Human Services and other federal agencies as well as recommendations to legislators. The Strategic Plan highlights the need for faculty development and the impact that routine substance abuse screening and intervention by generalist health professionals can have in linking patients and family members to services to facilitate treatment and recovery. The Strategic Plan identifies additional methods for building a national infrastructure for faculty development in substance abuse. Most recently, AMERSA was invited by the White House Office of National Drug Control Policy (ONDCP) to participate with national experts in the 2004 Leadership Conference on Medical Education in Substance Abuse. An ONDCP report is expected in 2005 that will outline a strategy that, in part, builds upon the 2002 Strategic Plan for Interdisciplinary Faculty Development.

SUBSTANCE ABUSE: THE OFFICIAL PUBLICATION

AMERSA publishes *Substance Abuse*, a peer-reviewed, quarterly journal that emphasizes research on health professional education in substance abuse and also includes original clinical and prevention research. It is a recognized source of empirical findings for health professionals and addiction specialists in teaching, clinical care and service delivery. It features original research and review articles on a variety of related topics: the education and training of health professionals in substance abuse; clinical care for substance abusers in a variety of settings; the organization of substance abuse treatment services; pre-clinical and clinical research, including therapeutic interventions and behavioral studies; medical complications associated with drug abuse; substance abuse among specific groups or populations; applied science research; and policy issues. The journal publishes timely editorials and book reviews, as well as abstracts from the AMERSA National Conference. *Substance Abuse* is distributed to all AMERSA members. The journal has a multi-disciplinary Editorial Board that represents the

full strength and range of AMERSA's experience and teaching.

ANNUAL CONFERENCE

The annual AMERSA National Conference has been the central exceptional product of the organization, as it brings together researchers and health professional educators to learn about scientific advances and exemplary teaching approaches. The conference fosters collaboration of health professionals within and among diverse disciplines, backgrounds and professional environments in a particularly supportive atmosphere encouraging peer mentoring and career development. It attracts presenters with national and international reputations to share new developments in substance abuse education, treatment, prevention and research. New research presented in both poster and oral formats is subsequently published as abstracts in *Substance Abuse*. This national meeting is held regularly during the fall in the Washington, DC area to take advantage of speakers from NIAAA and NIDA as well as to enhance networking with leaders of the National Institutes of Health (NIH) and the Substance Abuse and Mental Health Services Administration (SAMHSA).

AWARDS SPONSORED BY AMERSA

AMERSA sponsors several awards to support and recognize outstanding individual achievements in the field of substance abuse. The premier awards given to members or non-members of the organization are The John P. McGovern Award for Excellence in Medical Education and The Betty Ford Award. The John P. McGovern Award is given to an individual who has made important contributions to substance abuse education and research. The Betty Ford Award is given to an individual who has played a significant role in the treatment and recovery of drug-dependent individuals, particularly women. Each Ford and McGovern awardee is invited to speak at the national conference. The New Investigator/Educator Award is given to an AMERSA member who has made significant contributions to substance abuse education or research at an early stage in his or her career, and demonstrates the potential for future achievements in the field. The Excellence in Mentorship Award is given to an AMERSA member from any discipline who has provided outstanding mentoring to junior faculty and/or trainees, resulting in those individuals' increased scholastic productivity and career advancement in the area of substance abuse education or research.

SOURCES OF FUNDING

Sources of funding are primarily through membership dues and registration fees from the annual conference. In recent years other funds have been obtained from foundation and federal grants (the Endowment of the John P. McGovern Foundation, CSAT, HRSA, NIAAA and NIDA), most of which are directed at improving health professional substance abuse training. Funding for the annual conferences has included support from NIDA and NIAAA to ensure high quality presentations for plenary sessions and recruitment of attendees who are promising diverse health professional faculty.

MEMBERSHIP

As described previously, AMERSA was comprised originally of medical school faculty. During the early years of the organization's development, however, members realized the need to involve a broader spectrum of health professionals in order to have a more substantial impact on the care of patients with addictive disorders. Addressing substance abuse issues among patients required multidisciplinary efforts and thus multi-disciplinary training was required to achieve this goal. The recognition that other health-care professions had a direct stake in clinical education led to the broadening of AMERSA's multidisciplinary base to faculty in all medical, nursing, social work and other health professional training programs. Gradually, nurses, social workers, dentists, allied health professionals and others became part of the organization. They started as active participants in the annual conference, then active members, and then active Executive Committee members—the leaders of the organization. Being multi-disciplinary is one of the great strengths of the organization, distinguishing AMERSA from organizations with physician-only or psychologist-only membership. This organizational hallmark encourages clinicians to take a patient-centered or family-centered perspective and enables members to discuss interdisciplinary training, a focus not consistently pursued by other substance abuse organizations. AMERSA's members come from a range of disciplines and health professions and membership has grown to over 300; the organization's President (2003–05) is a senior faculty leader in a School of Social Work.

Leadership structure

AMERSA's Executive Committee, represented by a variety of professions, is responsible for setting the direction of the organization. The Executive Committee consists of President, Vice President, Immediate Past President,

Secretary, Treasurer and *Substance Abuse* Journal Editor-in-Chief, four Members-at-Large, Director and two Co-Directors. Officers take office at the conclusion of the national meeting following an election that has occurred a few months prior, and serve for a period of 2 years. No officer can serve on the Executive Committee for more than 8 consecutive years, excluding the 2-year term as Immediate Past President. For a current listing of officers see the website www.amersa.org.

Joining AMERSA

Full membership is open to people engaging in substance abuse research or education and to faculty of health professional schools. AMERSA also offers associate, corporate and emeritus membership. Members' range of benefits include the following: reduced rates for the annual national conference; a subscription to *Substance Abuse*; and a national voice supporting academic programs in universities, professional schools, and organizations that emphasize substance abuse education and research.

CHALLENGES

As in many non-profit organizations, AMERSA faces the ongoing challenge of limited financial resources. In general, support in the addiction area is directed at treatment and research rather than education and training. Training is likely to be conducted by people involved in treatment and research, but only limited sources of funding have been traditionally available specifically for teaching efforts. Because of this, AMERSA members, through their commitment to the organization's educational mission, must creatively garner cooperation from many faculty members and operate with limited resources to fulfill their goals.

Compared to other organizations in the United States that focus on substance abuse AMERSA has always had a relatively small membership, in part reflecting funding for educational efforts. The organization's strength is that this group is committed, talented, collaborative and imbued with the spirit to provide guidance to junior and peer colleagues. It has survived and flourished, in part, because it is the only organization that focuses on the educational mission in the way that it does. Many within its committed membership are the leaders within health professional schools nation-wide; they are the teachers of substance abuse at the nation's major universities, hospitals and health-care institutions. They educate and mentor future clinicians, researchers and educators, creating an impact well beyond their direct sphere of influence. They are in the forefront leading this effort. Even though AMERSA's numbers are not in the thousands, the organization has a big ripple effect on health professional substance abuse professional training.

FUTURE OF AMERSA

Members of AMERSA will continue to pursue the organization's education and training goals, including the development of a national infrastructure for interdisciplinary faculty development. Teaching about substance abuse needs to become mainstream and rooted securely in health professional schools. New faculty at these institutions must be inspired, well trained and supported so that students have a respected faculty source and role model for integrating substance abuse prevention, intervention and treatment into their daily work. These goals are the essence of what AMERSA will pursue in the coming years through its conferences, journal and training programs.

AMERSA was a leading participant in the development of a strategic planning document [7] to guide the improvement of health professional education on substance abuse. The future of AMERSA depends on the wide recognition of the problems of alcohol and drug abuse and dependence in society. It depends on the shift away from stigmatizing and towards understanding these problems as health issues, which has been occurring over the last 30 years. The efforts towards competency-based education [16], with a focus towards outcomes rather than process, should bring more attention to the field because the problems of substance abuse are so commonplace in clinical practice. The high prevalence of alcohol and other drug problems in both hospital and ambulatory practice will be a potent motivating force, as it has not yet been addressed adequately. As the trend toward skills training and competency based health professional education continues and the stigma of alcohol and drug abuse decreases, it is hoped that the training of doctors, nurses, social workers, dentists, allied health professionals and other clinicians about substance abuse will be more widely recognized as essential to a quality education in these disciplines. At that time, the human resources that the AMERSA membership and organization provide will become even more valued.

The current younger generation of students seems to recognize the importance of appropriate training in substance abuse. A recent and encouraging illustration is the effort by students to form their own multi-disciplinary group called Health Professional Students for Substance Abuse Training. They have taken initiatives to expand substance abuse education in their institutions and have created their own website (www.HPSSAT.ORG) in order to provide curriculum and training opportunities. These

students are now forming an alliance with AMERSA. Thus it is anticipated that AMERSA will see a cohort of younger members in the next 5 years that will eventually become the leadership. AMERSA welcomes such change as the organization views mentorship as a core organizational value. The new generation of leaders will come from a different style of health professional education than the current generation of AMERSA members. The continuing challenge is to integrate substance use disorders effectively into the traditional curriculum so that students will gain competence in these common problems that take a heavy toll on the health of individuals and families in our society.

Acknowledgements

We gratefully acknowledge the contributions of Richard Saitz MD, MPH, President-Elect, AMERSA, Maryann Amodeo MSW, PhD, President, AMERSA and Isabel Vieira Co-Director, AMERSA who either critically reviewed or assisted with the gathering of material for the preparation of this manuscript.

Author's note

To contact AMERSA or to request reprints, the address is Association for Medical Education and Research in Substance Abuse, 125 Whipple Street, 3rd Floor, Suite 300, Providence, RI 02908, USA. The website address is http://www.amersa.org

References

1. Schnoll, S. & Horvatich, P. (1989) A training model for student research in alcohol and other drug problems. *Alcohol Health and Research World*, **13**, 20–23.
2. Project SAEFP, Executive Committee and Working Group (1991) *Project SAEFP Workshop Materials*. Kansas City: Society of Teachers of Family Medicine.
3. Galanter, M. (1989) Alcohol and drug abuse as a subspecialty: credentialing and specialization. *Alcohol Health and Research World*, **13**, 42–47.
4. Goldstein, M. G. & Dube, C. E. (1989) *Substance Abuse Prevention: Instructor's Guide. The Project ADEPT Curriculum for Primary Care Physician Training*. Providence: Brown University.
5. Lewis, D. (1989) Putting training about alcohol and other drugs into the mainstream of medical education. *Alcohol Health and Research World*, **13**, 8–13.
6. Fleming, M., Cotter, F. & Talboy, E. (1997) *Training Physicians in Techniques for Alcohol Screening and Brief Intervention*. Rockville, MD: NIAAA.
7. Association for Medical Education and Research in Substance Abuse (AMERSA) (2002) Strategic plan for interdisciplinary faculty development: arming the nation's health professional workforce for a new approach to substance use disorders. *Substance Abuse*, **23** (Supplement), 1–345. Also available at: http://www.projectmainstream.net/mainstream/supportdata/spacdocfinal.pdf. Accessed 5 May 2005.
8. Galanter, M. (1980) *Alcohol and Drug Abuse in Medical Education*. Washington, DC: Government Printing Office.
9. Association for Medical Education and Research in Substance Abuse (AMERSA) (2004) *Organizational Profile*. Available at: http://www.amersa.org/OrgProf.asp. Accessed 22 November, 2004.
10. Josiah Macy Jr Foundation (1995) Training about alcohol and substance abuse for all primary care physicians. *Proceedings of a Macy Conference held 2–5 October 1994, New York*.
11. Alcohol, Drug Abuse and Mental Health Administration (ADAMHA) (1988) Report on a Symposium cosponsored by the National Institute on Alcohol Abuse and Alcoholism, and the National Institute on Drug Abuse. *ADAMHA News*, 14 (Supplement).
12. Alcohol, Drug Abuse and Mental Health Administration (ADAMHA) (1985) Consensus statements from the Conference Alcohol, Drugs, and Primary Care Physician Education: Issues, Roles, Responsibilities, 12–15 November 1985, Rancho Mirage, California. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.
13. World Health Organization (WHO) (1992) *The Role of General Practice Settings in the Prevention and Management of the Harm Done by Alcohol Use*. Report on a WHO Meeting. Copenhagen: WHO Regional Office for Europe. Available at: http://whqlibdoc.who.International/euro/-1993/EUR_ICP_ADA_038.pdf.
14. Stoil, M. (1989) *Alcohol-Related Topics in Medical Education: Activites at the* Karolinska Institute. *Alcohol Health & Research World*, 13, 88–89. Also available at: http://www.findarticles.com/p./articles/mi_m0847/is_n1_v13/ai_7650857.
15. Rankin, J. (2000) Conversation with Jim Rankin. *Addiction*, **95**, 667–676.
16. Accreditation Council for Graduate Medical Education (2000) *Outcome Project. Enhancing Residency Education Through Outcomes Assessment*. Available at: http://www.acgme.org/Outcome/. Accessed 26 January 2005.

A randomized controlled trial to enhance antiretroviral therapy adherence in patients with a history of alcohol problems

Jeffrey H Samet[1–3], Nicholas J Horton[4], Seville Meli[1], Kim Dukes[5], Tara Tripp[5], Lisa Sullivan[6] and Kenneth A Freedberg[7,8]*

[1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, [2]Departments of Medicine, [3]Social and Behavioral Sciences, [6]Biostatistics and [7]Epidemiology, Boston Medical Center, Boston University Schools of Medicine and Public Health, Boston, MA, USA
[4]Department of Mathematics, Smith College, Northampton, MA, USA
[5]DM-STAT, Medford, MA, USA
[8]Division of General Medicine and Partners AIDS Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

*Corresponding author: Tel: +1 617 414 7288; Fax: +1 617 414 4676; E-mail: jsamet@bu.edu

Objective: To assess the effectiveness of an individualized multicomponent intervention to promote adherence to antiretroviral therapy (ART) in a cohort of HIV-infected individuals with a history of alcohol problems.
Design: We conducted a randomized controlled trial to compare the usual medical follow-up with an adherence intervention.
Setting: The principal enrolment site was Boston Medical Center, a private, not-for-profit, academic medical institution.
Subjects: HIV-infected patients with a history of alcohol problems on ART. A total of 151 were enrolled and 141 (93%) were assessed at follow-up.
Intervention: A nurse, trained in motivational interviewing, completed the following over 3 months in four encounters: addressed alcohol problems; provided a watch with a programmable timer to facilitate pill taking; enhanced perception of treatment efficacy; and delivered individually tailored assistance to facilitate medication use.
Main outcome measures: Prior 30-day adherence ≥95%, prior 3-day adherence of 100%, CD4 cell count, HIV RNA and alcohol consumption, each at both short- and long-term follow-up.
Results: At follow-up, no significant differences in medication adherence, CD4 cell count, HIV RNA or alcohol consumption were found (all P values >0.25).
Conclusions: A multicomponent intervention to enhance adherence among HIV-infected individuals with a history of alcohol problems was not associated with changes in medication adherence, alcohol consumption or markers of HIV disease progression. The failure to change adherence in a group at high risk for poor adherence, despite utilizing an intensive individual-focused patient intervention, supports the idea of addressing medication adherence with supervised medication delivery or markedly simplified dosing regimens.

Introduction

Poor medication adherence is a problem across the spectrum of medical conditions and its degree varies with both illness and patient population [1]. In general, medication adherence has been reported as being between 50% and 65% for long-term regimens and as low as 18% for acute illnesses [2,3]. Historically, several variables have been associated with medication adherence: severity of the disease, complexity of the prescribed regimen, physician–patient relationship, potential for painful or undesirable side effects, patient's lifestyle and substance abuse [4]. Prior to the AIDS epidemic, adherence research demonstrated that effective intervention to enhance medication adherence must address multiple dimensions in a patient's life [5].

Antiretroviral therapy (ART) has significantly reduced morbidity and mortality from HIV infection [6]. Medication regimens are most effective when adherence is at a level of 95% or higher [7,8]. Decreased adherence to HIV medications has been associated with lower CD4 cell counts, higher HIV RNA levels and more rapid disease progression [8–11]. Decreased adherence can increase the development of drug-resistant strains of

HIV [12,13]. Antiretroviral therapy may involve complicated regimens with potentially minor or major toxicity, factors that can promote non-adherence [14].

Alcohol abuse has also been associated with poorer medication adherence in HIV-infected individuals and heavy drinking is associated with taking ART off schedule [15]. One study found that among HIV-infected individuals with a history of alcohol problems, current alcohol consumption was the factor most strongly associated with decreased adherence [16].

Medication adherence has assumed a central role among HIV therapeutic recommendations [17,18]. A critical need exists to find methods to achieve optimal adherence and to provide methodologically sound assessments of these strategies [19]. Randomized controlled trials in general populations with HIV infection have tested psychoeducational interventions to promote adherence with varying results [10,20,21]. In an intention-to-treat analysis by Tuldra *et al.*, there was no significant difference in medication adherence between the experimental and control groups, although subjects receiving the intervention reported increased adherence [10]. Goujard *et al.* achieved significant increases in adherence but did not see improvements in clinical indicators [20]. Pradier *et al.* not only achieved significant improvements in adherence, but also decreases in viral load [21]. As yet, no randomized controlled trial to enhance medication adherence has been reported in an HIV-infected population with a history of alcohol problems. Our objective was to assess the effectiveness of an individualized multicomponent intervention to promote adherence to ART in a cohort of HIV-infected individuals with a history of alcohol problems.

Methods

Study design and recruitment

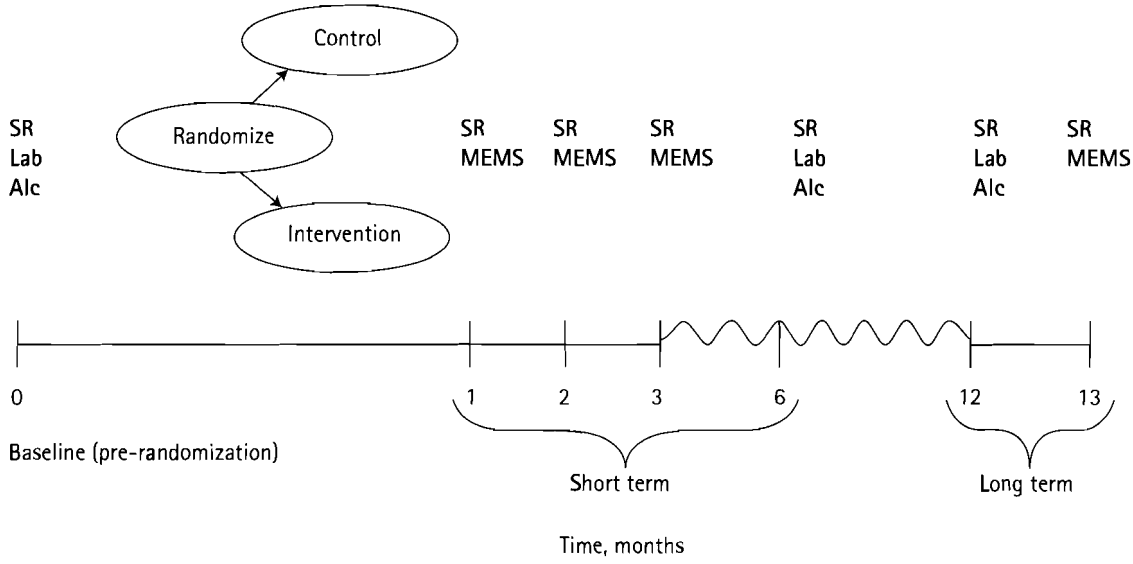
This study, ADHERE (Adherence to Drugs for HIV, an Experimental Randomized Enhancement), employed a randomized controlled trial design to study a multicomponent intervention to enhance adherence to antiretroviral therapy in HIV-infected patients with a history of alcohol problems. Recruitment of subjects from several sites occurred from July 1997 to December 2000. The principal enrolment site was a Boston Medical Center clinic in which patients initiating HIV medical care underwent an extensive clinical evaluation [22].

Subjects in a previously described longitudinal cohort study, HIV-ALC (HIV Alcohol Longitudinal Cohort), were eligible for this trial if they were taking antiretroviral medication [23]. HIV-ALC subjects were HIV-infected and had a history of alcohol problems identified by explicit criteria. HIV infection was confirmed by a positive laboratory test result demonstrating infection or antiretroviral medication use. Alcohol problems were defined as current or lifetime history of alcohol abuse or dependence and were determined by two or more positive responses to the CAGE screening questionnaire [24,25]. Patients recruited from the principal enrolment site who did not meet CAGE criteria were eligible if one of two attending physicians (JHS, KAF) made a clinical diagnosis of alcohol abuse or dependence. Other entry criteria included the following: fluency in English or Spanish, Mini-Mental State Examination score ≥21 [26] and no plans to move from the Boston area in the following 2 years. Subjects meeting eligibility criteria completed initial assessments and were recruited to participate in the randomized controlled trial (Figure 1). Subjects were randomly assigned to either the control or intervention group and balance between groups was ensured after every four subjects were enrolled. The Institutional Review Boards of the Boston Medical Center and the Beth Israel Deaconess Medical Center approved this study.

Subject assessment

After providing informed consent for the ADHERE study, all subjects were interviewed by trained research associates. Interviews occurred at seven time points over 13 months and took place at the hospitals' General Clinical Research Centers. Prior to randomization, a comprehensive evaluation ascertained demographic, HIV risk behaviour, alcohol and drug consumption, depressive symptoms (CES-D) [27], quality of life (SF-36) [28] and self-reported 3- and 30-day ART adherence information [29]. At the 6- and 12-month time points, alcohol consumption was reassessed. Interviews were conducted 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, back-translated to check for accuracy and then corrected. We used routine clinical results of CD4 cell counts and HIV RNA levels when available at the pre-randomization, 6- and 12-month evaluations. When clinical values within a 3-month window were unavailable, we obtained results during the research interview. A survey was administered at the 0-, 1-, 2-, 3-, 6-, 12- and 13-month visits to assess the effectiveness of the intervention. To corroborate self-report of ART adherence, two Medication Event Monitoring System (MEMS) caps were distributed to each subject at the baseline contact, collected at the 3-month contact, redistributed at the 12-month contact and recollected at the 13-month contact. MEMS caps record the precise date and time of every bottle opening [30]. Assuming that a bottle opening represents a dose

Figure 1. Timeline for data recorded during the study



SR, self-reported medication adherence; Lab, clinical laboratory data (for example, CD4 and HIV RNA); Alc, alcohol consumption; MEMS, electronic recording of medication adherence [Medication Event Monitoring System (MEMS) caps were distributed at baseline, collected at month 3, redistributed at month 12 and recollected at month 13].

taken, this approach theoretically provides continuous and accurate dose monitoring.

ADHERE intervention

The ADHERE intervention incorporated four distinct components: i) assessment and discussion of the patient's alcohol and other substance use based on stage of readiness for behavioural change [31]; ii) use of a watch that served as a medication timer device, a practical aid to improving adherence; iii) enhancement of perceived efficacy of medications; and iv) individualized HIV counselling and exploration of ways to tailor medication use to specific circumstances (Table 1). Subjects randomized to the intervention group were scheduled for an initial 60-min individual appointment (within 2 weeks after randomization), a follow-up home visit within the first 3 weeks and two subsequent 15–30-min appointments at 1 month and 3 months with the nurse interventionist who delivered the adherence enhancement intervention. The nurse had prior HIV and substance abuse expertise and was specifically trained in motivational interviewing to address substance abuse and HIV medication adherence [32]. A structured approach was used by the nurse to assess the subject's substance use based on the individual's readiness to change [31]. The nurse instructed the subject on the use of the medication timer device provided by the study – a programmable watch that sounded an auditory or vibratory alarm at the time scheduled for the subject's ART dose. The intervention included a review of the efficacy of HIV medications, answering subject's HIV-related questions and exploring ways to tailor medication use to specific circumstances (for example, homelessness, access to refrigeration, non-disclosure of HIV status and social support). The purpose of the home visit was for the nurse interventionist to observe the subject's living situation so that his or her particular barriers to medication adherence could be addressed. During the intervention visits, the nurse re-assessed barriers to medication adherence and counselled about strategies to overcome these barriers. Subjects in the intervention group who received at least one, but not all four nurse intervention encounters, were identified as having received a 'partial intervention'.

The same nurse delivered all nursing interventions and, after each session, recorded if the pre-specified components of the adherence intervention had been delivered. A second nurse consultant experienced in HIV and substance abuse observed three interventions to assess and provide feedback to the nurse interventionist and verify that all pre-specified elements of the intervention were addressed.

Subjects randomized to the control group received standard care for HIV infection. During this period standard practice with regard to ART was evolving, and typically included verbal or written instructions

Table 1. Components of the ADHERE intervention

	Initial visit	Follow-up visit (during first 3 weeks in the intervention)	1-month visit	3-month visit
Site	At Boston Medical Center or Beth Israel Deaconess Medical Center	At patient's home	At Boston Medical Center or Beth Israel Deaconess Medical Center	At Boston Medical Center or Beth Israel Deaconess Medical Center
Time	60 min	30–45 min	15–30 min	15–30 min
Intervention	Component 1: Assessment and discussion of the patient's alcohol and other substance use based on stage of readiness for behavioural change	Facilitation with and practical implementation of ART medication	Reinforcement of components addressed during initial visit (for example, enhancement of perceived efficacy of medications, provision of individualized HIV counselling)	Reinforcement of components addressed during initial visit (for example, enhancement of perceived efficacy of medications, provision of individualized HIV counselling)
	Component 2: Distribution of and education on medication timer device	Reassessment of the patient's personal situation and experience with medications and timer use over the prior week	Reassessment of the patient's personal situation and experience with medications and timer use over the prior week	Reassessment of the patient's personal situation and experience with medications and timer use over the prior week
	Component 3: Enhancement of perceived efficacy of medications	Provision of positive feedback or development of a more suitable plan of action by strategizing with the patient	Provision of positive feedback or development of a more suitable plan of action by strategizing with the patient	Provision of positive feedback or development of a more suitable plan of action by strategizing with the patient
	Component 4: Provision of individualized HIV counselling and exploration of ways to tailor medication use to specific circumstances			

about optimal medication adherence strategies [33–35]. Subjects continued to receive their regular medical care for HIV infection.

Primary outcome

The primary study outcome was adherence to ART. Both 3-day and 30-day self-reported adherence were measured using the AIDS Clinical Trials Group (ACTG) scale [29]. We defined desired adherence as ≥95% adherence over the previous 30 days and 100% adherence over the previous 3 days. In addition to the dichotomous measure of adherence, additional analyses were conducted using a continuous measure of adherence for the past 30 days. This measure is the actual proportion of doses taken versus doses prescribed (0.0–1.0).

We attempted to obtain both short- and long-term measurements for each of the outcomes. Short-term follow-up was defined as the 6-month interview result if available, otherwise the most proximal measurement was used (that is, the 3-month measurement if available; if not, then the 2-month one). Long-term follow-up was defined as the 13-month interview if available, otherwise the 12-month interview was used.

We corroborated self-reported adherence results using MEMS caps [30]. At the time of medication adherence assessment, subjects were asked about their pill-taking routine. Subjects who reported no change in HIV medications during the assessed interval and did not use a pill organizing container were included in the analyses corroborating self-report of adherence.

Secondary outcomes

We examined several secondary outcomes at both 6 and 12 months: CD4 cell count, log HIV RNA, and alcohol severity and consumption using both the Addiction Severity Index [36] and quantity and frequency questions assessing the previous 30 days (that is, the number of drinking days and maximum

number of drinks in a day). These responses were used to calculate the average number of drinks per day and dichotomous measures of abstinence and hazardous drinking using National Institute on Alcohol Abuse and Alcoholism guidelines [37].

Statistical analysis

To test for group differences in baseline characteristics, we carried out two-sample *t*-tests for continuous variables and Fisher's exact tests for categorical variables. Measures of agreement between self-report and MEMS were calculated using the kappa statistic for dichotomous reports of adherence and intraclass correlation for continuous reports.

Analyses of primary and secondary outcomes were conducted using generalized estimating equations [38] to test for intervention effects at both short-term and long-term follow-up. These analyses controlled for baseline measurements of adherence for models of adherence, laboratory values for CD4 cell count and HIV RNA models, and alcohol consumption for alcohol models. These longitudinal regression models account for the correlation between observations taken over time on the same subject, using an exchangeable working correlation matrix and empirical variance estimator [39]. For each outcome, two models were fit, one with an interaction between the intervention group and time (to test for differences between short-term and long-term intervention effect) and one with main effects for time and group (which averages the intervention effect over time). In addition to the main analyses, a series of secondary analyses were performed to assess homogeneity of effect. Stratified analyses were performed to assess differences in important subgroups defined at baseline: gender, hazardous drinking, adherence ≥95%, injection drug use in the previous 6 months, undetectable HIV RNA (<500) and CD4 cell count ≤350 cells/mm^3.

All analyses were performed by the intention-to-treat method. Reported *P* values are two-tailed and a *P* value <0.05 was considered statistically significant. No adjustment for multiple comparisons was made. The study was designed to enrol 240 subjects (120 per group). Power calculations assumed that 10% of subjects would be lost at 6 months and an additional 10% at 12 months (n=108 per group and n=97 per group, respectively) and that the standard deviation of adherence was 21%. The study was designed to have 80% power to detect a difference in adherence of 8% at short-term follow-up between groups (for example, 84% in intervention group vs 76% in controls) and 84% power to detect a difference of 9% at long-term follow-up. All analyses were performed using SAS statistical software v8.2 (SAS Institute, Inc., Cary, NC, USA) [39].

Results

A total of 151 ADHERE subjects were randomized to either the control or intervention group (Figure 2 and Table 2). Of these, 67 subjects were recruited from the principal enrolment site with an additional 29 from Boston Medical Center's primary care practices, three from Beth Israel Deaconess Medical Center, 10 from a respite facility for homeless individuals, five from a methadone clinic and 37 referred by friends or posted flyers. At baseline, the control group and the adherence intervention group were comparable in terms of sex, race, age and mean log HIV RNA. Intervention subjects had significantly higher mean CD4 cell counts than control subjects (480 vs 364 cells/mm^3; P=0.01). Among the intervention subjects, 93% received at least one component of the intervention and 76% received all four pre-specified encounters of the intervention.

Follow-up outcome data were obtained in a high percentage of subjects for the short-term time point in both control and intervention subjects (95% and 88%, respectively) and in a moderate percentage for the long-term time point (58% and 66%, respectively). No significant differences in the primary outcome (that is, medication adherence) or other secondary outcomes (CD4 cell count, HIV RNA or alcohol consumption) were found using the longitudinal regression models either at short-term or long-term follow-up (all *P* values >0.25). Subgroup analyses were consistent with primary analyses.

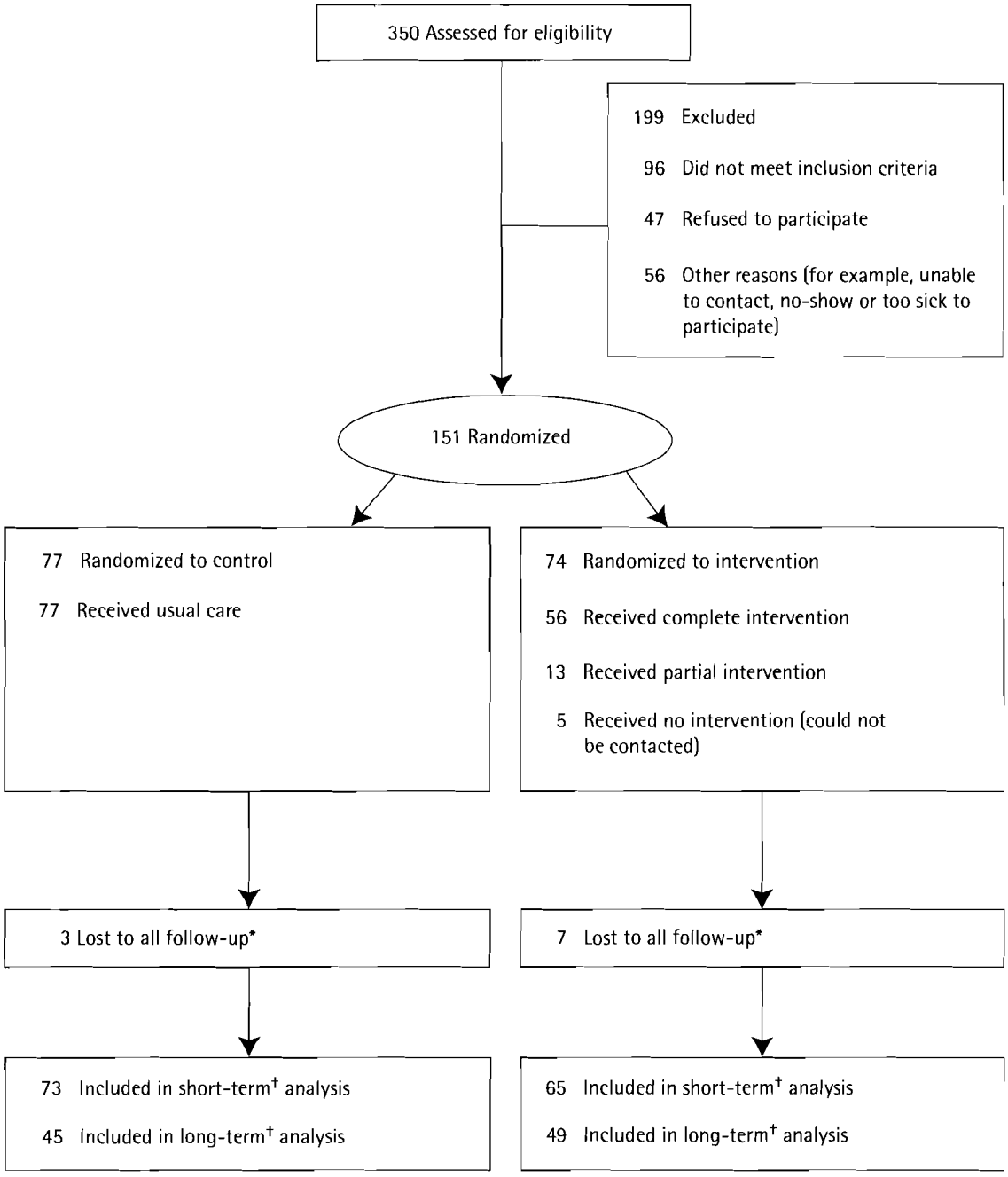
Because of challenges in subject recruitment, the study enrolled only 151 subjects, not the 240 planned, and as a result the study had less power than originally designed. Using the assumptions described in the Methods section with this smaller sample size, the expected power to detect differences between groups was 60% at short-term and 65% at long-term follow-up.

Because a large number of subjects used pill-counting devices or switched medications during the study, only 29% of subjects had usable baseline MEMS data for corroboration with self-reported data. This was a previously reported limitation of MEMS use in community-based clinical trials [40]. Similar to Arnsten *et al.* [41], we found among this subset of subjects that, although self-report of medication adherence at baseline was higher than that calculated using MEMS data (P<0.001), there was a slight association between 30-day self-report data and MEMS data (intraclass correlation=0.26; kappa=0.19) [42].

Discussion

Improving adherence to ART in patients at high risk for poor adherence is challenging [3]. We developed and assessed an individually-tailored multicomponent

Figure 2. Enrolment and follow-up of subjects



*Reasons for being lost to follow-up: not interested, no time, enrolled in another study. †Short term: months 1–6; long-term: months 12–13.

intervention to enhance adherence to ART among HIV-infected individuals with a history of alcohol problems. Using behavioural science theories, the intervention was designed to be potent, utilizing interviewing principles of motivational enhancement to promote behaviour change, as well as Health Belief Model principles to support the benefit and need for therapy [43]. Despite consistently delivering this pragmatic theory-based intervention, we found no significant difference in the primary outcomes of either short-term or long-term adherence between the intervention and control groups. In addition, subjects

Table 2. Characteristics of control and intervention subjects in a randomized controlled trial of a multicomponent intervention to enhance adherence to antiretroviral treatment (ADHERE Study)

	Control (n=77)	Intervention (n=74)	P value
Male	84%	78%	0.40
Age in years, mean (SD)	43.2 (7.4)	42.5 (7.9)	0.53
Race/ethnicity			0.44
White	26%	34%	
Black	52%	42%	
Other	22%	24%	
HIV transmission risk			0.67
Heterosexual sex	21%	15%*	
Injection drug user	57%	60%*	
Men who have sex with men	22%	25%*	
Homeless	25%	20%	0.56
CES-D score, mean (SD)	22.6 (13.0)	21.9 (13.3)	0.74
Pillbox user			0.77
None	59%	53%	
Daily/>daily	30%	36%	
Weekly/other	11%	11%	
CD4 cell count, cells/mm^3, mean (SD)	364 (263)*	480 (256)*	0.01
Log HIV RNA, copies/ml, mean (SD)	2.2 (1.7)*	1.9 (1.8)*	0.26
Undetectable HIV RNA (<500)	28%	30%	0.14
AIDS†	31%	24%	0.24
ART doses per day, mean (SD)	4.6 (1.7)	4.8 (1.6)*	0.53
On protease inhibitors‡	26%	27%	0.63
Number of ambulatory care visits, prior 6 months, mean (SD)	8.4 (11.4)	7.6 (6.5)	0.61
Received substance abuse treatment, prior 6 months	61%	65%	0.74
Cocaine use, prior 30 days	13%	11%	0.57
Heroin use, prior 30 days	5%	4%	1.00
Injected drugs, prior 6 months	6%	7%	0.81

*n range from 74–76 for control and 67–73 for intervention subjects. †Defined per CDC guidelines [55]. ‡Amprenavir, indinavir, saquinavir, lopinavir/ritonavir, ritonavir or nelfinavir. SD, standard deviation. CES-D, Center for Epidemiological Studies-Depression.

randomized to the intervention had no demonstrable benefit in terms of the secondary outcomes of CD4 cell count, HIV RNA or alcohol use. The absence of any difference between groups, comparing all primary and secondary outcomes, is robust, as we measured outcomes at both short-term follow-up and long-term over the course of 13 months and corroborated adherence self-report data with available MEMS data.

Previous adherence research not involving HIV-infected individuals describes several useful strategies for interventions to increase medication adherence. Traditional methods focus on communication between the physician and patient, written instruction, patient education about the illness and medication, physician advice, special drug packaging and tailoring medications to a patient's daily schedule [3]. Early adherence research in HIV disease found that adherence to zidovudine was associated with a patient's belief that this medication would prolong life [44]. Recent factors associated with HIV adherence include the patient's correct understanding of the effect of HIV medications on viral load [45], alcohol use and active drug use [46], race [47], perceived self-efficacy to comply with the prescribed regimen, physician–patient relationship, older age, lifestyle and other psychosocial factors [48]. According to the Health Belief Model, the decision to adhere to medications depends upon the patient's feeling of susceptibility to the disease, belief in the severity of the disease, belief in the efficacy of the medication to prevent disease progression and perceived barriers to initiating or continuing medication adherence, such as side effects [43].

Despite the rational construction and effective delivery of the adherence intervention in this study, we found no effect of its use. These results argue against a strategy of strengthening the skills of individual patients as the major approach to improving antiretroviral medication adherence. It is possible that the efficacy of our adherence intervention was underestimated. Approximately a quarter (24%) of the

intervention group received only a partial intervention or no intervention. Although this represents only a minority of the intervention group, this incomplete exposure of the intervention to subjects randomized to that group would reduce the demonstrated potency of the intervention. Another potential explanation for the lack of efficacy of the intervention is that the target enrolment was not met. Thus, the study was not completely powered to detect differences in the primary outcome. However, the lack of any trend towards a difference in any of the primary or secondary short-term or long-term outcomes argues against the likelihood that enrolling additional subjects in each group would have yielded different results.

Study entry criteria yielded cohort characteristics that worked against the demonstration of potential effectiveness of this intervention. The focus on HIV-infected individuals with a history of alcohol problems was important, as alcohol has consistently been implicated as a factor associated with poor ART adherence. We hypothesized that an effective adherence intervention would need to directly address the impact of alcohol on adherence in these patients. However, this approach limited available subjects. In order to facilitate recruitment, study entry criteria were not limited with regard to current adherence, CD4 cell count, HIV RNA or alcohol use. This strategy allowed recruitment of a sizable cohort but risked both ceiling and floor effects that might limit the ability to demonstrate a change in study outcomes. For example, as reflected in Table 3, 30-day adherence at baseline was already ≥95% in 69% of control subjects and 68% of intervention subjects. Thus, the potential to demonstrate an improvement in this outcome relied on the 32% of subjects with baseline adherence being in need of substantial improvement. Comparable floor effects were encountered for HIV RNA, for which 54% of control and 64% of intervention subjects were undetectable at baseline.

Of note, during the time this study was carried out (1998–2001), adherence was increasingly being recognized as an issue requiring clinical attention. Standard medical care was beginning to focus on improving adherence, making it more challenging for a specific intervention to demonstrate even better adherence.

Improving medication adherence is difficult. Past research has shown behavioural interventions can increase medication adherence from low to moderate [1]. However, with antiretroviral therapy, the aim is to achieve greater than 95% adherence, an extremely high level in comparison with that needed for successful treatment of other illnesses [8]. It remains unclear what it will take to achieve this objective in HIV care. The ability to improve patients' knowledge about medication adherence has been previously demonstrated [49]. However, as in our study, it has also been shown that improving patients' knowledge about HIV/AIDS and its treatment may not be enough to change behaviour and increase medication adherence [49]. Intensive adherence interventions in the setting of HIV infection may be appropriate, as recent cost-effectiveness analyses have illustrated that it would even be cost-effective to institute expensive interventions, if improved adherence resulting in better viral suppression can be achieved [50]. Although the suggestion of implementing focused interventions to address risk factors for non-adherence in HIV-infected patients [51,52] is theoretically sound, attempts to improve adherence in this way have not been demonstrable. In the current study, a defined, repeated, multicomponent intervention to enhance adherence in HIV-infected individuals with a history of alcohol problems was not able to improve adherence or to significantly affect markers of HIV disease progression or alcohol consumption.

To enhance adherence to ART in challenging populations effectively, systematic interventions may be more effective than individually focused interventions. Directly observed therapy with the availability of additional services has been shown to improve adherence during supervised medication administration in HIV-infected drug users [53]. Another systematic intervention is regimen simplification. In a study of self-report perceptions, HIV-infected patients perceived that a twice-daily two-pill medication regimen with no dietary restrictions would yield higher adherence than more complicated regimens [54]. Future efforts to enhance adherence to antiretroviral medications in challenging populations, such as those with a history of alcohol problems, should focus on directly observed therapy and/or markedly simplified dosing regimens.

Grant support

Grant support for this study came from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) of the NIH: RO1-AA11785 (Medication Adherence in Alcohol Abusing HIV Patients) and 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.

Acknowledgements

The authors appreciate the contributions of the clinical staff of the HIV Diagnostic Evaluation Unit at Boston Medical Center, including Colleen LaBelle, RN (nurse specialist) and Michael O'Connor, RN (research interventionist), as well as assistance

Table 3. Observed adherence, laboratory and alcohol consumption outcomes at baseline and over time

	Control	Intervention	*P* value*
Adherence			
% Reporting 100%			
3-day adherence			
Baseline	65% (*n*=77)	58% (*n*=72)	0.50
Short-term	63% (*n*=73)	65% (*n*=65)	0.86
Long-term	62% (*n*=45)	71% (*n*=49)	0.39
% Reporting ≥95%			
30-day adherence			
Baseline	69% (*n*=77)	68% (*n*=72)	1.00
Short-term	62% (*n*=73)	63% (*n*=65)	1.00
Long-term	64% (*n*=45)	67% (*n*=49)	0.83
30-day adherence†, mean (SD)			
Baseline	0.92 (SD=0.18, *n*=77)	0.92 (SD=0.15, *n*=72)	0.94
Short-term	0.83 (SD=0.31, *n*=73)	0.87 (SD=0.23, *n*=65)	0.42
Long-term	0.91 (SD=0.15, *n*=45)	0.91 (SD=0.17, *n*=49)	0.95
Laboratory values			
CD4 cell count, mean (SD)			
Baseline	364 (SD=263, *n*=74)	480 (SD=256, *n*=67)	0.01
Short-term	374 (SD=222, *n*=61)	479 (SD=283, *n*=52)	0.03
Long-term	362 (SD=227, *n*=47)	512 (SD=341, *n*=47)	0.01
Log HIV RNA, mean (SD)			
Baseline	2.2 (SD=1.7, *n*=76)	1.9 (SD=1.8, *n*=69)	0.26
Short-term	2.4 (SD=1.7, *n*=61)	2.0 (SD=1.7, *n*=52)	0.29
Long-term	2.5 (SD=1.7, *n*=45)	2.7 (SD=1.7, *n*=48)	0.44
Drinking			
Drinks/day, mean (SD)			
Baseline	2.1 (SD=6.0, *n*=77)	1.2 (SD=3.2, *n*=73)	0.26
Short-term	1.8 (SD=4.1, *n*=65)	1.9 (SD=4.5, *n*=52)	0.91
Long-term	1.6 (SD=3.3, *n*=47)	2.0 (SD=4.9, *n*=50)	0.62
% Reporting any drinking			
Baseline	47% (*n*=77)	50% (*n*=74)	0.75
Short-term	48% (*n*=65)	54% (*n*=54)	0.58
Long-term	60% (*n*=47)	54% (*n*=50)	0.68
% Reporting any hazardous drinking			
Baseline	34% (*n*=77)	37% (*n*=73)	0.73
Short-term	38% (*n*=65)	40% (*n*=52)	0.85
Long-term	34% (*n*=47)	40% (*n*=50)	0.67

*Univariate *t*-test or Fisher's exact test at each time point. †Reported proportion adherence (0.1–1.0) during prior 30 days. SD, standard deviation.

provided by research associates on the project. We also acknowledge intellectual contributions to this project by Emily Williams, MPH and Carly Bridden, MA, MPH. Finally we appreciate the efforts of Dr Howard Libman and the clinical staff at the Beth Israel Deaconess Medical Center.

References

1. Haynes RB, Taylor DW & Sackett DL (Editors). *Compliance in Health Care*. 1979. Baltimore: Johns Hopkins University Press.
2. Bergman AB & Werner RJ. Failure of children to receive penicillin by mouth. *New England Journal of Medicine* 1963; **268**:1334–1338.
3. Melkinow J & Kiefe C. Patient compliance and medical research. *Journal of General Internal Medicine* 1994; **9**:96–105.
4. Matthews D & Hingson R. Improving patient compliance: a guide for physicians. *Medical Clinics of North America* 1977; **61**:879–889.
5. Haynes RB, McKibbon KA & Kanani R. Systematic review of randomised trials of interventions to assist patients to follow prescriptions for medications. *Lancet* 1996; **348**:383–386.
6. Hogg RS, Yip B, Kully C, Craib KJ, O'Shaughnessy MV, Schechter MT & Montaner JS. Improved survival among HIV-infected patients after initiation of triple-drug antiretroviral regimens. *Canadian Medical Association Journal* 1999; **160**:659–665.
7. Bartlett JA. Addressing the challenges of adherence. *Journal of Acquired Immune Deficiency Syndromes* 2002; **29**:S2–10.

8. Paterson DL, Swindells S, Mohr J, Brester M, Vergis EN, Squier C, Wagener MM & Singh N. Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. *Annals of Internal Medicine* 2000; **133**:21–30.
9. Garcia de Olalla P, Knobel H, Carmona A, Guelar A, Lopez-Colomes JL & Cayla JA. Impact of adherence and highly active antiretroviral therapy on survival in HIV-infected patients. *Journal of Acquired Immune Deficiency Syndromes* 2002; **30**:105–110.
10. Tuldra A, Fumaz CR, Ferrer MJ, Bayes R, Arno A, Balague M, Bonjoch A, Jou A, Negredo E, Paredes R, Ruiz L, Romeu J, Sirera G, Tural C, Burger D & Clotet B. Prospective randomized two-arm controlled study to determine the efficacy of a specific intervention to improve long-term adherence to highly active antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes* 2000; **25**:221–228.
11. Wood E, Hogg RS, Yip B, Harrigan PR, O'Shaughnessy MV & Montaner JS. Effect of medication adherence on survival of HIV-infected adults who start highly active antiretroviral therapy when the CD4+ cell count is 0.200 to 0.350 x 10(9) cells/L. *Annals of Internal Medicine* 2003; **139**:810–816.
12. Hecht FM, Grant RM, Petropoulos CJ, Dillon B, Chesney MA, Tian H & Ewing JA. Sexual transmission of an HIV-1 variant resistant to multiple reverse-transcriptase and protease inhibitors. *New England Journal of Medicine* 1998; **339**:307–311.
13. Richman DD. Antiretroviral drug resistance: mechanisms, pathogenesis, clinical significance. *Advances in Experimental Medicine & Biology* 1996; **394**:383–395.
14. Heath KV, Singer J, O'Shaughnessy MV, Montaner JS & Hogg RS. Intentional nonadherence due to adverse symptoms associated with antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes* 2002; **31**:211–217.
15. Cook RL, Sereika SM, Hunt SC, Woodward WC, Erlen JA & Conigliaro J. Problem drinking and medication adherence among persons with HIV infection. *Journal of General Internal Medicine* 2001; **16**:83–88.
16. Samet JH, Horton NJ, Meli S, Freedberg KA & Palepu A. Alcohol consumption and antiretroviral adherence among HIV-infected persons with alcohol problems. *Alcoholism, Clinical & Experimental Research* 2004; **28**:572–577.
17. Department of Health and Human Services. *Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents*. 2003. Available at http://aidsinfo.nih.gov Accessed 23 January 2004.
18. Yeni PG, Hammer SM, Carpenter CC, Cooper DA, Fischl MA, Gatell JM, Gazzard BG, Hirsch MS, Jacobsen DM, Katzenstein DA, Montaner JS, Richman DD, Saag MS, Schechter M, Schooley RT, Thompson MA, Vella S & Volberding PA. Antiretroviral treatment for adult HIV infection in 2002: updated recommendations of the International AIDS Society-USA Panel. *Journal of the American Medical Association* 2002; **288**:222–235.
19. Simoni JM, Frick PA & Pantalone DW. Strategies to enhance HAART adherence: a review of randomized controlled intervention trials. *eJournal of the International AIDS Society* 2004; **1**:WePeB5819.
20. Goujard C, Bernard N, Sohier N, Peyramond D, Lancon F, Chwalow J, Arnould B & Delfraissy JF. Impact of a patient education program on adherence to HIV medication: a randomized clinical trial. *Journal of Acquired Immune Deficiency Syndromes* 2003; **34**:191–194.
21. Pradier C, Bentz L, Spire B, Tourette-Turgis C, Morin M, Souville M, Rebillon M, Fuzibet JG, Pesce A, Dellamonica P & Moatti JP. Efficacy of an educational and counseling intervention on adherence to highly active antiretroviral therapy: French prospective controlled study. *HIV Clinical Trials* 2003; **4**:121–131.
22. Samet JH, Libman H, LaBelle C, Steger K, Lewis R, Craven DE & Freedberg KA. A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. *Archives of Internal Medicine* 1995; **155**:1629–1633.
23. Samet JH, Horton NJ, Traphagen ET, Lyon SM & Freedberg KA. Alcohol consumption and HIV disease progression: are they related? *Alcoholism, Clinical & Experimental Research* 2003; **27**:862–867.
24. Buchsbaum DG, Buchanan RG, Centor RM, Schnoll SH & Lawton MJ. Screening for alcohol abuse using CAGE scores and likelihood ratios. *Annals of Internal Medicine* 1991; **115**:774–777.
25. Ewing JA. Detecting alcoholism. The CAGE questionnaire. *Journal of the American Medical Association* 1984; **252**:1905–1907.
26. 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.
27. Radloff L. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977; **1**:385–401.
28. Ware JE. *The SF-36 Health Survey. Manual and Interpretation Guide*. 1993. Boston: The Health Institute.
29. 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. Patient Care Committee & Adherence Working Group of the Outcomes Committee of the Adult AIDS Clinical Trials Group (AACTG). *AIDS Care* 2000; **12**:255–266.
30. Averbuch M, Weintraub M & Pollock DJ. Compliance assessment in clinical trials: the MEMS devices. *Journal of Clinical Research Pharmacoepidemiology* 1990; **4**:199–204.
31. Samet JH, Rollnick S & Barnes H. Beyond CAGE. A brief clinical approach after detection of substance abuse. *Archives of Internal Medicine* 1996; **156**:2287–2293.
32. Miller WR & Rollnick S. *Motivational Interviewing: Preparing People to Change Addictive Behavior*. 1991. New York: The Guilford Press.
33. Centers for Disease Control and Prevention. Report of the NIH panel to defining principles of therapy of HIV infection and guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. *Morbidity & Mortality Weekly Report. Recommendations & Reports* 1998; **47**:48.
34. Department of Health and Human Services. *Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents*. 2000. Available at http://aidsinfo.nih.gov Accessed 27 January 2004.
35. Carpenter CC, Cooper DA, Fischl MA, Gatell JM, Gazzard BG, Hammer SM, Hirsch MS, Jacobsen DM, Katzenstein DA, Montaner JS, Richman DD, Saag MS, Schechter M, Schooley RT, Thompson MA, Vella S, Yeni PG & Volberding PA. Antiretroviral therapy in adults: updated recommendations of the International AIDS Society-USA Panel. *Journal of the American Medical Association* 2000; **283**:381–390.
36. McLellan 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.
37. NIAAA. *The Physician's Guide to Helping Patients with Alcohol Problems*. 1995. Washington, DC: National Institutes of Health.
38. Zeger SL & Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; **42**:121–130.
39. SAS/STAT Software: SAS version 8.2, 2001. Cary, NC: SAS Institute Inc.
40. Samet JH, Sullivan LM, Traphagen ET & Ickovics JR. Measuring adherence among HIV-infected persons: is MEMS consummate technology? *AIDS & Behavior* 2001; **1**:21–29.

41. Arnsten JH, Demas PA, Farzadegan H, Grant RW, Gourevitch MN, Chang CJ, Buono D, Eckholdt H, Howard AA & Schoenbaum EE. Antiretroviral therapy adherence and viral suppression in HIV-infected drug users: comparison of self-report and electronic monitoring. *Clinical Infectious Diseases* 2001; **33**:1417–1423.

42. Landis JR & Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**:159–174.

43. Becker MH. *The Health Belief Model and Personal Health Behavior.* 1974. Thorofare, NJ: Charles B. Slack.

44. Samet JH, Libman H, Steger KA, Dhawan RK, Chen J, Shevitz AH, Dewees-Dunk R, Levenson S, Kufe D & Craven DE. Compliance with zidovudine therapy in patients infected with human immunodeficiency virus, type 1: a cross-sectional study in a municipal hospital clinic. *American Journal of Medicine* 1992; **92**:495–502.

45. Durante AJ, Bova CA, Fennie KP, Danvers KA, Holness DR, Burgess JD & Williams AB. Home-based study of anti-HIV drug regimen adherence among HIV-infected women: feasibility and preliminary results. *AIDS Care* 2003; **15**:103–115.

46. Golin CE, Liu H, Hays RD, Miller LG, Beck CK, Ickovics J, Kaplan AH & Wenger NS. A prospective study of predictors of adherence to combination antiretroviral medication. *Journal of General Internal Medicine* 2002; **17**:756–765.

47. Heckman BD, Catz SL, Heckman TG, Miller JG & Kalichman SC. Adherence to antiretroviral therapy in rural persons living with HIV disease in the United States. *AIDS Care* 2004; **16**:219–230.

48. Ammassari A, Trotta MP, Murri R, Castelli F, Narciso P, Noto P, Vecchiet J, D'Arminio Monforte A, Wu AW & Antinori A; AdICoNA Study Group. Correlates and predictors of adherence to highly active antiretroviral therapy: overview of published literature. *Journal of Acquired Immune Deficiency Syndromes* 2002; **31(Suppl 3)**:S123–127.

49. van Servellen G, Carpio F, Lopez M, Garcia-Teague L, Herrera G, Monterrosa F, Gomez R & Lombardi E. Program to enhance health literacy and treatment adherence in low-income HIV-infected Latino men and women. *AIDS Patient Care & STDS* 2003; **17**:581–594.

50. Goldie SJ, Paltiel AD, Weinstein MC, Losina E, Seage GR, III, Kimmel AD, Walensky RP, Sax PE & Freedberg KA. Projecting the cost-effectiveness of adherence interventions in persons with human immunodeficiency virus infection. *American Journal of Medicine* 2003; **115**:632–641.

51. Mohammed H, Kieltyka L, Richardson-Alston G, Magnus M, Fawal H, Vermund SH, Rice J & Kissinger P. Adherence to HAART among HIV-infected persons in rural Louisiana. *AIDS Patient Care & STDS* 2004; **18**:289–296.

52. Reynolds NR, Testa MA, Marc LG, Chesney MA, Neidig JL, Smith SR, Vella S & Robbins GK; Protocol Teams of ACTG 384, ACTG 731 and A5031s. Factors influencing medication adherence beliefs and self-efficacy in persons naive to antiretroviral therapy: a multicenter, cross-sectional study. *AIDS & Behavior* 2004; **8**:141–150.

53. 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. *Clinical Infectious Diseases* 2004; **38(Suppl 5)**:S376–S387.

54. Stone VE, Jordan J, Tolson J, Miller R & Pilon T. Perspectives on adherence and simplicity for HIV-infected patients on antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes* 2004; **36**:808–816.

55. Centers for Disease Control and Prevention. 1993 Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *Morbidity & Mortality Weekly Report. Recommendations & Reports* 1992; **41**:1–19.

Received 21 July 2004, accepted 22 October 2004

ELSEVIER

Journal of Substance Abuse Treatment 29 (2005) 117–123

Journal of Substance Abuse Treatment

Relationship of depressive symptoms and mental health functioning to repeat detoxification

Christopher W. Shanahan, (M.D., M.P.H.)[a,*], Alisa Lincoln, (Ph.D., M.P.H.)[b,c], Nicholas J. Horton, (Sc.D.)[d], Richard Saitz, (M.D., M.P.H.)[a,e,f], Michael Winter, (M.P.H.)[e,g], Jeffrey H. Samet, (M.D., M.A., M.P.H.)[a,b]

[a]*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*
[b]*Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA 02118, USA*
[c]*Department of Psychiatry, Boston University School of Medicine, Boston, MA 02118, USA*
[d]*Department of Mathematics, Smith College, Northampton, MA 01060, USA*
[e]*Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA 02118, USA*
[f]*Department of Epidemiology, Boston University School of Public Health, Boston, MA 02118, USA*
[g]*Data Coordinating Center, Boston University School of Public Health, Boston, MA 02118, USA*

Received 24 March 2005; received in revised form 19 May 2005; accepted 27 May 2005

Abstract

To better understand residential detoxification use, we assessed the roles of depressive symptoms (DS) and mental health functioning (MHF) on repeat detoxification. A prospective cohort of residential detoxification patients ($N = 400$) without primary medical care was followed over 2 years at 6-month intervals. Subsequent detoxification admissions were examined using a statewide administrative database and DS (Center for Epidemiologic Studies Depression Scale) and MHF (SF-36 mental component summary subscale) measurements at follow-up. Incidence rate ratios of return to detoxification were estimated using multivariable longitudinal Poisson regression. In separate analyses, greater DS and worse MHF predicted higher detoxification use rates. Clinically significant worsening (10 points) of DS and MHF on objective scales predicted a 20% increased rate of detoxification readmission. Male sex, heroin as a problem substance, and race/ethnicity each predicted detoxification use. These data suggest that identifying individuals with DS or worse MHF after detoxification may provide opportunities for clinical intervention to reduce recurrent residential detoxification.

Keywords: Detoxification; Treatment use; Substance abuse; Depressive symptoms; Readmission

1. Introduction

Detoxification services contribute approximately $4.7 billion (39.5%) of about $11.9 billion spent annually on substance abuse treatment in the United States (Harwood, Kallinis, & Liu, 2001; Lynskey, 1998; Mark et al., 2000). Some individuals use detoxification services frequently, accounting for a disproportionately large share of overall use, and do so without appropriate aftercare or treatment. In one older study, 4% of patients accounted for nearly 25% of detoxifications, with nearly 51% of individuals receiving no continuing mental health and/or substance abuse treatment within 30 days of discharge (Richman, 1983). However, the substance abuse treatment system has made limited progress in addressing these challenges.

Compared with those who do not have any addictions, substance-dependent individuals are more likely to have comorbid psychiatric illnesses, including major depression (Curran, Flynn, Kirchner, & Booth, 2000; Depression Guideline Panel, 1993; Greenfield et al., 1998; Hesselbrock, Meyer, & Keener, 1985; Olfson et al., 2000; Powell, Penick, Othmer, Bingham, & Rice, 1982; Ross, Glaser, & German-

* Corresponding author. 91 East Concord St., Suite 200, Boston, MA 02118, USA. Tel.: +1 617 414 4562; fax: +1 617 414 4676.
E-mail address: cshanaha@bu.edu (C.W. Shanahan).

0740-5472/05/$ – see front matter
doi:10.1016/j.jsat.2005.05.005

son, 1988). The medical, psychosocial, and financial costs of substance abuse are amplified by psychiatric comorbidity (Burns & Teesson, 2002; Davis, Bush, Kivlahan, Dobie, & Bradley, 2003; Jerrell, Wieduwilt, & Macey, 2002; Kessler et al., 1996; Stein, O'Sullivan, Ellis, Perrin, & Wartenberg, 1993). Not surprisingly, comorbid mental health problems among people undergoing substance abuse treatment are associated with higher levels of impairment (Najavits, Weiss, & Shaw, 1999; Stoffelmayr, Benishek, Humphreys, Lee, & Mavis, 1989) and/or worse treatment outcomes (Amaro, 1999; Grella, 1996; Safer, 1987). Psychotic disorders, particularly severe Axis I disorders, predict treatment dropout and relapse (Brown et al., 1995; Curran, Kirchner, Worley, Rookey, & Booth, 2002). Likewise, Axis II-related disorders such as symptoms of depression, somatization, and hostility are associated with treatment dropout; both the type and severity of psychopathology predict treatment attrition (Bovasso, 2001; Haller, Miles, & Dawson, 2002; Hasin & Grant, 2002). In addition, other factors may contribute to the probability that an individual will return for detoxification. For example, among homeless individuals, exposure to postdetoxification stabilization programs predicted delay in resumption of substance use after detoxification (Kertesz, Horton, Friedmann, Saitz, & Samet, 2003). Although this body of research suggests that comorbid mental health problems may increase the likelihood of repeat use of detoxification services, this health services research question has not been explicitly addressed.

In this study, we examined depressive symptoms (DS) and mental health functioning (MHF) after residential detoxification to determine their ability to predict recurrent use of detoxification services. Potentially, patient characteristics might help identify individuals at risk for repeat detoxification. Interventions to reduce the likelihood of future repeat detoxifications could then be targeted toward high-risk patients. We hypothesized that the presence of greater DS or worse MHF assessed after initial detoxification predicts repeat residential detoxification.

2. Materials and methods

2.1. Study design and population

We conducted a secondary analysis of data obtained prospectively from the cohort of patients enrolled in the Health Evaluation and Linkage to Primary care (HELP) study who returned for interviews on one to four occasions during a 24-month follow-up period. The study, conducted in an urban short-term inpatient detoxification unit, was a randomized controlled trial of a multidisciplinary health evaluation session (the HELP intervention) designed to link substance-dependent persons without primary care physicians to outpatient primary medical care (Samet et al., 2003). The randomization allocated 205 and 195 subjects eligible for this analysis into the intervention and control groups, respectively. Eligibility criteria included the following: inpatient detoxification admission; being older than 17 years; and having alcohol, heroin, or cocaine as the substance of first or second choice. Exclusion criteria were as follows: at least one visit to a primary care provider in the preceding 2 years; pregnancy; a Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975) score lower than 21; lack of fluency in either English or Spanish; less than three contacts available to facilitate follow-up; and having specific plans to leave the Boston area within 24 months.

Trained research associates identified 642 trial-eligible subjects, of whom 470 (73%) consented to be in the HELP trial, within 24–72 hours of initial admission for detoxification. Research associates performed a 60- to 90-minute interview with all subjects prior to their discharge. Follow-up interviews were attempted with subjects every 6 months after the baseline interview for up to 24 months. Data from 995 follow-up interviews among 85% of the enrolled subjects (N = 400) were available. Interviews were conducted in English or Spanish. The Spanish instruments were translated from English, back translated by a different individual, reviewed for discrepancies, and corrected. Subjects were modestly compensated for their time in the interview. As previously reported, we used exhaustive techniques to track subjects enrolled in the HELP trial over the follow-up period (Samet et al., 2003). The institutional review board of the Boston Medical Center approved this research. All subjects provided informed consent and all the procedures followed were in accord with the standards of the institutional review board.

2.2. Administrative data

In Massachusetts, all facilities receiving any public funding must report admission and discharge data for substance abuse treatment episodes to the Massachusetts Department of Public Health, Bureau of Substance Abuse Services. Subject use data were linked and extracted using a unique identifier. Each identifier-study subject match was validated by cross-checking the name code, sex, and birth date. Patients were also asked about addiction treatment during follow-up interviews to uncover additional treatment episodes (e.g., treatment from out-of-state facilities, general medical staff, self-help groups, or facilities receiving only private funds). For this study, a subset of administrative data collected from all Massachusetts addiction treatment use was extracted and matched to all subjects for the entire study period. These data records included information about substance abuse treatment use including detoxification services; outpatient and partial-day treatment programs; short- and long-term residential treatments; and hospitalization.

2.3. Outcome variable

For this prospective cohort analysis, the outcome variable of interest was the number of detoxification admissions per

6-month period following the initial detoxification admission for up to 24 months as recorded in the state administrative data. To ensure standard identification and enumeration of discrete detoxification episodes and minimize contamination by billing irregularities, we defined a single detoxification admission for a single subject as a 3-day window of time. No repeat discharge could occur within 72 hours of previous discharge. This definition permitted consolidation of very short-term admission/discharge activities.

2.4. Explanatory variables

2.4.1. Depressive symptoms and mental health functioning

The predictor variables examined were DS and MHF collected at each follow-up interview. We used the Center for Epidemiologic Studies Depression Scale (CES-D), a measure of DS. We also used the mental component summary (MCS) subscale of the SF-36 to characterize MHF. Higher CES-D scores (range, 0–60) signify increased levels of DS and, conversely, higher MCS scores (range, 0–100) indicate better MHF. It is accepted that the manifestations of DS overlap substantially with the acute withdrawal syndrome. Furthermore, it is widely acknowledged that most individuals experience withdrawal symptoms during acute detoxification. For this reason, our analysis did not include baseline DS and MHF information collected at study entry during the initial detoxification admission.

2.4.2. Covariates

To adjust for suspected confounders, we selected covariates for inclusion in the multivariable analyses based on evidence from the literature and clinical experience. These covariates included age, race/ethnicity (White [reference group], African American, Hispanic, and Other), homelessness (report of a night spent on the street or in a shelter in the past 6 months; Kertesz et al., 2003), completion of high school education, employment (full time, part time, unemployed, other), any health insurance, current smoking, and problem substance (alcohol, heroin, cocaine) as well as a dichotomous indicator of randomization assignment relating to the original trial. We defined a substance as a "problem substance" if subjects reported either 5 or more days of its use in the last 30 days or 1 or more years of its use three or more times per week in their lifetime (DeAlba, Samet, & Saitz, 2004).

2.5. Analyses

Descriptive statistics were first used to characterize the sample. For the analysis, we modeled the outcome variable (detoxification admissions) as a count of each subject's individual detoxification admissions over each 6-month follow-up period to determine a rate during each period. Multivariable longitudinal Poisson regression (log link) models were constructed to assess the incidence rate ratio (IRR) of admissions for detoxification. Separate models were fit for DS (Model A, containing CES-D) and MHF (Model B, containing MCS) as the main independent predictor of interest because CES-D and MCS are related but not identical measures. Updated information on CES-D and MCS from each follow-up interview was included in the model as a time-varying covariate. Generalized estimating equations were used to account for the clustering within subjects over time in the Poisson models (Zeger & Liang, 1986). An independent working correlation model was assumed, along with an empirical variance estimator. The multivariable longitudinal Poisson regression models controlled for time (6, 12, 18, or 24 months) and other covariate predictors listed previously to provide 6-month IRRs. We first constructed parsimonious models focusing on either CES-D or MCS versus outcome (number of subsequent detoxifications/past 6 months) while adjusting for core predictors (time [6, 12, 18, or 24 months], age, sex, race, randomization group, alcohol to intoxication as problem substance, heroin as problem substance, cocaine as problem substance). We tested an interaction term between sex and DS (CES-D) and between sex and MHF (MCS) in corresponding adjusted models to detect and characterize possible differential sex effects. The two final fully adjusted models included additional predictors (homelessness, high school graduation, employment status, health insurance status, and smoking) to more thoroughly account for potential confounding. The descriptive analysis and Poisson regression modeling were performed using SAS/STAT software (Version 8.2; Cary, NC). Sensitivity analysis using zero-inflated Poisson regression models was conducted with Stata Software (Version 8.2; College Station, TX).

Table 1
Sociodemographic characteristics of drug- and alcohol-dependent patients without primary medical care in a detoxification unit (N = 400)

Covariate	M	SD	Range
CES-D	33.2	12.4	1–60
MCS	31.4	12.7	6.8–62.2
Age (years)	36.0	7.7	18–60
Covariate	**Attribute**	**N**	**%**
Sex	Male	303	76
Race/Ethnicity	Black	201	50
	White	135	34
	Hispanic	40	10
	Other	24	6
High school graduate		278	70
Employment	Full time	163	41
	Part time	73	18
	Unemployed	152	38
	Other	12	3
Homeless		187	47
Any health insurance		157	39
Current smoking		345	86
Problem substance	Alcohol to intoxication	340	85
	Cocaine	306	77
	Heroin	153	38

3. Results

3.1. Cohort characterization: descriptive statistics

The demographic characteristics of the cohort are described in Table 1. The subjects had an average age of 36 years, 76% were male, 50% were Black, 34% were White, 10% were Hispanic, 70% completed high school, 47% reported homelessness, 39% had health insurance, and 86% currently smoked tobacco (Table 1). Of the 400 study subjects, 394 had 990 separate observations for Model A (CES-D) and 395 had 992 separate observations for Model B (MCS). Criteria for inclusion into either model were identical. The number of subjects and that of observations differ because of missing CES-D or MCS data. Over the entire 2-year observation period, 169 subjects (42%) were found to have one or more detoxification admissions.

3.2. Multivariable longitudinal poisson regression models (Model A, CES-D; Model B, MCS)

In separate models, more severe DS and worse MHF were significantly associated with a higher rate of detoxification use following the initial detoxification admission (Table 2). The IRR for a 10-point increase in the CES-D was 1.2 (95% confidence interval [CI], 1.0–1.4) and that for a 10-point decrease in the MCS was likewise 1.2 (95% CI, 1.0–1.3); both are 20% increases in detoxification rate. Clinically significant worsening (10 points) of DS and MHF on objective scales predicted a 20% increased rate of detoxification readmission. The likelihood of repeat detoxification was 50% greater for men and subjects with heroin as problem substance whereas it was 60% less for Hispanic subjects. Comparison of the parsimonious and fully adjusted versions of both Model A and Model B revealed minimal to no difference in the calculated IRRs. Finally, the likelihood of repeat detoxification was more than 50% greater for homeless individuals in both models (A and B). The results from only the fully adjusted models are presented in Table 2. No significant association was found between the number of detoxifications in either Model A or Model B and any of the additional suspected confounding variables added in the fully adjusted models, nor were significant interactions detected.

Additional regression models for counts (zero-inflated Poisson, negative binomial, and zero-inflated negative binomial) were fit to assess the sensitivity of the overdispersed Poisson model. The zero inflation models allowed the proportion of zeroes to vary by the CES-D score and the MCS score, respectively. All models yielded consistent inferences regarding the association of CES-D or MCS with the number of detoxification admissions. The results of these analyses are not reported in this article.

Table 2
Multivariable models of the association of repeat detoxification with DS (Model A), MHF (Model B), and other factors

	IRR (95% CI)	
Covariate	Model A	Model B
CES-D[a,b] (10-point) increase	**1.18 (1.02–1.36)**	NA
MCS[c] (10-point) decrease	NA	**1.16 (1.03–1.31)**
Age (decade) increase	0.9 (0.7–1.1)	0.9 (0.7–1.1)
Female sex	**0.5 (0.3–0.7)**	**0.5 (0.3–0.7)**
Race/Ethnicity		
White	1.0	1.0
African American	0.7 (0.5–1.1)	0.7 (0.5–1.0)
Hispanic	**0.4 (0.2–0.7)***	**0.4 (0.2–0.7)****
Other	**0.3 (0.2–0.5)***	**0.3 (0.2–0.6)****
Group[d]	0.8 (0.6–1.3)	0.8 (0.6–1.2)
High school graduate	1.0 (0.7–1.4)	1.0 (0.7–1.4)
Employed	1.18 (0.99–1.41)	1.17 (0.99–1.4)
Homeless	**1.51 (1.00–2.26)**	**1.53 (1.02–2.28)**
Any health insurance	1.1 (0.7–1.7)	1.1 (0.7–1.7)
Current smoking	1.2 (0.8–1.8)	1.2 (0.8–1.8)
Problem substance		
Alcohol to intoxication	1.0 (0.6–1.6)	1.0 (0.6–1.5)
Cocaine	0.8 (0.5–1.1)	0.7 (0.5–1.1)
Heroin	**1.5 (1.1–2.2)**	**1.5 (1.0–2.1)**

Note. NA indicates not applicable.

[a] Bolded values indicate that 95% CIs do not include 1.0.

[b] Response range: 0–60, where higher scores indicate more symptoms.

[c] Response range: 0–100, where higher scores indicate better function.

[d] Refers to random subject assignment to the control group (usual care) and the intervention group in a controlled trial of a multidisciplinary health evaluation session linking substance-dependent persons without primary care physicians to outpatient primary medical care.

* $p = .008$ for overall association between race and detoxification.

** $p = .001$ for overall association between race and detoxification.

4. Discussion

DS and MHF each, in separate models, impacted detoxification admission rates over a 2-year period. These effects were substantial, consistent, and sustained, thus indicating that greater detoxification use can be found among people with a greater degree of DS or a lower level of MHF status. This finding was demonstrated similarly in a population-based study in which DS predicted greater overall medical care use despite adjustment for several measures of medical severity (Rowan, Davidson, Campbell, Dobrez, & MacLean, 2002). However, the mechanism for this phenomenon is not entirely clear. In one study, elevated risk of substance abuse relapse after treatment discharge was related to patients diagnosed with depression (Curran et al., 2000). It is plausible that this group of patients might be more likely to seek help and consequently have increased detoxification use. However, this seems unlikely based on evidence that depressed patients are three times more likely than nondepressed patients to be noncompliant with medical treatment recommendations and perhaps, by extension, health-seeking behaviors in general (DiMatteo, Lepper, & Croghan, 2000). Alternatively, barriers to treatment (e.g., perceived separation between mental health and general health, stigma, cost, lack of time or confidentiality, and fear of documentation or unwanted intervention; Givens

& Tjia, 2002; Van Hook, 1999) might explain delayed but eventually greater overall use of detoxification services. Further studies will be necessary to elucidate this mechanism.

Prior studies demonstrate that DS and MHF are strong but somewhat erratic predictors of several important substance abuse treatment outcomes. For example, DS increase the likelihood of relapse in abstinent individuals during the transitional period between short- and long-term treatment (Curran, Flynn, et al., 2000; Curran, Kirchner, et al., 2002). Furthermore, in substance abusers, depression decreases the functional health status and subjective quality of life (Rudolf & Priebe, 2002) and significantly increases risk for suicide (Cornelius et al., 1995; Dhossche, Meloukheia, & Chakravorty, 2000). However, in contrast, depression was favorably associated with treatment outcomes (longer duration of abstinence and greater decreases in symptomatology) in a study of individuals interviewed within 3 months of starting outpatient substance abuse treatment (Charney, Paraherakis, Negrete, & Gill, 1998). Our study tested the hypothesis that DS increase the likelihood of return to detoxification after an initial detoxification. Direct associations were detected between repeat detoxification and sex, race/ethnicity, homelessness, and heroin as problem substance (Table 2); however, no significant interaction was detected either between sex and DS or between sex and MHF.

The IRR for repeat detoxification admissions among Hispanic subjects was 60% less than that found in Whites. Although it is possible that race may represent a proxy for unmeasured sociodemographic characteristics, we included several potential confounders to minimize the likelihood that the adjusted models would misattribute the effects of any variable. No significant relationship was detected between these additional potential confounders to detoxification use in the fully adjusted models. This ethnic disparity for detoxification readmission remains unexplained by our investigation. However, other investigators have noted differences in treatment use across ethnic lines (Arroyo, Westerberg, & Tonigan, 1998; Brecht, von Mayrhauser, & Anglin, 2000; Cherpitel, 2001).

Subjects reporting heroin as a problem substance were 50% more likely than those not reporting this to be admitted for detoxification within a 6-month period. Current research has still not provided a consistent understanding of the influence of specific problem substances on substance abuse treatment use. For example, while opioid-dependent persons have higher rates of general health care use (Darke, Ross, Teesson, & Lynskey, 2003), paradoxically, long-term cocaine or heroin users are more likely than nonusers to not want health care treatment and to put off seeking needed health care (McCoy, Metsch, Chitwood, & Miles, 2001). Although this latter study demonstrates a negative relationship between problem substance and treatment use, it is equally plausible that delayed preventive care and treatment results in more urgent presentation for more intensive treatment.

In persons presenting for detoxification, the likelihood of detoxification readmission being enhanced by increased DS or impaired MHF could be explained by several plausible mechanisms. First, a lowered threshold for care seeking could lead to an increased frequency of treatment seeking. Second, an increased likelihood for inappropriate or illicit self-medication of DS may hasten detoxification readmission. Finally, repeat detoxification could result from a negative impact on the protective behaviors that lessen the risk of substance abuse behaviors. Examples include both low-level substance use heightening the risk of relapse and the need for treatment as well as decreased substance abuse treatment adherence leading to a diminished likelihood for therapeutic success after detoxification. Further studies on the type and intensity of services rendered, the acuity of dependence, withdrawal, and comorbid psychiatric burden at presentation as well as detailed longitudinal evaluations of treatment may make it possible to distinguish between these possible mechanisms.

Up until the present, inconsistent effects have been detected in the relationship between depression and substance abuse treatment use. However, these analyses have had limitations. Most studies use subject self-report outcomes. In addition, most studies define the dependent variable as a single time event, "time to first drink" (Greenfield et al., 1998). However, without actual counts of detoxification admissions, accurate rates of treatment use are not calculable. This study uses a multivariable Poisson regression analysis, which, as a more comprehensive modeling strategy, permits more accurate and efficient use of a state's substance abuse treatment use data. This modeling method not only validates the existence of a contemporaneous relationship between psychiatric health status and substance abuse treatment use but also provides an estimate of the relative contributions of relevant predictors. Moreover, the magnitude and consistency of the relationship as demonstrated by this analytic approach using two separate but related measures (CES-D and MCS) suggest that both DS and MHF have clinically meaningful effects on detoxification use.

These data also support efforts to identify patients with a history of recent detoxification who are coincidently burdened with excessive comorbid DS or particularly low MHF and provide a rationale for testing interventions aimed at optimizing substance abuse treatment in individuals at risk for future repeat detoxifications (Novacek & Raskin, 1998). Several instruments have been developed to screen for depression in general medical settings (Williams, Noël, Cordes, Ramirez, & Pignone, 2002). In addition to the CES-D, such instruments include the Beck Depression Inventory, Primary Care Evaluation of Mental Disorders (PRIME-MD), PRIME-MD Patient Health Questionnaire, and the Zung Self-Rating Depression Scale. Although treatment of comorbid DS among substance-dependent persons appears to have some benefit (Cornelius, Bukstein, et al., 2004; Cornelius, Maisto, et al., 2004; Nunes & Levin, 2004),

it is not known whether this salutary effect would impact the sizeable subset of individuals who continue to cycle through detoxification programs. These data provide a rationale for future investigations to evaluate specific treatment interventions aimed at individuals exhibiting DS to determine whether a reduction in rates of repeat detoxification use can be achieved.

Although not measured by the present study, this research calls to mind interventions that would reduce many of the negative consequences of depression (e.g., lost productivity, suicide, and overall increased health care use), which in turn could provide considerable additional benefit to this population (Dhossche et al., 2000; Frasure-Smith et al., 2000; Goldman, Nielsen, & Champion, 1999; Jackson, Manning, & Wells, 1995; Kessler et al., 1996; Pearson et al., 1999; Preuss et al., 2002; Rowan et al., 2002; Simon, VonKorff, & Barlow, 1995).

4.1. Limitations

This investigation has some limitations. Despite adjusting for several known potential confounders and using extensive data collected prospectively, it remains possible that there are additional unaccounted confounders that may explain our findings.

It is also possible that all detoxification admissions might not have been captured in the administrative database. However, it is unlikely that the study population would seek detoxification from programs not receiving public funding. It is reassuring however that 100% of the study subjects were found in the administrative database based on their index detoxification admission.

Finally, there are limits to the mental health measures that we used. For example, CES-D DS do not correlate well with a diagnosis of depression. In addition, the DS and MHF measures used in this investigation have not been validated explicitly on a population with alcohol or drug dependence.

4.2. Conclusions and implications

Greater DS or worse MHF is associated with future repeat detoxification use. These findings suggest that assessing patients for increased levels of DS and worse MHF after discharge from detoxification can identify individuals at a higher risk for repeat detoxification. Mental health interventions targeting these individuals might decrease subsequent detoxification service use and costs and thus represent an appropriate focus of future studies.

Acknowledgments

This work was supported by the National Institute of Drug Abuse (DA10019) and the National Institute of Alcohol Abuse and Alcoholism (AA10870). Subject interviews took place at the General Clinical Research Center, Boston University School of Medicine (RR00533). We gratefully acknowledge the assistance of Mary Jo Larson, PhD, of the New England Research Institutes (Watertown, MA) in designing the study and Emily Shapiro of Smith College (Northampton, MA) in preparing the analyses.

References

Amaro, H. (1999). An expensive policy: The impact of inadequate funding for substance abuse treatment. *American Journal of Public Health, 89*, 657–659.

Arroyo, J. A., Westerberg, V. S., & Tonigan, J. S. (1998). Comparison of treatment utilization and outcome for Hispanics and non-Hispanic Whites. *Journal of Studies on Alcohol, 59*, 286–291.

Bovasso, G. (2001). The long-term treatment outcomes of depression and anxiety comorbid with substance abuse. *Journal of Behavioral Health Services & Research, 28*, 42–57.

Brecht, M. L., von Mayrhauser, C., & Anglin, M. D. (2000). Predictors of relapse after treatment for methamphetamine use. *Journal of Psychoactive Drugs, 32*, 211–220.

Brown, S. A., Inaba, R. K., Gillin, J. C., Schuckit, M. A., Stewart, M. A., & Irwin, M. R. (1995). Alcoholism and affective disorder: Clinical course of depressive symptoms. *American Journal of Psychiatry, 152*, 45–52.

Burns, L., & Teesson, M. (2002). Alcohol use disorders comorbid with anxiety, depression and drug use disorders. Findings from the Australian National Survey of Mental Health and Well Being. *Drug and Alcohol Dependence, 68*, 299–307.

Charney, D. A., Paraherakis, A. M., Negrete, J. C., & Gill, K. J. (1998). The impact of depression on the outcome of addictions treatment. *Journal of Substance Abuse Treatment, 15*, 123–130.

Cherpitel, C. J. (2001). Differences in services utilization between White and Mexican American DUI arrestees. *Alcoholism, Clinical and Experimental Research, 25*, 122–127.

Cornelius, J. R., Bukstein, O. G., Salloum, I. M., Kelly, T. M., Wood, D. S., & Clark, D. B. (2004a). Fluoxetine in depressed AUD adolescents: A 1-year follow-up evaluation. *Journal of Child and Adolescent Psychopharmacology, 14*, 33–38.

Cornelius, J. R., Maisto, S. A., Martin, C. S., Bukstein, O. G., Salloum, I. M., Daley, D. C., et al. (2004b). Major depression associated with earlier alcohol relapse in treated teens with AUD. *Addictive Behaviors, 29*, 1035–1038.

Cornelius, J. R., Salloum, I. M., Mezzich, J., Cornelius, M. D., Jr., Fabrega, H., Ehler, J. G., et al. (1995). Disproportionate suicidality in patients with comorbid major depression and alcoholism. *American Journal of Psychiatry, 152*, 358–364.

Curran, G. M., Flynn, H. A., Kirchner, J., & Booth, B. M. (2000). Depression after alcohol treatment as a risk factor for relapse among male veterans. *Journal of Substance Abuse Treatment, 19*, 259–265.

Curran, G. M., Kirchner, J. E., Worley, M., Rookey, C., & Booth, B. M. (2002). Depressive symptomatology and early attrition from intensive outpatient substance use treatment. *Journal of Behavioral Health Services & Research, 29*, 138–143.

Darke, S., Ross, J., Teesson, M., & Lynskey, M. (2003). Health service utilization and benzodiazepine use among heroin users: Findings from the Australian Treatment Outcome Study (ATOS). *Addiction, 98*, 1129–1135.

Davis, T. A., Bush, K. R., Kivlahan, D. R., Dobie, D. J., & Bradley, K. A. (2003). Screening for substance abuse and psychiatric disorders among women patients in a VA health care system. *Psychiatric Services, 54*, 214–218.

DeAlba, I., Samet, J. H., & Saitz, R. (2004). Burden of medical illness in drug- and alcohol-dependent persons without primary care. *American Journal on Addictions, 13*, 33–45.

Depression Guideline Panel. (1993). Depression in primary care: I. Detection and diagnosis: Clinical Practice Guideline No. 5. Rockville, MD: US Dept of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

Dhossche, D. M., Meloukheia, A. M., & Chakravorty, S. (2000). The association of suicide attempts and comorbid depression and substance abuse in psychiatric consultation patients. *General Hospital Psychiatry, 22*, 281–288.

DiMatteo, M. R., Lepper, H. S., & Croghan, T. W. (2000). Depression is a risk factor for noncompliance with medical treatment: Meta-analysis of the effects of anxiety and depression on patient adherence. *Archives of Internal Medicine, 160*, 2101–2107.

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.

Frasure-Smith, N., Lesperance, F., Gravel, G., Masson, A., Juneau, M., Talajic, M., et al. (2000). Depression and health-care costs during the first year following myocardial infarction. *Journal of Psychosomatic Research, 48*, 471–478.

Givens, J. L., & Tjia, J. (2002). Depressed medical students' use of mental health services and barriers to use. *Academic Medicine, 77*, 918–921.

Goldman, L. S., Nielsen, N. H., & Champion, H. C. (1999). Awareness, diagnosis, and treatment of depression. *Journal of General Internal Medicine, 14*, 569–580.

Greenfield, S. F., Weiss, R. D., Muenz, L. R., Vagge, L. M., Kelly, J. F., Bello, L. R., et al. (1998). The effect of depression on return to drinking. *Archives of General Psychiatry, 55*, 259–265.

Grella, C. E. (1996). Background and overview of mental health and substance abuse treatment systems: Meeting the needs of women who are pregnant or parenting. *Journal of Psychoactive Drugs, 28*, 319–343.

Haller, D. L., Miles, D. R., & Dawson, K. S. (2002). Psychopathology influences treatment retention among drug-dependent women. *Journal of Substance Abuse Treatment, 23*, 431–436.

Harwood, H. J., Kallinis, S., & Liu, C. (2001). The cost and components of substance abuse treatment. Center for Substance Abuse Treatment, Caliber/NEDS: Department of Health and Human Services.

Hasin, D. S., & Grant, B. F. (2002). Major depression in 6050 former drinkers: Association with past alcohol dependence. *Archives of General Psychiatry, 59*, 794–800.

Hesselbrock, M. N., Meyer, R. E., & Keener, J. J. (1985). Psychopathology in hospitalized alcoholics. *Archives of General Psychiatry, 42*, 1050–1055.

Jackson, C. A., Manning, W. G., Jr., & Wells, K. B. (1995). Impact of prior and current alcohol use on use of services by patients with depression and chronic medical illnesses. *Health Services Research, 30*, 687–705.

Jerrell, J. M., Wieduwilt, K. M., & Macey, D. V. (2002). Use and costs of public-sector behavioral health services for African-American and White women. *Psychiatric Services, 53*, 195–200.

Kertesz, S. G., Horton, N. J., Friedmann, P. D., Saitz, R., & Samet, J. H. (2003). Slowing the revolving door: Stabilization programs reduce homeless persons' substance use after detoxification. *Journal of Substance Abuse Treatment, 24*, 197–207.

Kessler, R. C., Nelson, C. B., McGonagle, K. A., Edlund, M. J., Frank, R. G., & Leaf, P. J. (1996). The epidemiology of co-occurring addictive and mental disorders: Implications for prevalence and service utilization. *American Journal of Orthopsychiatry, 66*, 17–31.

Lynskey, M. T. (1998). The comorbidity of alcohol dependence and affective disorders: Treatment implications. *Drug and Alcohol Dependence, 52*, 201–209.

Mark, T. L., Coffey, R. M., King, E., Harwood, H., McKusick, D., Genuardi, J., et al. (2000). Spending on mental health and substance abuse treatment, 1987–1997. *Health Affairs, 19*, 108–120.

McCoy, C. B., Metsch, L. R., Chitwood, D. D., & Miles, C. (2001). Drug use and barriers to use of health care services. *Substance Use & Misuse, 36*, 789–806.

Najavits, L. M., Weiss, R. D., & Shaw, S. R. (1999). A clinical profile of women with PTSD and substance dependence. *Psychology of Addictive Behaviors, 13*, 98–104.

Novacek, J., & Raskin, R. (1998). Recognition of warning signs: A consideration for cost–effective treatment of severe mental illness. *Psychiatric Services, 49*, 376–378.

Nunes, E. V., & Levin, F. R. (2004). Treatment of depression in patients with alcohol or other drug dependence: A meta-analysis [Review]. *JAMA, 291*, 1887–1896.

Olfson, M., Shea, S., Feder, A., Fuentes, M., Nomura, Y., Gameroff, M., et al. (2000). Prevalence of anxiety, depression, and substance use disorders in an urban general medicine practice. *Archives of Family Medicine, 9*, 876–883.

Pearson, S. D., Katzelnick, D. J., Simon, G. E., Manning, W. G., Helstad, C. P., & Henk, H. J. (1999). Depression among high utilizers of medical care. *Journal of General Internal Medicine, 14*, 461–468.

Powell, B. J., Penick, E. C., Othmer, E., Bingham, S. F., & Rice, A. S. (1982). Prevalence of additional psychiatric syndromes among male alcoholics. *Journal of Clinical Psychiatry, 43*, 404–407.

Preuss, U. W., Schuckit, M. A., Smith, T. L., Danko, G. R., Dasher, A. C., Hesselbrock, M. N., et al. (2002). A comparison of alcohol-induced and independent depression in alcoholics with histories of suicide attempts. *Journal of Studies on Alcohol, 63*, 498–502.

Richman, A. (1983). Cost/effectiveness analysis of alcoholism and drug abuse treatment programs: The relevance of recidivism and resource absorption. *Evaluation & Program Planning, 6*, 49–52.

Ross, H. E., Glaser, F. B., & Germanson, T. (1988). The prevalence of psychiatric disorders in patients with alcohol and other drug problems. *Archives of General Psychiatry, 45*, 1023–1031.

Rowan, P. J., Davidson, K., Campbell, J. A., Dobrez, D. G., & MacLean, D. R. (2002). Depressive symptoms predict medical care utilization in a population-based sample. *Psychological Medicine, 32*, 903–908.

Rudolf, H., & Priebe, S. (2002). Subjective quality of life and depressive symptoms in women with alcoholism during detoxification treatment. *Drug and Alcohol Dependence, 66*, 71–76.

Safer, D. J. (1987). Substance abuse by young adult chronic patients. *Hospital & Community Psychiatry, 38*, 511–514.

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 multidisciplinary health evaluation in a detoxification unit (The Health Evaluation and Linkage to Primary care [HELP]). *Addiction, 98*, 509–516.

Simon, G. E., VonKorff, M., & Barlow, W. (1995). Health care costs of primary care patients with recognized depression. *Archives of General Psychiatry, 52*, 850–856.

Stein, M. D., O'Sullivan, P. S., Ellis, P., Perrin, H., & Wartenberg, A. (1993). Utilization of medical services by drug abusers in detoxification. *Journal of Substance Abuse Treatment, 5*, 187–193.

Stoffelmayr, B. E., Benishek, L. A., Humphreys, K., Lee, J. A., & Mavis, B. E. (1989). Substance abuse prognosis with an additional psychiatric diagnosis: Understanding the relationship. *Journal of Psychoactive Drugs, 21*, 145–152.

Van Hook, M. P. (1999). Women's help-seeking patterns for depression. *Social Work in Health Care, 29*, 15–34.

Williams, J. W., Jr., Noël, P. H., Cordes, J. A., Ramirez, G., & Pignone, M. (2002). Is this patient clinically depressed? *JAMA, 287*, 1160–1170.

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics, 42*, 121–130.

Do Variations in Disease Prevalence Limit the Usefulness of Population-Based Hospitalization Rates for Studying Variations in Hospital Admissions?

Michael Shwartz, PhD, Erol A. Peköz, PhD,* Arlene S. Ash, PhD,† Michael A. Posner, MS,‡ Joseph D. Restuccia, DrPH,* and Lisa I. Iezzoni, MD§*

Background: Studies of geographic variation in hospitalizations commonly examine age- and gender-adjusted population-based hospitalization rates (ie, the numbers of persons hospitalized relative to what is expected given the age/gender distributions in the area population).

Objective: To determine whether areas identified as extreme using population-based hospitalization rates remain extreme when ranked by disease-based hospitalization rates (the numbers of persons hospitalized relative to what is expected given the amount of disease in the area).

Design: The authors examined 1997 Medicare data on both inpatient admissions and outpatient visits of patients 65 years and older in each of 71 small areas in Massachusetts for 15 medical conditions. For each area, the number of people having each condition was calculated as the sum of those hospitalized plus those treated as outpatients only. The authors used hierarchical Bayesian modeling to estimate area-specific population-based hospitalization rates, disease-based hospitalization rates (DHRs), and disease prevalence.

Main Outcome Measure: The extent to which the same areas were identified as extreme based on population-based hospitalization rates versus DHRs.

Results: Area-specific population-based hospitalization rates, DHRs, and disease prevalence varied substantially. Areas identified as extreme using population-based hospitalization rates often were not extreme when ranked by DHRs. For 11 of the 15 conditions, 5 or more of the 14 areas ranked in top and bottom deciles by population-based hospitalization rates were more likely than not (ie, with probability ≥ 0.50) to be at least 2 deciles less extreme when ranked by DHRs.

Conclusion: Differences in disease prevalence can limit the usefulness of population-based hospitalization rates for studying variations in hospital admissions.

Key Words: small-area variations, hospital utilization, disease prevalence

(*Med Care* 2005;43: 4–11)

Many studies have reported large differences in age- and gender-adjusted rates of hospitalization across small geographic areas.[1–13] Studies of surgery and procedures[14–19] and medical conditions[20] have found that higher rates of inappropriateness do not explain higher hospitalization rates. Despite this, some suggest that savings are possible without sacrificing quality of care by reducing rates in high-rate areas to levels in lower rate areas.[21,22] This suggestion assumes that much of the variation is due to "practice style" differences, either "uncertainty" about best treatment practices[23] or "enthusiasm" for one type of treatment over another,[24] which only minimally affect health care outcomes.

Area-specific hospitalization rates are population based because denominators used in calculating the rates reflect the populations in the areas rather than the numbers of people within areas who have the disease. Differences in age- and gender-adjusted population-based hospitalization rates might be due to differences in the likelihood that patients with the disease are admitted to the hospital. Alternatively, they may reflect differences in disease prevalence.

The most widely used approach to account for differences in disease burden across areas is to adjust area-specific rates further for differences in hospitalization rates for selected marker conditions chosen as proxies for the underlying burden of illness in the population (eg, hospitalizations for hip fracture, colon or lung cancer treated surgically, acute myocardial infarction, and stroke).[25] This approach hypothesizes that hospitalization rates for the marker conditions are highly correlated with the total amount of other diseases in an

From the *School of Management, Boston University; the †Department of General Internal Medicine, Boston University School of Medicine; the ‡Boston University School of Public Health; and the §Harvard Medical School, Boston, Massachusetts.

Supported in part by grant no. RO3 HS09832-01 from the Agency for Healthcare Research and Quality.

Reprints: Michael Shwartz, School of Management, Boston University, 595 Commonwealth Avenue, Boston, MA 02215. E-mail: mshwartz@bu.edu.

ISSN: 0025-7079/05/4301-0004

area. The hypothesis has some merit, as suggested by the finding that variation in such marker conditions explains about 25% of the variation in age-, gender-, and race-adjusted Medicare spending across regions.[22] However, use of marker conditions as a proxy for the prevalence of other diseases has not been validated. Also, when examining variations in hospitalization rates for specific conditions, the conceptual justification for using markers as a measure of prevalence is weaker.

Over the last decade, several groups have developed risk assessment models that use both inpatient and ambulatory claims to identify diseases and predict costs.[26] Their success indicates that although diagnosis codes on claims forms are imperfect indicators of true disease prevalence, they contain useful information about the medical problems that are present. Especially in a Medicare population, in which most beneficiaries are insured for both inpatient and outpatient care, differences across areas in amount of disease as determined from claims forms should reflect, reasonably well, differences in underlying disease prevalence.

In this study we used Medicare data from both inpatient and outpatient claims to identify people with any of 15 medical conditions. For each condition and geographic area, we calculated both the population-based hospitalization rate (the number of people hospitalized relative to the number expected given the age/gender distribution of the population in the area) and the disease-based hospitalization rate (the number of people hospitalized relative to the number expected given the amount of disease in the area). We then examined the extent to which the same areas were ranked as either particularly high or low using the 2 measures.

METHODS

Database

We studied hospitalizations and outpatient treatment of Medicare patients older than 65 years of age in Massachusetts in 1997. Hospitalization data were obtained from the Center for Medicare and Medicaid Services (CMS, formerly the Health Care Financing Administration) MedPAR file. Outpatient data were obtained from the CMS 1997 Carrier File (claims data for part B physician/supplier services) and Outpatient File (claims data for outpatient facility charges at hospitals and other institutions).

Conditions

Table 1 shows the 15 medical conditions in our study, defined initially by diagnosis-related group (DRG). As described elsewhere,[27] within most DRGs we increased clinical homogeneity by considering only discharges with a principal diagnosis from selected ICD-9-CM codes. All 15 conditions had at least 1000 people hospitalized and 6000 treated as outpatients only in 1997.

Determining Numerators and Denominators

Disease prevalence cannot be inferred from claims for Medicare beneficiaries in health maintenance organizations (HMOs; because their outpatient claims are not submitted to CMS) or those not eligible for outpatient and physician office (part B) reimbursement. Thus, we wanted to include in our analyses only non-HMO part B-eligible enrollees. We did not have individual-level eligibility information. As a proxy for part B eligibility of hospitalized patients, we only counted hospitalized patients with at least 1 part B bill in 1997

TABLE 1. Number of People Hospitalized and Number Treated on an Outpatient-Only Basis, by Condition

DRG	Hospitalized	Outpatient-Only
15 Transient ischemic attack	2001	11,762
88 Chronic bronchitis and emphysema	4842	51,982
89 Bacterial pneumonia	8666	7899
127 Heart failure	11,359	37,187
130 Peripheral vascular disorders	1334	40,936
132 Ischemic heart disease	3432	111,838
138 Cardiac arrhythmia and conduction disorder	3856	74,109
140 Angina pectoris	1596	25,597
141 Syncope and collapse	2046	6002
143 Chest pain	2859	26,282
243 Medical back problems	1453	59,754
277 Cellulitis and abscess	1765	16,339
294 Diabetes	1009	119,713
296 Fluid and electrolyte disorder	4351	6451
320 Kidney and urinary tract infections	3116	24,579

(slightly over 90% of hospitalized Medicare patients). We also eliminated from hospitalization counts the approximately 5% of discharges in the MedPAR file in which an HMO was the payer.

We identified outpatient visits, including office, nursing and rest home, and home visits from CPT codes.[27] Diagnosis coding for outpatient visits is generally less reliable than inpatient coding and is governed by somewhat different rules. To reflect the potential effect of different coding rules, we used 2 methods to assign outpatient visits to conditions: one based on a "series" of outpatient visits and a second based on a single visit. For inpatients, coding guidelines instruct abstractors to code a diagnosis accompanied by such phrases as "rule out," "suspect," or "question" as if the disease had actually occurred. For outpatients, coding guidelines stipulate that only confirmed diagnoses be coded to their highest level of specificity. As a result, a series of outpatient visits may carry multiple codes as the physician attempts to confirm a diagnosis. Because we wished to capture only "final" diagnoses, we proceeded as follows: We considered any outpatient visit within 6 weeks of a previous outpatient visit to be part of the same series of visits. Only diagnostic codes on the last visit in a series were used to identify the conditions being addressed. We required a gap of at least 8 weeks (ie, an additional 2-week buffer) to establish the beginning of a new series.

We examined the sensitivity of conclusions to an approach that ignored visit series and identified the conditions being addressed based on the presence of any relevant diagnosis at any single visit. Because findings concerning the concordance between population-based and disease-based rates were similar for both approaches, we only report analyses using the "visit series" method.

We considered the numbers of people with each condition, either treated in the hospital or as outpatient only, rather than the numbers of admissions or outpatient visits, primarily because the number of people who have a diagnosis is a more direct measure of underlying disease prevalence than service counts, which also reflect practice style.[28] Most variation in overall hospitalization rates is caused by variations in the number of people hospitalized.[27] By focusing on numbers of people rather than numbers of events, we can estimate the observed amount of disease in an area as the total of people hospitalized plus those treated as outpatients only.

For each 5-year age category from 65 years and older, and for each gender, we determined the number of Medicare enrollees in each zip code in Massachusetts from the Annual Zip Code Enrollment File.

Creating Small Geographic Areas

As described elsewhere,[27,29] we used Ward's clustering algorithm to create small geographic areas. Ward's clustering algorithm[30,31] creates areas by combining zip codes based on similarity in the proportion of total hospital discharges from the zip code that were from each hospital. Discharges of patients with the following characteristics were used in clustering: age 65 years or older, Massachusetts resident, and discharged in 1997 from a hospital in Massachusetts paid under Medicare's Prospective Payment System. The results of the clustering were 71 small areas with the following distribution of residents 65 years and older: 20 areas had less than 5000; 21 had 5000 to 9999; 19 had 10,000 to 19,999; and 11 areas had more than 20,000.

Analysis

We considered 3 types of area-specific "relative rates" (referred to more simply as "rates"), defined as observed counts divided by expected.

- Population-based hospitalization rate = number of people hospitalized relative to the number expected.
- Disease prevalence = sum of people hospitalized plus those treated as outpatients only (ie, the number with the disease) relative to the number expected (we use the term prevalence rather than rate to emphasize what this rate is measuring).
- Disease-based hospitalization rate = proportion of people with the disease that were hospitalized relative to the ratio of the expected number hospitalized to the expected number with the disease.

Let O_{ij} = number of people hospitalized (ie, treated as inpatients) in area j

O_{oj} = number of people treated as outpatients only in area j

E_{ij} = expected number of people hospitalized in area j

E_{oj} = expected number of people treated as outpatients only in area j

Both E_{ij} and E_{oj} are adjusted for age and gender distribution in the area using indirect standardization.[32] Observed relative rates are calculated as follows:

Observed population-based hospitalization rate = O_{ij}/E_{ij}

Observed disease rate = $(O_{ij} + O_{oj})/(E_{ij} + E_{oj})$

Observed disease-based hospitalization rate =

$$\frac{O_{ij}/(O_{ij} + O_{oj})}{E_{ij}/(E_{ij} + E_{oj})} = \frac{O_{ij}/E_{ij}}{(O_{ij} + O_{oj})/(E_{ij} + E_{oj})}$$

The rates are centered at 1 because over all areas the observed number of events is equal to the number expected. Because population-based hospitalization rate equals disease-based hospitalization rate multiplied by disease prevalence, population-based hospitalization rates are similar to disease-based hospitalization rates only when disease prevalence varies little across the area from what is expected based on age and gender alone.

We ranked areas from lowest to highest according to their population-based hospitalization rates. Areas in the first decile (ranks 1–7) had the lowest rates and those in the last

decile (ranks 65–71) had the highest. We call areas in these deciles extreme. For each extreme area, we examined its ranking according to its disease-based hospitalization rate. We present detailed results for heart failure (DRG 127), which has over 10,000 people treated as inpatients and nearly 40,000 treated as outpatients only.

Using the observed rates as calculated in the previous equations as estimates of "true" underlying rates does not explicitly take into account random variation of "true" rates across areas.[8] To estimate "true" rates more accurately, we used a hierarchical Bayesian model[33–38]—specifically, a Poisson model with area treated as a random effect. We validated the model by showing that, across the 71 areas, inpatient and outpatient counts generated from the model corresponded well to the observed counts. (Details of the model and validation are available from the first author.)

We estimated "true" rates in each area using Gibbs sampling as implemented in WINBUGS 1.4.[39] These estimates can be thought of as weighted averages of the observed rates in each area (as calculated from the previous formulas) and the average rate over all areas in the state (which is 1, because our rates are relative rates). Thus, the weighting "shrinks" each observed rate toward 1. We refer to these estimates as "shrunken" rates. The areas with the most shrinkage are those with the most extreme observed rates and those with the fewest people. Shrunken estimates are more accurate than traditional estimates in predicting small-area hospitalization rates.[29]

Based on the posterior means from Gibbs sampling, we ranked each area using both population-based and disease-based rates. Ranks, even when based on shrunken estimates, are inherently unstable.[40] To describe the extent to which areas were ranked differently when using population-based versus disease-based rates, we estimated how often (in repeated Gibbs samples from the appropriate posterior distribution) the rank produced by the disease-based rate differed from the rank produced by the population-based rate by at least 2 deciles (14 or more ranks). In summarizing results across conditions, we report the number of the 14 extreme areas in which it was "more likely than not" (ie, probability ≥ 0.50) that the disease-based rank differed by more than 2 deciles from the population-based rank.

RESULTS

Table 1 shows the number of people hospitalized and the number treated on an outpatient-only basis for each condition.

We use heart failure (DRG 127) to illustrate the analyses. Table 2 shows the effect of shrinkage on both the population-based and the disease-based hospitalization rates for those areas in the lowest and highest decile based on their observed rates. For example, although the area with the smallest observed population-based hospitalization rate had an observed (relative) rate of 0.43, its shrunken rate was 0.62. The area ranked 65th according to its observed population-based rate was particularly small. Its shrunken rate was pulled a lot toward 1 (from 1.30–1.08), such that after shrinkage it was ranked only 49th. Although shrunken estimates were less spread out than the observed rates, shrinkage rarely caused large changes in rank. In fact, among the 28 observed-versus-shrunken rank comparisons in Table 2, 22 changed by 3 ranks or less and only 2 comparisons (both relating to the very small area with 909 residents) changed ranks by 14 or more.

TABLE 2. Heart Failure (DRG 127) Example: Effect of Shrinkage on Rates in Areas With the Most Extreme Observed Population-Based and Disease-Based Hospitalization Rates

Population-Based Rate (Rank)			Disease-Based Rate (Rank)		
Population	**Observed**	**Shrunk**	**Population**	**Observed**	**Shrunk**
5183	0.43 (1)	0.62 (1)	5183	0.55 (1)	0.78 (3)
3973	0.50 (2)	0.66 (4)	3384	0.67 (2)	0.83 (5)
1291	0.53 (3)	0.66 (3)	11,474	0.67 (3)	0.74 (1)
3776	0.62 (4)	0.71 (7)	3973	0.72 (4)	0.89 (10)
7943	0.62 (5)	0.65 (2)	33,791	0.76 (5)	0.78 (2)
15,271	0.63 (6)	0.67 (5)	6693	0.79 (6)	0.85 (6)
3383	0.66 (7)	0.71 (6)	30,476	0.81 (7)	0.83 (4)
909	1.30 (65)	1.08 (49)	11,904	1.21 (65)	1.17 (66)
11,904	1.32 (66)	1.26 (66)	909	1.22 (66)	1.05 (42)
23,750	1.32 (67)	1.30 (68)	2178	1.22 (67)	1.09 (57)
2178	1.33 (68)	1.16 (58)	4978	1.23 (68)	1.14 (64)
7751	1.41 (69)	1.32 (70)	7828	1.24 (69)	1.17 (69)
3527	1.44 (70)	1.31 (69)	7751	1.25 (70)	1.19 (70)
2899	1.65 (71)	1.44 (71)	8261	1.36 (71)	1.26 (71)

We were primarily interested in the extent to which areas identified as extreme (ie, in the top and bottom deciles) according to their population-based hospitalization rate were also extreme according to their disease-based hospitalization rate. Table 3 shows for heart failure the shrunken population-based hospitalization rate and rank, and the shrunken disease-based hospitalization rate and rank, for the most extreme 14 areas according to their shrunken population-based rates. The area with the lowest population-based hospitalization rate (62% of expected, rank 1) also had a low disease-based hospitalization rate (78% of expected, rank 3). We call the area with the second lowest population-based hospitalization rate (rank 2) "area A" (we refer to it again later). Although area A's population-based hospitalization rate was 65% of expected, based on its disease-based hospitalization rate (105% of expected), it was ranked 41st. Of the 7 areas ranked in the first decile (ranks 1–7) based on their population-based hospitalization rate, 4 were ranked in the 3rd decile or higher (rank 22 or higher) based on their disease-based hospitaliza-

TABLE 3. Heart Failure (DRG 127) Example: Shrunken Rates and Ranks* of 3 Measures For Areas With the Most Extreme Population-Based Hospitalization Rates

Population-Based Hospitalization		Disease-Based Hospitalization		Disease	
Rate	**Rank**	**Rate**	**Rank**	**Prevalence**	**Rank**
0.62	1	0.78	3	0.79	11
0.65	2	1.05	41	0.62	1
0.66	3	0.95	22	0.69	4
0.66	4	0.89	10	0.74	7
0.67	5	0.92	13	0.73	6
0.71	6	1.03	38	0.68	3
0.71	7	0.96	23	0.74	8
1.25	65	1.01	33	1.24	66
1.26	66	1.17	66	1.08	53
1.29	67	0.90	11	1.44	71
1.30	68	1.09	58	1.19	63
1.31	69	1.08	52	1.22	65
1.32	70	1.19	70	1.12	59
1.44	71	1.07	50	1.35	69

*Ranks are based on the mean of the posterior distribution of the relative rates estimated by the Gibbs sampler.

tion rate. Of the 7 areas in the 10th decile (ranks 65–71) based on their population-based hospitalization rate, 3 were in the 7th decile or lower (ranks 50 or lower) based on their disease-based hospitalization rate. Notably, when using the disease-based hospitalization rate, 2 areas in the highest population-based hospitalization rate decile (ranks 65 and 67) were ranked lower (ranks 33 and 11 respectively) than 2 of the areas in the lowest population-based hospitalization rate decile (population-based hospitalization rate ranks 2 and 6, which were ranked 41st and 38th based on disease-based hospitalization rate).

The right 2 columns in Table 3, which show estimates of disease prevalence, demonstrate why population-based hospitalization rates and disease-based hospitalization rates can be so different. Area A, discussed earlier, had 62% of the amount of disease expected (the lowest estimated disease prevalence). Its low population-based hospitalization rate was largely a reflection of the low disease prevalence in the area. After taking disease prevalence into account (by using the disease-based hospitalization rate), the area actually had 5% more hospitalizations than expected. All 7 areas with the lowest population-based hospitalization rates had very low levels of disease and, to a large extent, this explains their very low population-based hospitalization rates. The highest decile demonstrates the same phenomenon. The area ranked 67th in terms of its population-based hospitalization rate had 44% more disease than expected. When this was taken into account, the proportion hospitalized in the area was 90% of expected rather than the 29% more than expected indicated by the population-based hospitalization rate.

Figure 1 is a box plot of shrunken population-based and disease-based hospitalization rates for the 15 conditions. Many of the conditions exhibited more variability in rates than heart failure. For most of the conditions, population-based rates varied from around 50% of expected to 50% more than expected. Disease-based rates also varied considerably, although in most conditions somewhat less than population-based rates.

Table 4 shows for each of the 15 conditions for areas in the top decile of population-based hospitalization rates, what their rank would be according to their disease-based hospitalization rate. To illustrate, for transient ischemic attacks (TIAs, top row) the area ranked highest (71st) based on its population-based hospitalization rate was ranked 62nd based on its disease-based hospitalization rate; the area ranked second highest based on its population-based hospitalization rate (70th) was ranked 49th based on its disease-based hospitalization rate. The numbers in parentheses indicate the fraction of the time in repeated Gibbs samples from the posterior distributions of the ranks that the 2 ranks differed by at least 2 deciles (shown if the fraction was ≥ 0.50). For example, for TIA for the highest ranked area based on its population-based hospitalization rate, there was a 0.50 chance that the disease-based hospitalization rate rank differed from the population-based hospitalization rate rank by at least 2 deciles. For the second highest ranked area based on its

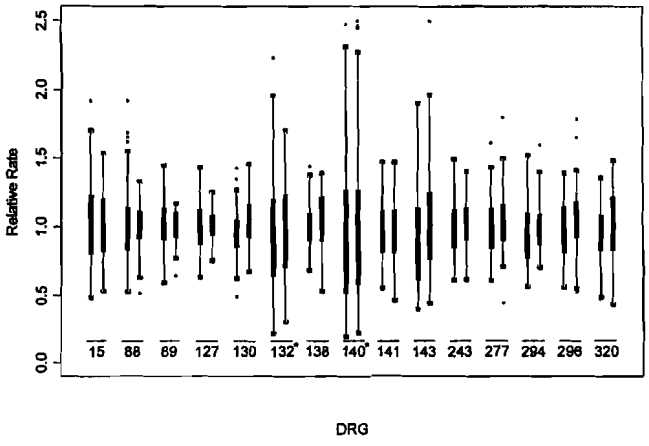


FIGURE 1. Box plot of shrunken population-based hospitalization rates (left) and disease-based hospitalization rate (right) by condition. To retain the graph's scale, we excluded extreme points, as follows:

DRG 132, right plot: 3.2

DRG 140, left plot: 2.7, 2.8, 3.1, 3.2, 3.4, 3.9, 5.4; right plot: 2.8, 4.0, 4.3

TABLE 4. For Areas With the Highest Ranked Shrunken Population-Based Hospitalization Rates, Rank of the Shrunken Disease-Based Hospitalization Rate and Probability That the 2 Ranks Differ by at Least 14*

	Rank of Population-Based Hospitalization Rate						
	71	70	69	68	67	66	65
Condition	**Rank of Disease-Based Hospitalization Rate (Probability*)**						
Transient ischemic attack	62 (0.50)	49 (0.80)	66	65	45 (0.74)	59	63
Chronic bronchitis and emphysema	64	68	52 (0.72)	69	34 (0.93)	71	58 (0.53)
Pneumonia	51 (0.61)	50 (0.70)	36 (0.99)	67	65	71	55
Heart failure	50 (0.74)	70	52 (0.66)	58 (0.65)	11 (0.99)	66	33 (0.93)
Peripheral vascular disorders	68	71	4 (0.99)	23 (0.98)	16 (0.99)	57	24 (0.92)
Ischemic heart disease	68	70	67	71	69	60	63
Arrhythmia	70	15 (0.99)	48 (0.62)	47 (0.70)	51 (0.52)	67	44 (0.84)
Angina	71	65	68	64	70	66	69
Syncope	67	71	66	70	55	69	65
Chest pain	55 (0.63)	70	71	69	68	66	50 (0.93)
Back problems	61 (0.52)	71	68	63	39 (0.84)	34 (0.94)	35 (0.83)
Cellulitis	67	69	64	60	31 (0.98)	65	19 (0.99)
Diabetes	69	68	62	13 (0.92)	39 (0.79)	58	52
Fluid and electrolyte disorder	5 (0.99)	59	63	70	65	68	66
Kidney and urinary tract infections	71	56	64	54	40 (0.95)	68	46 (0.61)

*Ranks are based on the means of the posterior distribution of the rates estimated by the Gibbs sampler. Probabilities are the fraction of the times that the 2 ranks, as sampled from the posterior distributions, differed by 14 or more.

population-based hospitalization rate, the probability was 0.80.

For 4 conditions (heart failure, peripheral vascular disorders, arrhythmia, and back problems), 10 or 11 of the 14 extreme areas according to their population-based hospitalization rate were "more likely than not" (ie, with a probability of 0.50 or greater) to differ by at least 2 deciles when ranked according to their disease-based hospitalization rate. For 7 conditions (TIA, chronic obstructive pulmonary disease, pneumonia, chest pain, cellulitis, diabetes, and electrolyte disorder), 5 or 6 of the 14 extreme areas based on the population-based hospitalization rate were "more likely than not" to differ by at least 2 deciles when ranked by their disease-based hospitalization rate. Only for ischemic heart disease, angina, and syncope was it more likely than not that most of the areas (12 or more) were within 2 deciles when ranked by each rate.

DISCUSSION

Population-based hospitalization rates varied widely across small areas in Massachusetts. However, disease prevalence also varied widely. Because of differences in disease prevalence, areas with extreme population-based hospitalization rates were not necessarily areas in which a particularly high or low proportion of those with the disease were admitted to the hospital, relative to what was expected given the age and gender distribution in the area.

Our approach relies on claims data to identify people with specific diseases. Because of concern about the validity of disease identification from outpatient codes, we conducted analyses using both a stringent definition (the series method) and a lenient definition (which considered all visits) for identifying disease from diagnostic codes on outpatient claims. Both analyses showed substantial differences in area ranks when calculated with population-based versus disease-based rates. However, claims data can be imperfect in many ways and are unlikely to capture all conditions noted on the medical record.[41] Our conclusions do not depend on accurately estimating actual disease prevalence, but on estimating relative disease prevalence (ie, how much more or less disease one area has compared with another). The key assumption behind our analysis is that outpatient coding is not systematically biased across areas. However, areas with better access to technology or more specialists may identify more disease than other areas with a similar disease burden, or they may code presentations to justify use of the technology or referral to specialists. We are currently exploring the potential for such bias by examining the relationship, within areas, between specialist physician supply and claims-based disease prevalence.

Another concern when identifying disease from claims is that higher rates of coded disease may reflect better access to primary care or may be a proxy for either physician or hospital bed supply. In a Medicare population eligible for

both hospital and outpatient care, however, financial access differences are at least somewhat muted. Also, by focusing on the number of people with any visit for a condition rather than total numbers of medical encounters by those with the condition, we reduce the effect of practice style on our measure of disease prevalence. If supply were the main factor driving demand, the same areas that have high rates of disease would have high rates of hospitalization among those with the disease. Correlations between disease prevalence and the proportion of those with the disease who were hospitalized were, in fact, statistically significant for 9 of the 15 conditions. However, in all 9, the correlation was negative. Finally, despite concerns about outpatient coding, CMS judges them sufficiently valid that, starting in 2004, it will accept diagnoses from either inpatient or outpatient claims to calculate health-based payments to Medicare+Choice HMOs for the Medicare beneficiaries they enroll.

As health care costs continue to outpace general inflation, pressure is mounting to revitalize certificate-of-need programs. For example, the *Wall Street Journal* recently wrote: "the Big Three [auto companies] have lobbied aggressively to keep certificate-of-need programs in states such as Missouri and have fought ardently for the establishment of programs in Ohio and Indiana."[42] Vermont recently issued a request for proposal for consultation services to develop a health resource allocation plan to guide health facility planning and capital expenditures. Hospital capacity is a major focus of such programs. Motivating decisions about hospital capacity are area-specific hospitalization rates. Our analyses suggest that disease prevalence rather than population counts provide a more appropriate denominator for such rates. Unfortunately, outpatient claims are not widely available. At a minimum, our study indicates that databases with outpatient as well as inpatient claims are needed to address better the research and policy questions raised by geographic variations.

Many current quality improvement efforts examine process measures for patients with specific conditions (eg, whether specifically identified patients with heart disease receive appropriate medications or diabetic patients receive appropriate assessment and preventive therapy). Whether patients with specific conditions are hospitalized is an important process measure, both because of iatrogenic events and costs. Although we have focused on variations across geographic areas, our approach applies as well to examining variations in hospitalizations across physician practices. Because of small samples, shrinkage estimators are even more important in this setting.

Our study examined only 15 medical conditions using 1 year of data from 1 state. Also, some Medicare beneficiaries use both Veteran Administration (VA) and non-VA facilities and we have not included diagnostic or utilization data from VA facilities.[43,44] Nonetheless, the large differences in disease prevalence across small areas that remain after adjusting for age and gender raise concerns about the value of population-based hospitalization rates for studying hospital utilization and drawing inferences about physician practice styles.

ACKNOWLEDGMENT

The manuscript has benefited greatly from suggestions from the anonymous reviewers and the editor.

REFERENCES

1. Barnes BA, O'Brien EZ, Comstock C, et al. Report on variation in rates of utilization of surgical services in the Commonwealth of Massachusetts. *JAMA*. 1985;254:371–375.
2. Carlisle DM, Valdez RB, Shapiro MF, et al. Geographic variation in rates of selected surgical procedures within Los Angeles County. *Health Serv Res*. 1995;30:27–42.
3. Connell FA, Day RW, LoGerfo JP. Hospitalization of Medicaid children: analysis of small area variations in admission rates. *Am J Public Health*. 1981;71:606–613.
4. Folland S, Stano M. Small area variations: a critical review of propositions, methods and evidence. *Med Care Rev*. 1990;47:419–465.
5. Gittelsohn AM, Powe NR. Small area variations in health care delivery in Maryland. *Health Serv Res*. 1995;30:295–317.
6. Lewis CE. Variations in the incidence of surgery. *N Engl J Med*. 1969;281:880–884.
7. McMahon LF, Wolfe RA, Tedeschi PJ. Variation in hospital admissions among small areas: a comparison of Maine and Michigan. *Med Care*. 1989;27:623–631.
8. McPherson S, Wennberg JE, Hovind OB, et al. Small-area variations in the use of common surgical procedures: an international comparison of New England, England and Norway. *N Engl J Med*. 1982;307:1310–1314.
9. Paul–Shaheen P, Clark JD, Williams D. Small area analysis: a review and analysis of the North American literature. *J Health Polit Policy Law*. 1987;12:741–809.
10. Roos NP, Roos LL Jr. Surgical rate variations: do they reflect the health or socioeconomic characteristics of the population. *Med Care*. 1982;20:945–958.
11. Roos NP. Hysterectomy: variations in rates across small areas and across physicians' practices. *Am J Public Health*. 1984;74:327–335.
12. Wennberg JE, Gittelsohn A. Health care delivery in Maine I: patterns of use of common surgical procedures. *J Maine Med Assoc*. 1975;66:123–130.
13. Wennberg JE, Gittelsohn A. Variations in medical care among small areas. *Sci Am*. 1982;246:120–133.
14. Berstein SJ, Hilborne LH, Leape LL, et al. The appropriateness of use of coronary angiography in New York State. *JAMA*. 1993;269:766–769.
15. Chassin MR, Kosecoff J, Park RE, et al. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *JAMA*. 1987;258:2533–2537.
16. Hilborne LH, Leape LL, Bernstein SJ, et al. The appropriateness in use of percutaneous transluminal coronary angioplasty in New York state. *JAMA*. 1993;269:761–765.
17. Leape LL, Park RE, Solomon DH, et al. Does inappropriate use explain small-area variations in the use of health care services? *JAMA*. 1990;263:669–672.
18. Leape LL, Hilborne LH, Park RE, et al. The appropriateness of use of coronary artery bypass graft surgery in New York state. *JAMA*. 1993;269:753–760.
19. Restuccia JD, Shwartz M, Kreger BE, et al. Does more "appropriateness" explain higher rates of cardiac procedures among patients hospitalized with coronary heart disease? *Med Care*. 2002;40:500–509.
20. Restuccia JD, Shwartz M, Ash A, et al. High hospital admission rates and inappropriate care. *Health Aff*. 1996;15:156–163.
21. Fisher ES, Welch HG, Wennberg JE. Prioritizing Oregon's hospital resources: an example based on variations in discretionary medical utilization. *JAMA*. 1992;267:1925–1931.

22. Wennberg JE, Fisher ES, Skinner JS. Geography and the debate over Medicare reform. *Health Aff* 13 February 2002. Available at www.healthaffairs.org/WebExclusives/Wennberg_Web_Excl_021302.htm.
23. Wennberg JE, Barnes BA, Zubkoff M. Professional uncertainty and the problem of supplier-induced demand. *Soc Sci Med.* 1982;16:811–824.
24. Chassin MR. Explaining geographic variations: the enthusiasm hypothesis. *Med Care* 1993;31:YS37–YS44.
25. Fisher ES, Wennberg JE, Stukel TA, et al. Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors. *Health Serv Res.* 2000;34:1351–1362.
26. Cumming RB, Cameron BA. *A Comparative Analysis of Claims-Based Methods of Health Risk Assessment for Commercial Populations.* Available at www.soa.org/sections/riskadjfinalreport1.pdf.
27. Shwartz M, Iezzoni LI, Ash AS, et al. *More Disease: How Major a Factor in Higher Utilization?* Final report submitted to the Agency for Healthcare Research and Quality, August 2001 (Grant No. R03 HS09832-01).
28. Wennberg JE. Population illness rates do not explain population hospitalization rates: a comment on Mark Blumberg's thesis that morbidity adjusters are needed to interpret small area variations. *Med Care* 1987; 25:339.
29. Shwartz M, Ash AS, Anderson J, et al. Small area variations in hospitalization rates: how much you see depends on how you look. *Med Care.* 1994;32:189–201.
30. Ward JH Jr. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc.* 1963;58:236–244.
31. Romesburg CH. *Cluster Analysis for Researchers.* London: Wadsworth; 1984.
32. Rothman KJ. *Modern Epidemiology.* Boston: Little, Brown; 1986.
33. Gelman A, Carlin JP, Stern HS, et al. *Bayesian Data Analysis.* 2nd edition. Boca Raton, FL: Chapman & Hall/CRC; 2003.
34. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis.* 2nd ed. Boca Raton, FL: Chapman & Hall/CRC; 2000.
35. Bryk AS, Raudenbush SW. *Hierarchical Linear Models: Applications and Data Analysis Methods. Advanced Quantitative Techniques in the Social Sciences.* Newbury Park: Sage Publications; 1992.
36. Gatsonis CA, Epstein AM, Newhouse JP, et al. Variations in the utilization of coronary angiography for elderly patients with an acute myocardial infarction: an analysis using hierarchical logistic regression. *Med Care.* 1995;33:625–642.
37. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med.* 1997;127:764–768.
38. Normand S-LT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. *J Am Stat Assoc.* 1997;92:803–814.
39. The software and documentation are available at www.mrc-bsu.cam.ac.uk/bugs/.
40. Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance [with discussion]. *J R Stat Soc A.* 1996;159:385–443.
41. Fowles JB, Lawthers AG, Weiner JP, et al. Agreement between physicians' office records and Medicare part B claims data. *Health Care Financ Rev.* 1995;16:189–199.
42. Hawkins L Jr. Detroit auto makers fight local hospitals' building plans. *The Wall Street Journal.* January 20, 2004. Available at: http://online.wsj.com.
43. Fleming C, Fisher ES, Chang CH, et al. Studying outcomes and hospital utilization in the elderly: the advantages of a merged data base for Medicare and Veterans Affairs hospitals. *Med Care.* 1992;30:377–391.
44. Ashton CM, Souchek J, Petersen NJ, et al. Hospital use and survival among Veterans Affairs beneficiaries. *N Engl J Med.* 2003;349:1637–1646.

ELSEVIER

Journal of Substance Abuse Treatment 28 (2005) 181–187

Journal of Substance Abuse Treatment

Regular article

Knowledge of and interest in hepatitis C treatment at a methadone clinic

Alexander Y. Walley, (M.D.)[a,*], Mary C. White, (R.N.), (M.P.H), (Ph.D)[b], Margot B. Kushel, (M.D.)[c], Yong S. Song, (Ph.D.)[d], and Jacqueline P. Tulsky, (M.D)[e]

[a]*Adult Medicine, Dimock Community Health Center, Boston, MA 02119, USA*
[b]*Community Health Systems, School of Nursing, University of California, San Francisco, San Francisco, CA 94110, USA*
[c]*Division of General Internal Medicine, San Francisco General Hospital, University of California, San Francisco, San Francisco, CA 94110, USA*
[d]*Department of Psychiatry, San Francisco General Hospital, University of California, San Francisco, San Francisco, CA 94110, USA*
[e]*Positive Health Program, San Francisco General Hospital, University of California, San Francisco, San Francisco, CA 94110, USA*

Received 18 February 2004; received in revised form 4 November 2004; accepted 1 December 2004

Abstract

This study describes knowledge about hepatitis C virus (HCV) infection and interest in treatment among 110 opiate dependent patients from an opiate dependence treatment program in San Francisco. None had been treated for HCV and only 30% had been evaluated for HCV treatment. While only 34% knew about HCV treatment, 54% of the sample became "definitely interested" in HCV treatment after hearing the risks and benefits. Men were approximately five times more likely than women to know of some HCV treatment. Whites were seven times and Latinos were about six times more likely than African-Americans to know about HCV treatment. Our findings suggest that methadone programs can play an important role in increasing access to HCV treatment through educating patients about treatment options.

Keywords: Hepatitis C; Methadone; Injection drug users; Knowledge of treatment; Interest in treatment

1. Introduction

Hepatitis C virus (HCV) infection is the most common chronic blood-borne infection in the United States. Injection drug users (IDU) account for 60% of new HCV infections (Centers for Disease Control and Prevention, 1998). Advancements in HCV treatment have proved effective at decreasing viral activity, producing sustained viral response rates of 40–80% depending on viral genotype (National Institutes of Health, 2002). However, the 6- to 12-month treatment course requires regular monitoring and management for significant side effects including cytopenias and depression (Ho et al., 2001; National Institutes of Health, 2002). Because of these monitoring requirements there have been concerns about treating IDUs for HCV. Would HCV-positive IDUs accept HIV treatment? Would they adhere to the treatment regimen? Could the side effects be monitored? Would HCV treatment exacerbate psychiatric illness? (Davis & Rodrigue, 2001; Johnson, Fisher, Fenaughty, & Theno, 1998).

Perhaps because of these concerns, early consensus guidelines on HCV treatment recommended delaying treatment for 6 months after patients who drank alcohol or used illicit drugs stopped (Centers for Disease Control and Prevention, 1998; National Institutes of Health, 1997). There have been ethical challenges to these guidelines recommending the withholding of HCV treatment from IDUs (Edlin et al., 2001), and two published case series have demonstrated that IDUs in drug treatment can be successfully

* Corresponding author. Adult Medicine, Dimock Community Health Center, 55 Dimock Street, Boston, MA 02119, USA. Tel.: +1 415 307 1532.

E-mail address: matiaswalley@mindspring.com (A.Y. Walley).

0740-5472/05/$ – see front matter
doi:10.1016/j.jsat.2004.12.004

treated for HCV (Backmund, Meyer, Von Zielonka, & Eichenlaub, 2001; Sylvestre, 2002). Thus the 2002 NIH Consensus Statement on HCV infection calls for increased availability of treatment to patients such as IDUs who were previously considered ineligible (National Institutes of Health, 2002).

About 20% of the 810,000 heroin users in the United States receive treatment for opiate dependence at methadone programs, and HCV prevalence among methadone clients is reported to be 66–93% (Best et al., 1999; Dhopesh, Taylor, & Burke, 2000; Murrill et al., 2002; Piccolo et al., 2002). Key components of drug treatment include education and counseling about transmission of blood borne pathogens, including hepatitis C (Astone, Strauss, Vassilev, & Des Jarlais, 2003; Pratt, Paone, Carter, & Layton, 2002; Sorensen, & Copeland, 2000; Strauss, Astone, Vassilev, Des Jarlais, & Hagan, 2003). Some methadone clinics have linked drug treatment, primary care and HIV treatment services (Friedmann, Alexander, & D'Aunno, 1999; Strauss et al., 2003; Umbricht-Schneiter, Ginn, Pabst, & Bigelow, 1994). Understanding what methadone patients know about hepatitis C treatment and learning whether they have been evaluated or treated are important first steps in determining the role for methadone clinics in hepatitis C screening, education, evaluation, and treatment. Therefore, in this study, we sought to determine patients' knowledge of and interest in HCV treatment among patients at a university-affiliated county hospital-based methadone clinic.

2. Methods

2.1. Design, setting, and sample

We conducted a cross sectional survey of clients receiving methadone at a university-affiliated, public hospital-based, opiate addiction treatment program between June and December 2002. This clinic offers approximately 460 treatment slots: 100 for methadone detoxification treatment (MDT) and 360 for methadone maintenance treatment (MMT). Serving the indigent population of the county, the clinic does not require insurance or exclude patients unable to pay. The clinic admits new MDT patients on a daily walk-in basis. A nurse practitioner examines admitted MDT patients and offers optional screening for syphilis and hepatitis A, B, and C. The hospital counseling and testing service screens patients for HIV. The MMT program recruits its patients from the MDT program. Demand for MMT is higher than available spots, so the MMT program selects patients based on medical severity; patients with a higher burden of co-existing illness, such as HIV or mental illness, receive priority for MMT. An addiction counselor meets regularly with each MDT and MMT patient. The clinic provides mental health treatment and HIV primary care for MMT patients who do not receive these services elsewhere. MDT lasts 21 to 90 days with a tapered methadone dose.

During data collection times, clinic staff informed each patient coming into clinic of the study and directed interested patients to research staff. Research staff conducted interviews on different days and during different times of day, including weekends to include patients with different schedules. All participants provided signed informed consent and the Committee on Human Research at the University of California, San Francisco, approved the study design.

2.2. Data collection

One of seven trained interviewers conducted a 20-min face-to-face interview with each participant. Interviewer training included several practice interviews with study authors and approximately five supervised interviews with study participants.

All data for the study were by self-report. Study staff did not review medical records, due to resource constraints. Demographic questions included age, gender, race/ethnicity, education, and employment. Health information included treatment status (MMT or MDT), HIV status ("ever had a test" and "what you were told were your results"), and HCV status (ever had a positive test or told by a health care professional that you were infected with hepatitis C). Three questions surveyed knowledge of HCV disease ("of those infected, how many get sick"; "how long does it take to become sick"; and "can someone die from hepatitis C"). We asked questions about knowledge of transmission ("do people get infected from sharing needles"; "... from sharing cookers, cottons or rinse water") and treatment ("do you know of any treatments or medications for hepatitis C infection").

We asked if subjects had ever been evaluated for HCV treatment ("have you been evaluated for hepatitis C treatment"; "have you ever had an appointment with a liver or GI specialist for hepatitis C infection"; "have you ever had a liver biopsy for hepatitis C infection") and if they were undergoing or had completed treatment for HCV. Following these questions, we provided information about HCV infection and treatment, and then we asked about interest in treatment (Box). Subjects answered using a five-point Likert scale from "definitely not interested" to "definitely interested". Upon completion of the interview the participant received $5 and an information pamphlet about HCV infection.

2.3. Analysis

We describe the sample in terms of demographic and health characteristics by treatment status (MDT vs. MMT). We made statistical comparisons between these groups and the overall clinic population (using Student's t, Chi Square and Fisher's Exact test) to determine how similar the sample was to the clinic population. Information on gender, race/ethnicity, education, and HIV status was available for

Box. Interest in HCV treatment question.

Hepatitis C is a viral infection of the liver. There are medications to treat hepatitis C infection, called interferon and ribavirin. I'm going to tell you a few things about the virus and the medications and then ask you how interested you think you might be in getting treatment.

- People with chronic hepatitis C who don't take any medications have about a 1 in 5 chance of getting cirrhosis (severe liver disease) at some point in their life.
- Usually a liver biopsy is required before starting treatment. A liver biopsy is a common test to check for damage caused by hepatitis C. It is done by taking out some cells from the liver using a needle placed in your right side just below the rib cage.
- The medications usually include shots and pills. The shots usually need to be taken once a week for at least 6 months and often a full year. The pills need to be taken two times every day.
- The medications make about half of the people who take them feel achy, like they have the flu—irritable, tired, or depressed, throughout the time they take the medicine.
- People who finish the medications have about a 50% chance of getting rid of the hepatitis C virus completely.
- If you do get rid of the virus, you can probably get it again if you share syringes, cookers, cottons, or rinse water with anyone.

Knowing these things AND assuming you been infected with the hepatitis C virus, how interested would you say you are in taking medications for the hepatitis C infection during the next year?

MMT clients enrolled at the clinic for approximately the same time period, July 2002 to March 2003, as the study. Information on the MDT clinic population was not available for this time period, and so the comparison was limited to MMT clients.

We collapsed certain variables for analysis. Gender, which included transgender categories, was collapsed into "birth sex," meaning sex at birth. HIV and HCV status were dichotomized into positive vs. all other responses (negative, don't know, declined). We defined evaluation for HCV treatment as those who answered that they had been evaluated. Those who reported that they had never been evaluated for HCV treatment ($n = 8$), but who were positive for HCV and reported either a GI appointment or a liver biopsy, were also coded as having been evaluated for HCV treatment. Interest in treatment was collapsed into "definitely interested" vs. all other responses.

We conducted separate analyses on the following outcomes: (1) knowledge of treatment; (2) having been evaluated for treatment; (3) receipt or completion of treatment; and (4) interest in treatment under an assumption of being infected with HCV. First, we described the proportion that was positive for each outcome. We then examined sociodemographic and methadone treatment status with each outcome in univariate analysis to determine variables that were significant for inclusion in a multivariable logistic regression. For each categorical variable, a reference was selected for comparisons. For example, African Americans were selected as the reference group for ethnicity, and each other group was compared to African Americans. All analyses were done using SPSS software, with alpha = 0.10

Table 1
Demographic and health-related characteristics of the study sample by methadone treatment category (methadone detoxification [MDT], methadone maintenance [MMT]), and characteristics of the methadone maintenance clinic population

Characteristic	Total study sample ($n = 110$)	MDT study sample ($n = 25$)	MMT study sample ($n = 85$)	MMT clinic ($n = 469$)
Age				
Mean	45	42	46	na
Median (range)	47 (19–62)	44 (19–57)	47 (21–62)	
Gender (% male)	66	68	66	64
Race/Ethnicity (%)				
African American	43	32	46	35
White	35	44	33	50
Hispanic/Latino	14	16	13	10
Other/Mixed/ Declined	8	8	8	6
Education (%)				
Less than high school	31	28	32	40
High School	50	56	48	38
Some college and higher	19	19	20	22
Employed (%)	9	20	6	na
HIV positive (%)	26	0	33	39
HCV positive (%)	81	64	86	na
Know about HCV treatment (%)	34	44	31	na
Ever evaluated for HCV treatment (%)	30	8	36	na
Undergoing or completed treatment for HCV (%)	0	0	0	na
Interested in HCV treatment (%)				
Definitely	54	80	46	na
Probably	22	12	25	
Not sure	16	0	21	
Probably not	5	4	6	
Definitely not	3	4	2	

na = not available.
HCV = hepatitis C virus.
HIV = human immunodeficiency virus.

Table 2
Univariate and multivariable logistic regression analyses for factors associated with knowledge of treatment for Hepatitis C

Variable	Univariate analysis OR (95% CI)	Multivariable logistic regression* OR (95% CI)	*p* value for logistic regression
Gender			
Male	3.8 (1.4–10.3)	5.0 (1.6–15.7)	.006
Female	Reference	Reference	
Hepatitis C status			
Positive	3.7 (1.0–13.5)	3.9 (0.9–16.6)	.069
Negative or unknown	Reference	Reference	
Race/Ethnicity			
White	9.0 (2.8–23.1)	7.4 (2.3–23.5)	.001
Hispanic/Latino	4.6 (1.2–17.4)	5.8 (1.4–24.8)	.017
Other/Mixed/ Declined	5.5 (1.1–26.3)	3.7 (0.7–20.8)	.137
African American	Reference	Reference	
Education level			
Less than high school	Reference	Reference	
High School	1.4 (0.5–3.9)	1.0 (0.3–3.2)	.981
At least some college	4.3 (1.3–14.0)	2.5 (0.6–9.7)	.187

* All variables from the univariate analysis were included in the logistic regression.

for inclusion in a logistic regression and 0.05 for determining statistical significance in each final model.

3. Results

3.1. Characteristics of the study population

A total of 110 clients participated in the study. Information is not available on the number or characteristics of those who refused. Demographic and health-related characteristics of the sample are shown on Table 1, for the total sample, by MDT and MMT status, and for the MMT clinic population. The MMT study sample had a larger proportion of African Americans, a smaller proportion of employed, and a larger proportions of HIV and HCV positive patients than the MDT study sample. There were no statistically significant differences between the MMT subset of the study sample and the clinic MMT population on the variables available for comparison.

3.2. Knowledge about hepatitis C

The accuracy of respondents' knowledge about the natural history of HCV infection varied. Nearly all (92%) believed a person could die of HCV. Forty-three percent answered it would take "many years" or "a lifetime" to become sick from the virus. Few, however, understood the likelihood of becoming ill once infected: 38% answered "most" or "all" would become sick, and 22% "did not know."

Knowledge of HCV transmission was more accurate and consistent. Most of the sample knew about modes of transmission of HCV; 97% answered that needles transmit the virus and 87% answered that sharing cookers, cottons, or rinse water was a source of transmission.

However, only one third (34%) knew that there was a treatment for HCV (Table 1). Table 2 shows analyses of the variables associated with treatment knowledge. In univariate analysis, men were more likely than women to know about treatment (43% vs. 16%, $p = 0.006$). Those who were positive for HCV were more likely to know about treatment (38% vs. 14%, $p = 0.037$). Among ethnic/racial categories, Whites were most likely to know about treatment (54%), followed by Mixed/Other/Declined (44%) and Hispanic/Latinos (40%), while African Americans were least likely to know (13%; $p = 0.001$). Those with high school (31%) or at least some college (57%) were more likely to know about treatment than those with less than high school (23%; $p = 0.031$). In multivariable logistic regression, men were five times as likely as women to know about treatment, and Whites were over seven times, and Hispanic/Latinos were nearly six times, as likely as African Americans to know about treatment. We found no statistically significant differences between men and women and among ethnicities on the transmission knowledge questions.

3.3. Evaluation for hepatitis C treatment

None of the study sample reported taking ribavirin or interferon alpha for hepatitis infection. Only 30% of the sample reported that they had been evaluated for HCV treatment. Those who had been evaluated were more likely

Table 3
Univariate and multivariable logistic regression analyses for factors associated with being "Definitely interested treatment for Hepatitis C", following information on virus and treatment

Variable	Univariate analysis OR (95% CI)	Multivariable logistic regression* OR (95% CI)	*p* value for logistic regression
Treatment status			
Detoxification	4.7 (1.6–13.7)	3.7 (1.2–11.7)	.023
Maintenance	Reference	Reference	
Hepatitis C status			
Positive	0.14 (0.04–0.52	0.17 (0.04–0.65)	.010
Negative or unknown	Reference	Reference	
Education level			
Less than high school	Reference	Reference	
High School	2.2 (0.9–5.1)	2.4 (0.9–6.2)	.073
At least some college	0.6 (0.2–1.7)	0.6 (0.2–2.2)	.456

* All variables from the univariate analysis were included.

to have some college as compared to those who were not evaluated, but in logistic regression, only treatment status, i.e. being in MMT, was associated with evaluation (OR 6.6, 95% CI 1.4–30.6, $p = 0.015$).

3.4. Interest in hepatitis C treatment

After being informed about some of the risks and benefits of HCV, half (54%) were "definitely interested" in treatment (Table 1). Logistic regression analyses (Table 3) indicate that being in detoxification treatment and reporting negative or unknown HCV status were associated with being definitely interested in treatment.

4. Discussion

Three important findings emerge from this study. First, no one in our sample was being treated for HCV and few were being evaluated for treatment. Three possible barriers to receiving HCV care include patients' lack of understanding of the seriousness of HCV and treatment options, patients' low interest in receiving treatment, and patients' lack of access to HCV evaluation and treatment. When asked directly, almost three quarters of our sample expressed interest in treatment; therefore, low interest is not a barrier to treatment for this group.

In terms of understanding HCV, our survey reveals that while almost all respondents believed HCV is a fatal disease and understood how it is transmitted, many did not understand the likelihood and time course of the illness. Fewer were aware that treatment is available. Thus, knowledge about illness from HCV and knowledge about HCV treatment likely are barriers. While our study was not designed to look at financial and institutional barriers to access to HCV care, significant financial barriers existed for a population with a large proportion of uninsured, and logistical barriers existed within a health system that had wait times up to 6 months to see an HCV treatment provider. The degree to which this lack of access to treatment applies to other methadone patients is not clear. One survey of New York methadone programs reported that while 85% of drug treatment programs refer patients for further evaluation, 12% provide evaluation and treatment on site (Pratt et al., 2002).

The second important finding is that while knowledge of transmission was high, knowledge of treatment for HCV was low, particularly among women and African Americans. Most methadone programs educate patients about the adverse consequences of drug use and its sequelae (Astone et al., 2003; Strauss et al., 2003). This education likely contributed to the number of correct responses to the HCV transmission questions. Education about HCV treatment should be included as part of this education and counseling curriculum. Without knowledge of treatment, patients cannot advocate for themselves to obtain treatment. The knowledge gap we noted among gender and ethnic categories may be part of underlying disparities in access to care that have been well-described (Smedley, Stith, & Nelson, 2002; Zierler & Krieger, 1997). This finding supports the need for education about treatment options among methadone patients, particularly women and African Americans.

Third, being HCV negative or not knowing one's HCV status and being in MDT predicted higher interest in treatment. These are both unexpected findings. We anticipated that patients in MMT, and those who knew they were infected, would be more stable and ready to take on a long course of treatment with significant side effects. Because preference for MMT placement was given to patients with greater medical severity, perhaps the sicker MMT patients were less interested in a treatment with potential side effects. Another explanation is that MMT patients may be more hesitant to embark on treatment because they have lived longer, both asymptomatically with HCV and in methadone treatment. Therefore, they may be less convinced about the benefit of treatment and more concerned about the side effects. Perhaps patients in MDT believed that treatment for HCV was a mechanism for access to placement in MMT, and their interest in MMT motivated their interest in HCV treatment. Further, those who were HCV-negative or with unknown HCV status could have been more likely than HCV-infected patients to express interest in treatment because the question was hypothetical.

Our findings agree with other studies of patients in substance abuse clinics, that while a high proportion of patients are HCV infected, few are evaluated for treatment (Murrill et al., 2002; National Institutes of Health, 2002; Pratt et al., 2002; Stein, Maksad, & Clarke, 2001). We conducted our study in the midst of a shift in recommendations about which patients should be treated. It is possible that our subjects were considered ineligible for HCV treatment—based on the earlier recommendations—because of recent drug use. Stein et al. (2001) surveyed motivation for treatment among 306 methadone maintenance patients in Rhode Island and found that 53% were "definitely" or "probably" interested in treatment after being informed of the risks and benefits. This can be compared to 71% of the MMT subset of our sample that were definitely or probably interested. How well these results reflect patients' true readiness to start therapy is unclear. With the majority of our sample not knowing that treatment existed until informed of such during the interview, it is unlikely that many had thought about treatment at all.

We are aware of several limitations to our study. The sample was not selected randomly. We believe that sampling over different days and times and over a 6-month period resulted in enrollment of subjects who were not different from the clinic population, which we confirmed for the MMT population. We were unable to confirm the self-reported HCV status or treatment history data by blood testing or chart review. However, in the case of HCV status,

previous studies of HCV prevalence among methadone patients show agreement between self-report and blood testing (Best et al.,1999; Stein et al., 2001). Drawn from a single clinic, our findings may be difficult to generalize. Specifically, unique barriers may exist at our institution, such as wait times around 6 months for hepatology appointments. However, the clinic is large, urban, and draws patients from throughout the city of San Francisco, including patients who are uninsured.

Drug treatment programs, such as methadone programs, may be fertile ground to move beyond preventing hepatitis C transmission to preventing disease progression and facilitating access to treatment (Novick, 2000; Samet, Friedmann, & Saitz, 2001; Sorensen, Masson, & Perlman, 2002; Strauss, Falkin, Vassilev, Des Jarlais, & Astone, 2002). Methadone programs require adherence with clinic visits and regular contact with program staff, which potentially could facilitate adherence to HCV evaluation, treatment and side effect monitoring. Some methadone programs provide linkage to mental health, facilitating screening and monitoring for depression.

Our findings reveal a gap between access to HCV treatment and interest in treatment. By increasing knowledge of treatment, drug treatment programs may not only empower patients to seek hepatitis C treatment for themselves, but also give patients another incentive to adhere to drug treatment.

Acknowledgments

This project was supported by a grant from the Academic Senate of the University of California, San Francisco, and a grant from the Population and Quality Improvement Program in the Department of General Internal Medicine at San Francisco General Hospital. We express our gratitude to the following who assisted with data collection: Enrique Menendez, M.D., Lia Gavari, M.D., Erin Matias, Laurie Agrillo, RN, Marilyn Barnes, Laura Kahn and to the following who assisted with questionnaire design: Chris Hall, M.D., Nancy Haug, Ph.D., Brian Edlin, M.D., and Karen Seal, M.D. Preliminary data was presented at the 26th Annual Meeting of the Society of General Internal Medicine, Vancouver, BC in May of 2003 and the 65th Annual Meeting of the College on Problems in Drug Dependence, Bal Harbour, FL, in June of 2003.

References

Astone, J., Strauss, S. M., Vassilev, Z. P., & Des Jarlais, D. C. (2003). Provision of hepatitis C education in a nationwide sample of drug treatment programs. *Journal of Drug Education*, *33*, 107–117.

Backmund, M., Meyer, K., Von Zielonka, M., & Eichenlaub, D. (2001). Treatment of hepatitis C infection in injection drug users. *Hepatology*, *34*, 188–193.

Best, D., Noble, A., Finch, E., Gossop, M., Sidwell, C., & Strang, J. (1999). Accuracy of perceptions of hepatitis B and C status: cross sectional investigation of opiate addicts in treatment. *British Medical Journal*, *319*, 290–291.

Centers for Disease Control and Prevention. (1998). Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. *Morbidity and Mortality Weekly Report*, *47*, 1–39.

Davis, G. L., & Rodrigue, J. R. (2001). Treatment of chronic hepatitis C in active drug users. *New England Journal of Medicine*, *345*, 215–217.

Dhopesh, V. P., Taylor, K. R., & Burke, W. M. (2000). Survey of hepatitis B and C in addiction treatment unit. *American Journal of Drug and Alcohol Abuse*, *26*, 703–707.

Edlin, B. R., Seal, K. H., Lorvick, J., Kral, A. H., Ciccarone, D. H., Moore, L. D., & Lo, B. (2001). Is it justifiable to withhold treatment for hepatitis C from illicit-drug users? *New England Journal of Medicine*, *345*, 211–215.

Friedmann, P. D., Alexander, J. A., & D'Aunno, T. A. (1999). Organizational correlates of access to primary care and mental health services in drug abuse treatment units. *Journal of Substance Abuse Treatment*, *16*, 71–80.

Ho, S. B., Nguyen, H., Tetrick, L. L., Opitz, G. A., Basara, M. L., & Dieperink, E. (2001). Influence of psychiatric diagnoses on interferon-alpha treatment for chronic hepatitis C in a veteran population. *American Journal of Gastroenterology*, *96*, 157–164.

Johnson, M. E., Fisher, D. G., Fenaughty, A., & Theno, S. A. (1998). Hepatitis C virus and depression in drug users. *American Journal of Gastroenterology*, *93*, 785–789.

Murrill, C. S., Weeks, H., Castrucci, B. C., Weinstock, H. S., Bell, B. P., Spruill, C., & Gwinn, M. (2002). Age-specific seroprevalence of HIV, hepatitis B virus, and hepatitis C virus infection among injection drug users admitted to drug treatment in six US cities. *American Journal of Public Health*, *92*, 385–387.

National Institutes of Health. (2002). National Institutes of Health consensus development conference statement on the management of hepatitis C: 2002. Available: consensus.nih.gov/cons/116/116cdc_intro.htm.

National Institutes of Health. (1997). National Institutes of Health consensus development conference statement on the management of hepatitis C. Available: consensus.nih.gov/cons/105/105_intro.htm.

Novick, D. M. (2000). The impact of hepatitis C virus infection on methadone maintenance treatment. *Mount Sinai Journal of Medicine*, *67*, 437–443.

Piccolo, P., Borg, L., Lin, A., Melia, D., Ho, A., & Kreek, M. J. (2002). Hepatitis C virus and human immunodeficiency virus-1 co-infection in former heroin addicts in methadone maintenance treatment. *Journal of Addictive Diseases*, *21*, 55–66.

Pratt, C. C., Paone, D., Carter, R. J., & Layton, M. C. (2002). Hepatitis C screening and management practices: a survey of drug treatment and syringe exchange programs in New York City. *American Journal of Public Health*, *92*, 1254–1256.

Samet, J. H., Friedmann, P., & Saitz, R. (2001). Benefits of linking primary medical care and substance abuse services: patient, provider, and societal perspectives. *Archives of Internal Medicine*, *161*, 85–91.

Sorensen, J. L., & Copeland, A. L. (2000). Drug abuse treatment as an HIV prevention strategy: a review. *Drug and Alcohol Dependence*, *59*, 17–31.

Sorensen, J. L., Masson, C. L., & Perlman, D. C. (2002). HIV/hepatitis prevention in drug abuse treatment programs: Guidance for Research. *Science and Practice Perspectives*, *1*, 4–11.

Smedley, B. D., Stith, A. Y., & Nelson, A. R. (Eds.). (2002). *Unequal Treatment: Confronting Ethnic and Racial Disparities in Healthcare*. Washington, DC: The National Academies Press.

Stein, M. D., Maksad, J., & Clarke, J. (2001). Hepatitis C disease among injection drug users: knowledge, perceived risk and willingness to receive treatment. *Drug and Alcohol Dependence*, *61*, 211–215.

Strauss, S. M., Astone, J., Vassilev, Z. P., Des Jarlais, D. C., & Hagan, H. (2003). Gaps in the drug-free and methadone treatment program response to Hepatitis C. *Journal of Substance Abuse Treatment*, *24*, 291–297.

Strauss, S. M., Falkin, G. P., Vassilev, Z., Des Jarlais, D. C., & Astone, J. (2002). A nationwide survey of hepatitis C services provided by drug treatment programs. *Journal of Substance Abuse Treatment*, *22*, 55–62.

Sylvestre, D. L. (2002). Treating hepatitis C in methadone maintenance patients: an interim analysis. *Drug and Alcohol Dependence*, *67*, 117–123.

Umbricht-Schneiter, A., Ginn, D. H., Pabst, K. M., & Bigelow, G. E. (1994). Providing medical care to methadone clinic patients: referral vs. on-site care. *American Journal Public Health*, *84*, 207–210.

Zierler, S., & Krieger, N. (1997). Reframing women's risk: Social inequalities and HIV infection. *Annual Review of Public Health*, *18*, 401–436.

Predicting Pharmacy Costs and Other Medical Costs Using Diagnoses and Drug Claims

Yang Zhao, PhD, Arlene S. Ash, PhD,*† Randall P. Ellis, PhD,*‡ John Z. Ayanian, MD, MPP,*§¶ Gregory C. Pope, MS,‖ Bruce Bowen, PhD,** and Lori Weyuker, ASA††*

Background: Predicting health care costs for individuals and populations is essential for managing care. However, the comparative power of diagnostic and drug data for predicting future costs has not been closely examined.

Objective: We sought to compare the predictive performance of claims-based models using diagnoses, drugs claims, and combined data to predict health care costs.

Subjects: More than 1 million commercially insured, nonelderly individuals in a national (MEDSTAT MarketScan®) research database comprised our sample.

Measures: We used 1997 and 1998 drug and diagnostic profiles to predict costs in 1998 and 1999, respectively. To assess model performance, we compared R^2 values and predictive ratios (predicted costs/actual costs) for important subgroups.

Results: Models using both drug and diagnostic data best predicted subsequent-year total health care costs (highest R^2 = 0.168 versus 0.116 and 0.146 for models based on drug or diagnostic data alone, respectively), with highly accurate predictive ratios (0.95–1.05) for subgroups of patients with major medical conditions. Models predicting pharmacy costs had substantially higher R^2 values than models predicting other medical costs (highest R^2 0.493 versus 0.124). Drug-based models predicted future pharmacy costs better than diagnosis-based models (highest R^2 = 0.482 versus 0.243), whereas diagnosis-based models predicted total costs (highest R^2 = 0.146 versus 0.116) and nonpharmacy costs (highest R^2 = 0.116 versus 0.071) more effectively than drug-based models. Newer models had markedly higher R^2 values than older ones, largely because of richer data rather than model refinements.

Conclusions: Combined drug and diagnostic data predicts total health care costs better than either type of data alone. Pharmacy spending is particularly predictable from drug data, whereas diagnoses are more useful than drugs for predicting other medical costs and total costs. Using even slightly more recent data can substantially boost model performance measures; thus, model comparisons should be conducted on the same dataset.

Key Words: risk adjustment, predictive models, population disease burden, pharmacy profile, actuarial prediction

(*Med Care* 2005;43: 34–43)

Medical and financial managers increasingly use current illness indicators to predict subsequent-year health care needs and costs.[1,2] Predictive models commonly rely on some combination of demographics (typically age and gender), diagnoses recorded during medical encounters, and prescription drug utilization data.

Models using diagnoses from claims to predict future health care costs were introduced in the 1980s.[3,4] In January 2004, to better link Medicare payments to health plans to the health status of their enrolled beneficiaries, the Centers for Medicare and Medicaid Services (CMS) introduced a claims-based Diagnostic Cost Group Hierarchical Condition Category (DCG/HCC) model to partially reimburse health plans that enroll Medicare beneficiaries. This CMS-HCC model uses demographics and a diagnosis-based medical profile captured during all clinician encounters—both inpatient and outpatient—to produce a health-based measure of future medical need.[5,6] Several states also make risk-adjusted Medicaid payments to providers.[7–9]

Many studies have confirmed the ability of diagnosis-based models to predict total health care costs in privately insured populations.[10–14] However, diagnoses from multiple sites of care often accrue slowly in centralized databases, whereas outpatient pharmacy claims are generated electron-

From*D_xCG, Inc., Boston, Massachusetts; the †Department of General Internal Medicine, Boston University, Boston, Massachusetts; the ‡Department of Economics, Boston University, Boston, Massachusetts; the §Division of General Medicine, Brigham and Women's Hospital, and the ¶Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts; the ‖Program on Health Economics and Financing, RTI International, Waltham, Massachusetts; **Multimetrics, Sisters, Oregon; and ††LW Consulting, Mill Valley, California.

Funded by DxCG, Inc, in Boston, Massachusetts, which develops and markets software to implement the predictive models described in this study.

Yang Zhao, Arlene S. Ash, Randall P. Ellis, and Gregory C. Pope are DxCG shareholders.

Dr. Zhao is now Senior Health Outcomes Scientist, Eli Lilly and Company, Indianapolis, Indiana.

Reprints: Yang Zhao, PhD, Eli Lilly and Company, Lilly Corporate Center, DC 1833, Indianapolis, IN 46285. E-mail: zhao_yang@lilly.com.

ISSN: 0025-7079/05/4301-0034

ically as prescriptions are filled. Pharmacy claims have been studied in United States privately insured, Medicaid, and veterans populations.[15–25] Since 2003, the Netherlands has used a drug-based model to reimburse sickness funds for their members.[26] Drug-based models predict future total costs well but less well than diagnosis-based models.[12,15,17,23] Accuracy may be improved by using both diagnoses and drug claims as predictors.[15,20] However, no studies have combined diagnoses from all sites with comprehensive outpatient drug claims to predict total health care costs. As more health plans have both diagnoses and drug claims available, it is important to assess how predictive accuracy is affected when both types of data are combined.

The extensive risk adjustment literature during the past 2 decades has largely been devoted to predicting future total health care costs. Other than modeling for the Medicare program, in which outpatient drug use is not reimbursed, few studies examined the predictability of medical costs excluding pharmacy costs[17] and pharmacy costs alone.[27–31] Predicting pharmacy costs is particularly important in light of the recent legislation authorizing Medicare prescription drug benefits.

Compared with earlier reports,[1,3–4,7,9–11,13,14,16,17,19] the predictive performance of newer predictive models has improved markedly.[8,12,15,18,20,23,32] However, the data used are more recent and more complete. No study has systematically distinguished improvements as the result of more refined predictive models versus newer data.

In this study, we comprehensively characterized the predictive performance of different sources of data to predict various future health care costs and systematically explored whether observed improvements in predictive power were caused by newer data or more refined predictive models. Using a large, nationally representative cohort of commercially insured people younger than the age of 65, we examined 2 widely-used families of claim-based models: RxGroups® Releases 1 and 2, which rely on pharmacy data,[15] and the Diagnostic Cost Group Hierarchical Condition Category (DCG/HCC) models Releases 5 and 6, which use diagnoses.[10,32] Both models were developed by D_xCG, Inc. (Boston, MA). The Society of Actuaries recently has compared the performance of early versions of the RxGroup and DCG/HCC models with other drug and diagnosis models to predict next year's total health care costs.[12] Focusing on these 2 types of models, we systematically evaluated their predictive accuracy across 4 dimensions:

1. Model types (predictions based on drug, diagnosis or both kinds of data combined).
2. Cost outcomes (pharmacy costs, other medical costs, or total costs).
3. Model versions (previously published models[10,15] or newer, more clinically-refined models).
4. Sequential years (either 1997 data used to predict 1998 costs or 1998 data used to predict 1999 costs).

METHODS

Data Source

We used the Commercial Claims and Encounters Database of MEDSTAT's MarketScan® Research Database, which includes inpatient, outpatient, and pharmacy claims for individuals enrolled in more than 100 health plans contracting with large employers, state and local governments, and public organizations in the United States. Both fee-for-service and capitated health plans that submitted encounter data are included. The 1998–1999 data include 1.3 million individuals younger than the age of 65 who were enrolled in a participating health plan for at least 1 month in both 1998 (year 1) and 1999 (year 2), with pharmacy coverage throughout their enrollment period. The analogous 1997–1998 MarketScan data included 1.1 million individuals, and 874,000 people appeared in both the 1997–1998 and 1998–1999 data. The overlap of subjects between the 2 years is a strength of our study, enabling us to assess whether changes in predictive accuracy were related to more recent data for a relatively stable population.

All outcomes are year-2 costs: total, pharmacy (outpatient only), and nonpharmacy (total minus pharmacy). Costs include deductibles, copayments, coinsurance, and coordination-of-benefits payments. For partial-year enrollees, we annualized expenditures (actual spending divided by eligibility fraction) and used eligibility fractions as modeling weights. For example, a person leaving a health plan in June 1999 with $3000 in health care costs during the previous 6 months contributes half of an observation with annualized spending of $6000.

Pharmacy Categories

We classified all National Drug Codes (NDCs), mainly according to the drug's "function" (therapeutic indication). RxGroups differ from the "Chronic Disease Score" pharmacy model and its refinements, whose categorizations are keyed to "inferred medical condition."[16–20,25,26] The previous RxGroup model version 1.0 (developed in 2000) mapped 58,000 NDCs into 127 mutually exclusive categories (called "RxGroups") primarily based on therapeutic indication.[15] Version 2.0 classified the January 1, 2002, listing of more than 76,000 NDCs into 155 RxGroups. To better distinguish severity level and likely medical problems being treated, the newer system categorized NDCs along 4 dimensions: active ingredient(s), strength, route of administration, and dose. For example, the RxGroup "lipid-lowering agents" in RxGroup model version 1.0 was split into 2 RxGroups based on active ingredients: statin versus other. We distinguished 3 "routes" for asthma/chronic obstructive pulmonary disease drugs: injectable, oral, and inhaled; and the inhaled category split further into 3 ingredient-based groups: beta agonist, steroid, or other. The different uses of leuprolide for men and women yielded 2 distinct RxGroups: leuprolide in men (prostate cancer) and women (endometriosis).

A risk-adjustment model used to calculate payments or allocate resources should be minimally sensitive to discretionary practice patterns or coding idiosyncrasies. To increase our model's robustness to common variations in prescribing, we imposed hierarchies among RxGroups used to treat the same medical problem. For example, when the higher-ranked RxGroup "insulin," is present, the lower-ranked "oral diabetic agents" is ignored (Fig. 1A). The hierarchy for ophthalmic problems is more complex (1B), with 6 RxGroups in 3 tiers, with the model "recognizing" as many groups that may be present in the highest tier. For example, a person's predicted cost could be based on as many as all 3 ophthalmic categories 90 through 92, but only if no drugs in RxGroups 87 through 89 are recorded.

Diagnostic Categories

The Diagnostic Cost Group Hierarchical Condition Category (DCG/HCC) model uses age, sex, and diagnoses from both inpatient hospital admissions and outpatient professional medical services to record the presence of multiple medical problems and predict health care costs. More than 15,000 diagnoses map to clinically homogenous groups, called DxGroups, which further cluster into broader Condition Categories (CCs).[10,32] In updating the model from Release 5 to Release 6, we incorporated all recently introduced diagnoses and increased the number (and thus specificity) of DxGroups from 541 to 781 and of CCs from 118 to 184.

The more clinically specific DxGroups and CCs better support disease management, especially in the areas of diabetes,

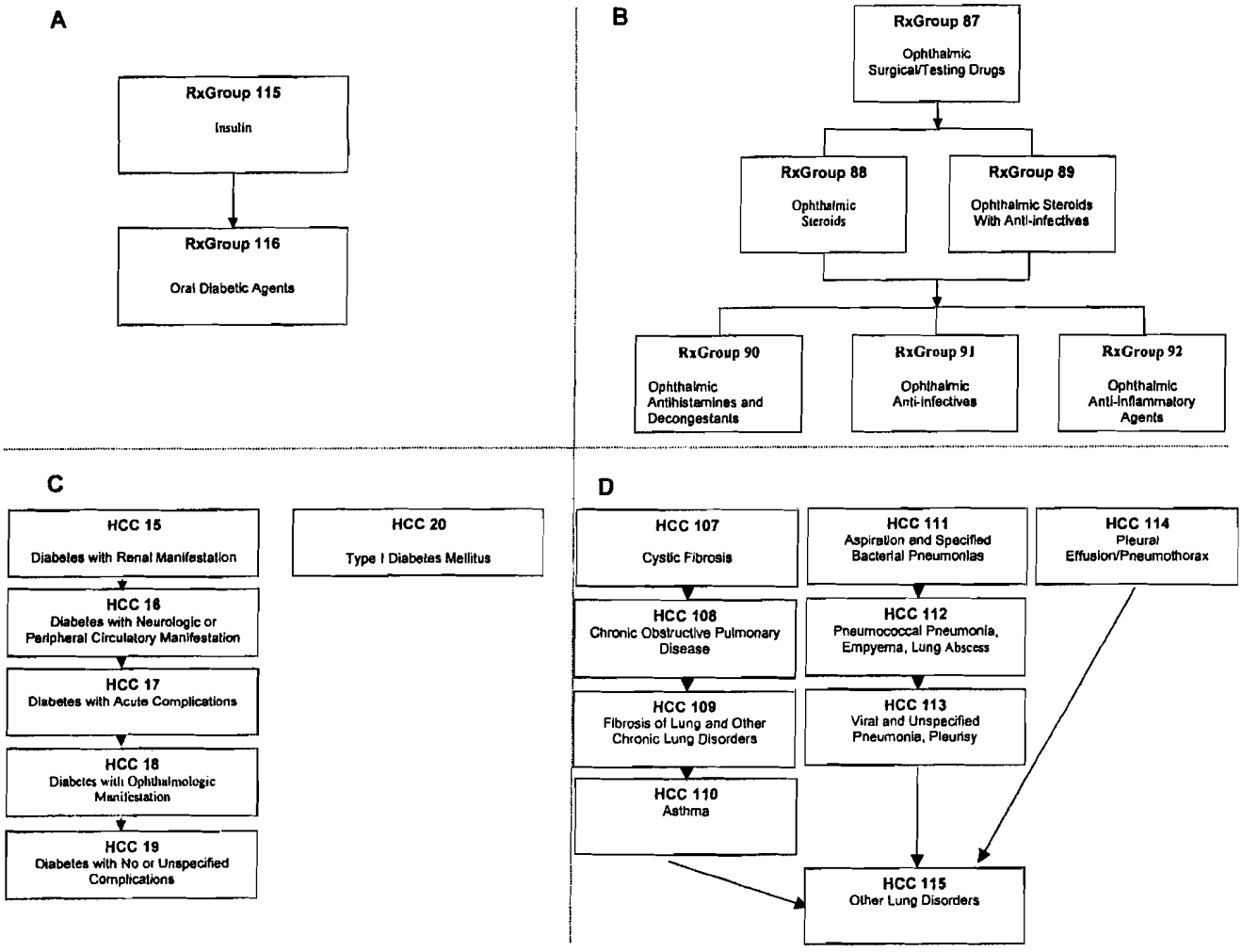


FIGURE 1. Sample RxGroups and condition category hierarchies. A, Diabetes drug hierarchy; B, ophthalmic drug hierarchy; C, diabetes condition hierarchy; D, pulmonary condition hierarchy.

TABLE 1. Demographics and Utilization Experience in Privately Insured Populations: 1997–1998 versus 1998–1999 samples*

	1997–1998	1998–1999	% Change†
Number of people	1,083,405	1,292,288	19.3
Percent female	50.2	50.7	1.0
Mean age in year 1	32.8	33.4	1.8
Age distribution			
0–17 years	26.2	25.9	−1.0
18–44 years	41.5	40.2	−3.0
45–64 years	32.3	33.8	4.7
Year-1 utilization statistics			
Percent with at least one diagnosis	71.9	73.9	2.8
Mean no. valid diagnoses per person	10.9	11.6	6.6
Mean no. distinct valid diagnoses per person	3.71	3.99	7.5
Mean no. HCCs per person	2.37	2.52	6.3
Percent with at least one prescription	63.9	66.4	3.9
Mean no. RxGroups per person	2.41	2.46	2.1
Year-2 total cost‡			
Mean	$1901	$2053	8.0
CV§	448	386	−13.9
Year-2 nonpharmacy cost			
Mean	$1531	$1601	4.6
CV§	539	471	−12.6
Year-2 pharmacy cost			
Mean	$370	$452	22.1
CV§	276	278	0.6

*For people with at least one month eligibility in both year 1 and year 2 in the MarketScan Research Database.
†All differences in means between years are significant at $P < 0.0001$ level after correcting for correlation induced by panel design.
‡Sum of inpatient, outpatient, and pharmacy costs.
§CV is the coefficient of variation defined as 100*standard deviation/mean.

liver disease, heart disease, and mental illness. We added a CC for type I Diabetes and mapped a few DxGroups, which were related to diabetes or congestive heart failure, to 2 CCs.

The CCs categorize a person's recorded medical conditions but may contain inconsistent or redundant information. Hierarchies are imposed on the CCs (leading to an HCC profile) to address clinical and statistical concerns. A hierarchy may replace a provisional diagnosis with a more definitive one (such as a specific cancer over "uncertain neoplasm" or asthma over "persistent cough"); mark the presence of or progression to a more severe disease state (such as metastatic versus nonmetastatic cancer or myocardial infarction versus angina); or resolve an inconsistency (such as, when a single individual receives codes indicating both moderate and severe developmental deficiencies). Thus, individual HCC markers are more informative than CC markers, and models based on HCC profiles are less apt to interpret redundant coding as evidence of greater medical need. Figure 1C illustrates the diabetes hierarchy; except for HCC 20 (type I diabetes mellitus), each CC dominates the next lower one; any individual can be assigned to at most one of HCCs 15 through 19. In contrast, the lung hierarchy has 3 subhierarchies (Fig. 1D); thus, a single person can be classified into as many as 3 lung HCCs. The complete CCs and hierarchies are listed elsewhere.[32]

Model Development

Using each classification system, we estimated models to predict each of our 3 cost outcomes (total, nonpharmacy and pharmacy only) in the subsequent year. We used weighted ordinary least squares regression; an observation's weight is the fraction of year 2 during which the person is eligible to incur costs. In all models, we added indicator variables to distinguish among 16 age/sex categories. In updating the RxGroups models, we included (interaction) terms when the joint effect of combinations of drugs on next year's costs was not additive. We considered all interactions explored in the earlier models,[15] and added additional interactions that clinical advisors deemed important. Twenty-seven interactions (2-, 3-, and 4-way) retained in the model pertained to at least 100 people (of 1.3 million). All were statistically significant at the 0.05 level and were judged to

TABLE 2. R^2 Values for Predicting Year-2 Total Costs in Privately Insured Populations: 1997–1998 versus 1998–1999 Samples*

	R^2 Values		% Change Attributable to	
	1997–1998	1998–1999	New Data	Both†
RxGroup model				
Old classification‡	0.084	0.115	38.2	
New classification§	0.084	0.116	38.3	
% Change attributable to new classification	0.7	0.8		39.3
DCG/HCC model				
Old classification¶	0.113	0.137	21.6	
New classification‖	0.117	0.146	25.0	
% Change attributable to new classification	3.6	6.6		29.6
Combined (Rx+DCG) model				
Old classifications‡¶	0.126	0.160	26.6	
New classifications§‖	0.131	0.168	28.9	
% Change attributable to new classification	3.5	5.3		33.4
% Change from (new) RxGroup to combined model	55.2	44.6		
% Change from (new) DCG/HCC to combined model	11.4	14.9		

*For people with at least 1 month of eligibility in both year 1 and year 2 in the MarketScan Research Database (1997–1998 sample: n = 1,083,405; 1998–1999 sample: n = 1,292,288). Total costs are the sum of inpatient, outpatient, and pharmacy costs.
†Better data and new classification(s).
‡Old drug-based model (RxGroup) predicts from 127 RxGroup drug categories.[15]
§Old diagnosis-based (DCG/HCC) model predicts from 118 hierarchical condition categories (HCCs).[10]
¶New drug-based model (RxGroup) predicts from 155 RxGroup drug categories, as described in the Methods section.
‖New diagnosis-based (DCG/HCC) model predicts from 184 hierarchical condition categories (HCCs), as described in the Methods section.

be clinically sensible. We also included interactions among coexisting medical conditions (HCCs) in the updated DCG/HCC models. We had previously examined interactions among 6 common chronic medical conditions for the Medicare program[32]: diabetes, cerebrovascular disease, congestive heart failure, vascular disease, coronary artery disease, and chronic obstructive pulmonary disease. In updating the DCG/HCC model, we explored 2- and 3-way interactions among 8 medical conditions: the original 6 plus uncompleted pregnancy and respiratory disease. To explore meaningful cost differences by age, we further tested interactions between age (younger than age 18) and all HCCs. As above, all 28 interactions retained in the model had at least 100 people and were both clinically reasonable and statistically significant ($P < 0.05$).

We excluded some RxGroups or HCCs from the models if they had a negative coefficient (although typically only slightly so), which would reduce predictions for people with such prescriptions or diagnoses. RxGroups that do not predict future costs (based on clinical judgment) also were dropped. These groups included drugs for "major diagnostic testing," drugs that are available in over-the-counter forms (ie, "OTC drugs"), and drugs commonly used for a range of typically minor medical problems such as fungal skin infection ("miscellaneous, recognized NDCs").

We enforced monotonicity in the models so that no HCC had a smaller coefficient than a less clinically severe HCC. This avoids "paying less" for a potentially more serious medical problem. For example, because respiratory arrest would otherwise have had a lower coefficient than cardiorespiratory failure and shock, we forced the model to calculate the same coefficient for both groups.

We also estimated a simple, additive "combination" model (Rx+DCG) for each outcome, using (in addition to age/sex indicators) both drug and disease predictors: RxGroups and their interactions, HCCs and their interactions, and interactions between age and HCCs. We did not explore interactions between RxGroups and HCCs for this model.

Data Analysis

For each population (1997–1998 and 1998–1999), we described its demographic distribution, year-1 diagnoses and drug use, and year-2 costs. We used t statistics for the statistical significance of the differences of means. To correct for the nonindependence of the 2 sets of MarketScan data, all tests were adjusted for the correlation between observations for the same person in the 2 data sets (see Appendix).

To measure the predictive power for each model type (RxGroup, DCG/HCC, and Rx+DCG), we assessed the mod-

TABLE 3. R^2 Values for Predicting Year-2 Total Costs Without Pharmacy in Privately Insured Populations: 1997–1998 versus 1998–1999 samples*

	R^2 Values		% Change Attributable to	
	1997–1998	1998–1999	New Data	Both†
RxGroup model				
Old classification‡	0.053	0.070	32.6	
New classification§	0.053	0.071	33.0	
% change attributable to new classification	1.1	1.4		34.5
DCG/HCC model				
Old classification¶	0.089	0.107	19.6	
New classification‖	0.094	0.116	23.1	
% Change attributable to new classification	5.5	8.5		29.8
Combined (Rx+DCG) model				
Old classification‡¶	0.096	0.115	20.6	
New classifications§‖	0.100	0.124	23.9	
% Change attributable to new classification	4.6	7.5		29.6
% Change from (new) RxGroup to combined models	88.1	75.4		
% Change from (new) DCG/HCC to combined models	6.2	6.9		

*For people with at least 1 month of eligibility in both year 1 and year 2 in the MarketScan Research Database (1997–1998 sample: n = 1,083,405; 1998–1999 sample: n = 1,292,288).
†Better data and new classification(s).
‡Old drug-based model (RxGroup) predicts from 127 RxGroup drug categories.[15]
§Old diagnosis-based (DCG/HCC) model predicts from 118 hierarchical condition categories (HCCs).[10]
¶New drug-based model (RxGroup) predicts from 155 RxGroup drug categories, as described in the Methods section.
‖New diagnosis-based (DCG/HCC) model predicts from 184 hierarchical condition categories (HCCs), as described in the Methods section.

els' R^2 values (percentage of variation in costs explained). To disentangle the effects of older and newer models and data, we examined the performance of both old and new models, for each model type, on both older and newer data. That is, for each model type (drug-based, diagnosis-based, and combined) and each of the 3 cost outcomes we applied both the older and newer versions of the model to both the older and newer datasets.

The ratio of predicted costs to actual costs within selected disease cohorts is widely used to assess model accuracy.[10–12] When a model predicts well for a group, this "predictive ratio," or PR, approximately equals 1.00; when it underpredicts, the PR is less than 1; PRs greater than 1 indicate overprediction. We used 1998–1999 as the validation sample, and applied models estimated from 1997–1998 to generate predictions for each person and PRs for 3 kinds of subgroups as identified in 1998: cohorts defined by the presence of a relevant diagnosis from either inpatient or outpatient settings; cohorts defined by the presence of a relevant drug claim; and cohorts defined by total health care costs in the initial year.

RESULTS

Demographic distributions and medical care costs changed little between 1997–1998 and 1998–1999 (Table 1). The later cohort was slightly older (mean, 33.4 versus 32.8 years). Total health care costs increased by 8.0%, reflecting a substantial increase in pharmacy costs (22.1%), and a smaller increase in other medical costs (4.6%). Relative variation, as measured by the coefficient of variation (CV, equal to 100 times the standard deviation divided by the mean) declined for total costs (−13.9%) and nonpharmacy costs (−12.6%) but not for pharmacy costs (0.6%). In the newer data, more people had at least one diagnosis; also, there were more valid diagnoses, distinct valid diagnoses, and HCCs per person. The proportion of people with any prescription and the number of distinct types of drugs (RxGroups) per person were also larger in the newer data.

The combined model (Rx+DCG) predicted total costs best, followed closely by the diagnosis model and more distantly by the drug model (Table 2). For example, with the new models and data, the respective R^2 values were 0.168 (Rx+DCG), 0.146 (DCG/HCC), and 0.116 (RxGroup). That is, moving from the RxGroup to the DCG/HCC to the combined model increased the R^2 first by 26% and then by an additional 15%. Moving from older to newer data (while holding the model fixed) also produced striking improvements. Among new models to predict total cost, the switch to new data always increased R^2 values by at least 25% and often by much more; the new drug model's R^2 was 0.084 in

the old data and 0.116 in the new, a 38% increase. In contrast, when holding the data set fixed and moving from older to newer predictive models, R^2 values always improved, although only minimally for the RxGroup models, and always modestly (the largest improvement was from 0.137 to 0.146, a 7% increase, for the DCG/HCC models in the new data).

Models predicting nonpharmacy costs (the most variable outcome) had lower R^2 values than the models predicting total costs (Table 3). The diagnosis models predicted these costs better than the drug models, and the combined models provided modestly better predictions than the diagnosis models. Models predicting pharmacy costs (the least variable outcome) had the highest R^2 values, between 0.47 and 0.49 for all models relying on drug data and between 0.21 and 0.24 for the diagnosis-based models alone (Table 4). Adding diagnoses to drug claims increased the predictive accuracy only minimally (between 1 and 3%). In contrast to the findings for predicting total and nonpharmacy costs, newer data did not yield consistently higher R^2 values (maximum increase was 0.7%), with the diagnosis-based models' R^2 values actually decreasing by 7 to 8%. Switching from older to newer models yielded only modest improvements in R^2 values (2 to 9%).

Table 5 shows means and predictive ratios for 1999 total costs (new models only) for 15 groups, 5 each based on specific kinds of information from the prior year: medical conditions, drug use, and costs. The most expensive of these groups (those who had an acute myocardial infarction or who were in the top 5% of spending during 1998) incurred costs more than 5 times as high as the 1999 average of $2053, whereas spending in the least expensive group examined (those with below median 1998 spending) had costs in 1999 that were about one third of this average.

All 3 models predicted these large cost differences with reasonable accuracy, with the largest deviations occurring for the RxGroup model in the most extreme cost-based groups. Specifically, the mean RxGroup prediction was $955 (= 1.36 times $702) for the below-median cost group, and $7407 (= 0.69 times $10,735) for the group with the highest 5% of prior-year costs. The DCG/HCC predictive ratios were 1.12 and 0.80, and the combined model predictive ratios were 1.00 and 0.88 for these same groups, respectively. The diagnosis-based model predicted costs very well for the diagnosis-identified groups (predictive ratios between 0.98 and 1.02), but somewhat underpredicted the groups defined by the use of drugs (predictive ratios of 0.81–0.90). Analogously, the drug-based model did well with the groups defined by the use of drugs (predictive ratios of 0.95–1.01) whereas underpredicting the costs of the medical condition cohorts (predictive ratios of 0.81–0.90). Only the combined model predicted group averages within 5% of actual costs for all diagnosis and drug-based subgroups.

TABLE 4. R^2 Values for Predicting Year-2 Pharmacy Costs in Privately Insured Populations: 1997–1998 versus 1998–1999 Samples*

	R^2 Values		% Change Attributable to	
	1997–1998	1998–1999	New Data	Both†
RxGroup model				
Old classification‡	0.472	0.474	0.4	
New classification§	0.479	0.482	0.7	
% Change attributable to new classification	1.6	1.9		2.3
DCG/HCC model				
Old classification¶	0.225	0.207	−8.2	
New classification‖	0.243	0.225	−7.3	
% Change attributable to new classification	7.9	8.9		0.0
Combined (Rx+DCG) model				
Old classifications‡¶	0.478	0.479	0.1	
New classifications§‖	0.491	0.493	0.3	
% Change attributable to new classification	2.8	3.0		3.1
% Change from (new) RxGroup to combined models	2.6	2.2		
% Change from (new) DCG/HCC to combined models	102.6	119.2		

*For people with at least 1 month of eligibility in both year 1 and year 2 in the MarketScan Research Database (1997–1998 sample: n = 1,083,405; 1998–1999 sample: n = 1,292,288).
†Better data and new classification(s).
‡Old drug-based model (RxGroup) predicts from 127 RxGroup drug categories.[15]
§Old diagnosis-based (DCG/HCC) model predicts from 118 hierarchical condition categories (HCCs).[10]
¶New drug-based model (RxGroup) predicts from 155 RxGroup drug categories, as described in the Methods section.
‖New diagnosis-based (DCG/HCC) model predicts from 184 hierarchical condition categories (HCCs), as described in the Methods section.

TABLE 5. Predictive Ratios for Next Year's Total Costs for Disease-, Drug Use- and Cost-Defined Groups

	n	Mean Total 1999 Cost	Predictive Ratios for New Models		
			RxGroup*	DCG/HCC†	Rx+DCG*†
Medical condition groups					
Acute myocardial infarction	2571	10,949	0.86	1.00	1.05
Asthma	38,361	3921	0.90	0.98	1.00
Chronic obstructive pulmonary disease (COPD)	35,603	5080	0.81	0.98	1.00
Depression	48,611	5167	0.85	1.01	1.01
Diabetes	33,083	7613	0.84	1.02	1.03
Drug utilization groups					
Antidepressants	90,335	5888	0.98	0.82	0.99
Asthma/COPD	83,877	3756	0.95	0.86	0.95
Diabetes	23,391	7450	1.01	0.90	1.03
Lipid-lowering	60,864	5933	1.01	0.86	1.01
Ulcer/gastroesophageal reflux disease (GERD)	80,239	6553	1.00	0.81	1.00
1998 spending percentiles					
Lowest 50%	646,144	702	1.36	1.12	1.00
Next highest 30%	387,686	1872	1.09	1.12	1.08
Next highest 10%	129,229	3638	0.99	1.01	1.04
Second highest 5%	64,615	4809	0.97	0.96	1.03
Highest 5%	64,614	10,735	0.69	0.80	0.88

Models were fit to MarketScan Research Database 1997–1998 data (n = 1,083,405) and validated on analogous 1998–1999 data (n = 1,292,288). Predictive ratios equal model-predicted 1999 costs for the specified group divided by actual costs in 1999. Medical condition groups consist of people with at least 1 relevant diagnosis in 1998 from any inpatient or outpatient setting; drug utilization groups, those with at least one relevant pharmacy fill in 1998.

*New drug-based model (RxGroup), using 155 RxGroup drug categories, as described in the Methods section.

†New diagnosis-based (DCG/HCC) model, using 184 hierarchical condition categories (HCCs), as described in the Methods section.

DISCUSSION

Predictive modeling based on claims data is an important tool for managing the financing and delivery of health care. Both the drug and diagnosis-based classification systems evaluated in this study are more clinically detailed than their precursors, making them more useful for cost profiling and disease management. Each model can identify and predict costs for clinically important subgroups. Using a recent, large, and nationally representative research database for privately insured individuals younger than the age of 65, this study compared the performance of models using different types of data (diagnoses, drugs, or both) and model versions (previously published models versus updated models) to predict different components of future health care costs (total, nonpharmacy, and pharmacy costs) in older and newer data. Combining diagnoses with drug claims substantially improved predictions of future total costs but only marginally improved on the ability of diagnosis-only models to predict nonpharmacy costs or drug-only models to predict pharmacy costs.

Drug costs were far more predictable than total or nonpharmacy cost. Relying only on the list of drugs ever used in a year (and not number of scripts, number of refills, dosages, or strength), RxGroup models explained nearly 50% of the variation in pharmacy cost during the subsequent year. Although it makes sense that current pharmaceutical costs and drug codes (NDCs) for a person are highly correlated with future pharmacy spending, our study is the first to quantify how accurately drug models can predict next year's pharmacy costs. It is likely that an RxGroup model that additionally tracks the number of prescriptions, or other volume measures, will provide even more robust predictions, especially among people who use drugs for chronic medical conditions.

Although we studied a privately insured population that was younger than the age of 65, this finding also has important implications for recently enacted pharmacy benefits in the Medicare program. Because elderly adults often have many chronic illnesses treated with drugs, their pharmacy costs may be even more predictable than what we found in a younger population. Therefore, stakeholders who have access to current patterns of drug use will be able to identify future high-cost users of pharmacy benefits. To address the potential problem of biased selection or "cherry picking" of Medicare beneficiaries with low predicted drug costs, CMS will need

access to the same drug utilization information that private health plans providing drug benefits have.

Although updated models were more predictive than earlier models for total or nonpharmacy costs, more improvement came from newer data (with more diagnoses and prescriptions) than from more clinically refined classifications. This finding is important because models developed on more recent data typically perform better than models developed from older data. Thus, credible comparisons of the performance of different models require evaluation on the same data. The more recent, richer diagnostic data provided more accurate predictions, but we cannot readily determine how much of the increased coding of diagnoses reflected a true shift in illness burden or better data collection. Increases in drug utilization can also be affected by changes in clinical guidelines. When coding and prescribing practices change rapidly, diagnosis- and drug-based models may not reliably identify true changes in need.

Current diagnosis- and drug-based models are powerful predictors of future cost. Each model captures population disease burden reasonably well and can be used to monitor or allocate use of health care resources. Drug and diagnosis models explain 12% and 15% of the variation in total cost, respectively, and the model combining both types of data explains 17% of this variation. Drug data are far superior for predicting pharmacy costs, whereas the diagnosis-based risk adjustment model better characterizes the population and more accurately predicts total and nonpharmacy cost. When more timely predictions are important, drug-based predictive models can provide an attractive alternative for predicting even total and nonpharmacy costs. As claims data become richer and more informative, the previously anticipated boundary of 20% for the explanatory power of claims-based models to predict total costs in general populations[4,33] may soon be surpassed.

ACKNOWLEDGMENTS

We are grateful to Marilyn Kramer of DxCG for helpful comments on an earlier draft of the manuscript.

REFERENCES

1. Pope GC, Ellis RP, Ash AS, et al. Principal inpatient diagnostic cost group model for Medicare risk adjustment. *Health Care Financ Rev.* 2000;21:93–118.
2. Tuohy C. A "new buzz" infects health care management. *Risk Insur.* 2004;Jan 1: http://www.riskandinsurance.com/040101_tech_1. asp.
3. Ash AS, Porell F, Gruenberg L, et al. Adjusting Medicare capitation payments using prior hospitalization. *Health Care Financ Rev.* 1989;10: 17–29.
4. Newhouse JP, Manning WG, Keeler EB, et al. Adjusting capitation rates using objective health measurers and prior utilization. *Health Care Financ Rev.* 1989;10:41–54.
5. Centers for Medicare and Medicaid Services. "Cover letter. Announcement of calendar year (CY) 2004 Medicare+Choice payment rates." Available at: http://cms.hhs.gov/healthplans/rates/2004/cover.asp.
6. Pope GC, Kautter J, Ellis RP, et al. Risk Adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev.* 2004;25:119–141.
7. Kronick R, Dreyfus T, Lee L, et al. Diagnostic risk adjustment for Medicaid: the disability payment system. *Health Care Financ Rev.* 1996;17:7–33.
8. Kronick R, Gilmer T, Dreyfus T, et al. Improving health-based payment for Medicaid beneficiaries: CDPS. *Health Care Financ Rev.* 2000;21: 29–64.
9. Weiner JP, Tucker AM, Collins AM, et al. The development of a risk-adjusted capitation payment system: the Maryland Medicaid model. *J Ambul Care Manage.* 1998;21:29–52.
10. Ash AS, Ellis RP, Pope GC, et al. Using diagnoses to describe populations and predict costs. *Health Care Financ Rev.* 2000;21:7–28.
11. Ash AS, Byrne-Logan S. How well do models work? predicting health care costs. Proceedings of the Section on Statistics in Epidemiology of the American Statistical Association, Dallas, 1998.
12. Cumming RB, Knutson D, Cameron BA, et al. A comparative analysis of claims-based methods of health risk assessment for commercial populations. Final report to the Society of Actuaries. Available at: http://www.soa.org/sections/riskadjfinalreport1. pdf. 2002.
13. Starfield B, Weiner JP, Mumford L, et al. Ambulatory care groups: a categorization of diagnoses for research and management. *Health Serv Res.* 1991;26:53–74.
14. Weiner JP, Starfield BH, Steinwachs DM, et al. Development and application of a population-oriented measure of ambulatory case-mix. *Med Care.* 1991;29:452–472.
15. Zhao Y, Ellis RP, Ash AS, et al. Measuring population health risks using inpatient diagnoses and outpatient pharmacy data. *Health Serv Res.* 2001;36:180–193.
16. Clark DO, Von Korff M, Saunders K, et al. A chronic disease score with empirically derived weights. *Med Care.* 1995;33:783–795.
17. Fishman PA, Shay DK. Development and estimation of a pediatric Chronic Disease Score using automated pharmacy data. *Med Care.* 1999;37:874–883.
18. Fishman PA, Goodman M, Hornbrook M, et al. Risk adjustment using automated pharmacy data: the RxRisk model. *Med Care.* 2003;41:84–99.
19. Johnson RE, Hornbrook MC, Nichols GA. Replicating the Chronic Disease Score (CDS) from automated pharmacy data. *J Clin Epidemiol.* 1994;47:1191–1199.
20. Gilmer T, Kronick R, Fishman P, et al. The Medicaid Rx model: drug-based risk adjustment for public programs. *Med Care.* 2001;39: 1189–1202.
21. Malone DC, Billups SJ, Valuck RJ, et al. Development of a chronic disease indicator score using a veterans affairs medical center medication database. *J Clin Epidemiol.* 1999;52:551–557.
22. Roblin DW. Physician profiling using outpatient pharmacy data as a source for case mix measurement and risk adjustment. *J Ambul Care Manage.* 1998;21:68–84.
23. Sales AE, Liu CF, Sloan KL, et al. Predicting costs of care using a pharmacy based measure risk adjustment in a veteran population. *Med Care.* 2003;41:753–760.
24. Sloan KL, Sales AE, Liu CF, et al. Construction and characteristics of the risk-V: a VA-adapted drug-based case-mix instrument. *Med Care.* 2003;41:761–774.
25. Von Korff M, Wagner EH, Saunders K. A Chronic Disease Score from automated pharmacy data. *J Clin Epidemiol.* 1992;45:197–203.
26. Lamers LM. Pharmacy Costs Groups: a risk-adjuster for capitation payment based on the use of prescribed drugs. *Med Care.* 1999;37:824–830.
27. Coulsen NE, Stuart BC. Persistence in the use of pharmaceuticals by the elderly: evidence from annual claims. *J Health Econ.* 1992;11:315–328.
28. Pauly M, Zeng Y. Adverse selection and the challenges to stand-alone prescription drug insurance. National Bureau of Economic Research working paper 9919, Cambridge, MA: NBER; August 2003.
29. Van Vliet RCJA. Predictability of individual health care expenditures. *J Risk Insur.* 1992;59:443–460.
30. Wouters AV. Disaggregated annual health care expenditures: their predictability and role as predictors. *Health Serv Res.* 1991;26:247–272.
31. Wrobel MV, Doshi J, Stuart BC, et al. Predictability of prescription drug expenditures for Medicare beneficiaries. *Health Care Financ Rev* 2003–2004;25:37–46.

32. Pope GC, Ellis RP, Ash AS, et al. Diagnostic Cost Group Hierarchical Condition Category models for Medicare risk adjustment. Final Report to the Health Care Financing Administration under Contract No. 500-95-048. Waltham, MA: Health Economics Research, Inc.; December 2000.
33. Newhouse JP. Reimbursing health plans and health providers: efficiency in production versus selection, *J Econ Lit.* 1996;34:1236–1263.

APPENDIX

Because our sample contains 874,000 people who occur in both the 1997–1998 and 1998–1999 cohorts, tests of statistical significance for differences in means between the 2 populations in Table 1 were calculated using the following formula, which takes into account the fact that measures are correlated between the 2 samples.

$$t = (\bar{X}_{98} - \bar{X}_{99}) \Big/ \sqrt{\left[\frac{s_{98}^2}{N_{98}} + \frac{s_{99}^2}{N_{99}} + \frac{874000^* s_{98} s_{99} \rho_{98,99}}{N_{98} N_{99}}\right]}$$

where $\bar{X}_i$, s_i, and N_i are the mean, standard deviation, and sample size, respectively, of some variable of interest for year $i = 98$ or 99 and $\rho_{98,99}$ is the correlation coefficient between the 1998 and 1999 samples for people who appear in both years. Because the sample sizes are large and the correlation coefficients are mostly small, tests remain powerful even after this correction. For our regression analysis developing the predictive models, we did not explicitly correct standard errors for this correlation, but instead used a higher significance threshold for deciding which variables to include.